

RESEARCH

Open Access

# Sparse covariance fitting for direction of arrival estimation

Luis Blanco<sup>1\*</sup> and Montse Nájjar<sup>1,2</sup>

## Abstract

This article proposes a new algorithm for finding the angles of arrival of multiple uncorrelated sources impinging on a uniform linear array of sensors. The method is based on sparse signal representation and does not require either the knowledge of the number of the sources or a previous initialization. The proposed technique considers a covariance matrix model based on overcomplete basis representation and tries to fit the unknown signal powers to the sample covariance matrix. Sparsity is enforced by means of a  $l_1$ -norm penalty. The final problem is reduced to an objective function with a non-negative constraint that can be solved efficiently using the LARS/homotopy algorithm. The method described herein is able to provide high resolution with a low computational burden. It proceeds in an iterative fashion solving at each iteration a small linear system of equations until a stopping condition is fulfilled. The proposed stopping criterion is based on the residual spectrum and arises in a natural way when the LARS/homotopy is applied to the considered objective function.

## 1. Introduction

### Brief summary of classical direction of arrival estimators

The estimation of the directions of arrival (DoA) of multiple sources using sensor arrays is an old problem and plays a key role in array signal processing. During the last five decades, a plethora of methods have been proposed for finding the DoA of different narrowband signals impinging on a passive array of sensors. These methods can be divided into two categories: parametric and nonparametric estimators.

Nonparametric methods include beamforming and subspace methods. The former relies on scanning the power from different locations. Exponents of this category are conventional beam-former [1] and Capon's method [2]. Conventional beamformer, a.k.a. Barlett beamformer, suffers from poor spatial resolution and cannot resolve sources within the Rayleigh resolution limit [1]. As it is well known, this lack of resolution can be mitigated only by increasing the number of sensors of the array because improving the SNR or increasing the number of time observations does not increase the resolution. On the contrary, Capon's minimum variance method can resolve sources within the Rayleigh cell if

the SNR is high enough, the number of observations is sufficient and the sources are not correlated. Unfortunately, in practice, Capon's power profile is strongly dependent on the beamwidth, which, on its turn, depends on the explored direction and in some scenarios this could lead to a resolution loss. To counteract this, an estimator of the spectral density obtained from the Capon's power estimate was derived in [3] achieving better resolution properties. Herein this method will be referred as Normalized Capon. Another well-known category of nonparametric DoA estimators is the one composed by subspace methods. These algorithms are able to provide high-resolution and outperform beamforming methods. The most prominent member of this family is MULTIPLE SIGNALS CLASSIFICATION (MUSIC) [4], it relies on an appropriate separation between signal and noise subspaces. This characterization is costly and needs a previous estimation of the number of incoming signals.

Parametric methods based on the maximum likelihood criterion [5] exhibit a good performance at expenses of a high computational cost. These techniques estimate the parameters of a given model instead of searching the maxima of the spatial spectrum. Unfortunately, they often lead to difficult multidimensional optimization problems with a heavy computational burden.

\* Correspondence: lblanco@cttc.es

<sup>1</sup>Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Castelldefels, Barcelona, Spain

Full list of author information is available at the end of the article

An interesting algorithm that lies in between the class of parametric and nonparametric techniques is the CLEAN algorithm. This method was first introduced by Högbom [6] and have applications in several areas: array signal processing, image processing, radar and astronomy. Recently, Stoica and Moses throw light on the semiparametric nature of the algorithm [7]. In broad outline, it operates in a recursive manner subtracting at each iteration a fraction of the strongest signal from the observed spatial spectrum.

For those readers interested on a more detailed and comprehensive summary of angle of arrival estimators, the authors refer them to [1,8].

### Sparse signal representation

Sparse representation of signals over redundant dictionaries is a hot topic that has attracted the interest of researchers in many fields during the last decade, such as image reconstruction [9], variable selection [10], and compressed sensing [11]. The most basic problem aims to find the sparsest vector  $\mathbf{x}$  such that  $\mathbf{y} = \mathbf{Ax}$ , where  $\mathbf{y}$  is the measured vector and  $\mathbf{A}$  is known. This matrix  $\mathbf{A}$  is called dictionary and is overcomplete, i.e., it has more columns than rows. As a consequence, without imposing a sparsity prior on  $\mathbf{x}$ , the set of equations  $\mathbf{y} = \mathbf{Ax}$  is underdetermined and admits many solutions. Formally, the objective is to minimize  $\|\mathbf{x}\|_0$  subject to  $\mathbf{y} = \mathbf{Ax}$ , where  $\|\cdot\|_0$  denotes the  $l_0$ -norm [12]. This is an intractable NP-hard combinatorial problem in general [13]. Fortunately, if the vector is sufficiently sparse the problem can be relaxed replacing the  $l_0$ -norm by a  $l_1$ -norm, defined as  $\|\mathbf{x}\|_1 = \sum_i |x_i|$ , leading to a convex optimization problem with a lower computational burden. The conditions that ensure the uniqueness of the solution were studied in [14].

In case of an observation vector contaminated by noise, a natural variation is to relax the equality constraint to allow some error tolerance  $\varepsilon \geq 0$ :

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{Ax}\|_2^2 \leq \varepsilon \quad (1)$$

or alternatively,

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \beta \quad (2)$$

where the constraint  $\|\mathbf{x}\|_1 \leq \beta$  with  $\beta \geq 0$  promotes sparsity. This formulation is known as Least Absolute Shrinkage and Selector Operator (LASSO) and was originally proposed by Tibishirani [15]. The augmented formulation of (2) is well-known in signal processing and is commonly called Basis Pursuit Denoising (BPDN) [16]:

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \tau \|\mathbf{x}\|_1 \quad \text{with} \quad \tau \geq 0 \quad (3)$$

The three formulations (1)-(3) are equivalent in the sense that the sets of solutions are the same for all the possible choices of the parameters  $\tau, \varepsilon, \beta$ . To go from one formulation to the other we only need a proper correspondence of the parameters. Nevertheless, even if the mapping between the regularization parameters exists, this correspondence is not trivial and it is possibly non-linear and discontinuous [17].

When the vector  $\mathbf{x}$  is real, the LASSO problem (2), or its equivalent formulation (3), can be solved with standard quadratic programming techniques [15]. However, these techniques are time demanding and faster methods are preferred. Osborne et al. [18] and later Efron et al. [19] proposed an efficient algorithm for solving the LASSO. This algorithm is known as ‘‘homotopy method’’ [18] or LARS (Least Angle Regression) [19]. In this article this technique will be referred to as LARS/homotopy. A variant of the traditional LASSO problem, that will be specially useful in the covariance fitting that will be addressed later on, is the so-called positive LASSO. In this case, an additional constraint over the entries of the vector  $\mathbf{x}$  is considered in the LASSO problem to enforce the components of the vector to be non-negative:

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 \\ \text{subject to} \quad \|\mathbf{x}\|_1 \leq \beta \quad \text{and} \quad x_i \geq 0 \quad \forall i \end{aligned} \quad (4)$$

The positive LASSO problem (4) can be solved in a efficient way introducing some slight modifications in the traditional LARS/homotopy. This approach was proposed by Efron et al. [19], but is not as widely known as the traditional one. Briefly, the algorithm starts with a very large value of  $\tau$ , and gradually decreases the regularization parameter, until the desired value is attained. As  $\tau$  evolves, the optimal solution for a given  $\tau$ ,  $\mathbf{x}(\tau)$ , moves on a piecewise affine path. As the minimizer  $\mathbf{x}(\tau)$  is a piecewise-linear function of  $\tau$  we only need to find the critical regularization parameters  $\tau_0, \tau_1, \tau_2, \dots, \tau_{\text{stop}}$  where the slope changes [17], these values are the so-called breakpoints. The algorithm starts with  $\mathbf{x} = \mathbf{0}$  and operates in an iterative fashion calculating the critical regularization parameters  $\tau_0 > \tau_1 > \dots > \tau_{\text{stop}} \geq 0$  and the associated minimizers  $\mathbf{x}(\tau_0), \mathbf{x}(\tau_1), \dots, \mathbf{x}(\tau_{\text{stop}})$  where an inactive component of  $\mathbf{x}$  becomes positive or an active element becomes equal to zero. Normally, the number of active components increases as  $\tau$  decreases. Nevertheless, this fact cannot be guaranteed: at some breakpoints, some entries may need to be removed from the active set.

### Sparse representation in source location

Although there are some pioneering studies carried out in the late nineties, e.g., [20,21], the application of sparse

representation to direction finding has gained noticeable interest during the last decade. Recent techniques based on sparse representation show promising results that outperform conventional high-resolution methods such as MUSIC. In [20] a recursive weighted minimum-norm algorithm called FOCUSS was presented. This algorithm considers a single snapshot and requires a proper initialization. The extension to the multiple-snapshot case was carried out in [22] and it is known as M-FOCUSS. Unfortunately, as it is described in [23], this technique is computationally expensive and requires the tuning of two hyperparameters that can affect the performance of the method significantly.

If multiple snapshots can be collected in an array of sensors, they can be used to improve the estimation of the angles of arrival. Several approaches for summarizing multiple observations have been proposed in the literature. The first of these approaches is the so-called  $l_1$ -SVD presented by Malioutov et al. [24]. This method is based on the application of a singular value decomposition (SVD) over the received data matrix and leads to a second-order cone optimization problem. This algorithm requires an initial estimation of the number of sources. Although it does not have to be exact, a small error is needed for a good performance. An underestimation or an overestimation of the number of sources provokes a degradation in the performance of the method. Even if the effect of an incorrect determination of the number of sources has no catastrophic consequences, such as the disappearance of the sources in MUSIC, the performance of the algorithm can be considerably degraded. Another important drawback is that  $l_1$ -SVD depends on a user-defined parameter which is not trivial to select. An alternative approach to summarize multiple snapshots is the use of mixed norms over multiple measurement vectors (MMV) that share the same sparsity pattern [22,25]. This formulation is useful in array signal processing, specially, when the number of snapshots is smaller than the number of sensors. If we assume that the snapshots are collected during the coherence time of the angles, the position of the sources keep identical among the snapshots; the only difference between them resides in the amplitudes of the impinging rays. Basically, this approach, which is out of the scope of the article, tries to combine multiple snapshots using the  $l_2$  norm and to promote sparsity on the spatial dimension by means of the  $l_1$ -norm. Unfortunately, this joint optimization problem is complex and requires a high computational burden. When the number of snapshots increases, the computational load becomes too high for practical real-time source location. Recently, new techniques based on a covariance matrix fitting approach have been considered to summarize multiple

snapshots, e.g., [26-28]. Basically, these methods try to fit the covariance matrix to a certain model. The main advantage of covariance fitting approaches is that they lead to convex optimization problems with an affordable computational burden. Moreover, they do not require a previous estimation of the number of incoming sources or heavy computations such as SVD of the data. It should be also pointed out that as these methods work directly with the covariance matrix less storage space is needed because they do not need to store huge amounts of time data. The technique proposed by Yardibi et al. [26] leads to an optimization problem that can be solved efficiently using Quadratic Programming (QP). In the case of the approach exposed by Picard and Weiss [27], the solution is obtained by means of linear programming (LP). The main drawback of this last method is that it depends on a user defined parameter that is difficult to adjust. In the same way, Liu et al. [29] propose a new method which is based on a hyperparameter that has been heuristically determined. On the contrary, Stoica et al. [28,30] propose an iterative algorithm named SParse Iterative Covariance-based Estimation approach (SPICE), that can be used in noisy data scenarios without the need for choosing any hyperparameter. The major drawback of this method is that it needs to be initialized.

#### Article contribution

This article proposes a simple, fast, and accurate algorithm for finding the angles of arrival of multiple sources impinging on a uniform linear array (ULA). In contrast to other methods in the literature, the proposed technique does not depend on user-defined parameters and does not require either the knowledge of the number of sources or initialization. It assumes white noise and that the point sources are uncorrelated.

The method considers a structured covariance matrix model based on over-complete basis representation and tries to fit the unknown signal powers of the model to the sample covariance. Sparsity is promoted by means of a  $l_1$ -norm penalty imposed on the powers. The final problem is reduced to an objective function with a non-negative constraint that can be solved efficiently using the LARS/homotopy algorithm, which is, in general, faster than QP [19] and LP [17]. The method described herein proceeds in an iterative manner solving at each iteration a small linear system of equations until a stopping condition is fulfilled. The proposed stopping criterion is based on the residual spectrum and arises in a natural way when the LARS/homotopy is applied to the considered objective function. From the best of our knowledge this stopping condition has never been considered before in sparse signal representation.

## 2. The proposed method: sparse covariance fitting for source location

Consider  $L$  narrowband signals  $\{x_i[k]\}_{i=1}^L$  impinging on an array of  $M$  sensors. The  $k$ th observation can be expressed as:

$$\mathbf{y}[k] = \mathbf{S}(\boldsymbol{\theta})\mathbf{x}[k] + \mathbf{w}[k] \quad k = 1, \dots, N \quad (5)$$

where  $\mathbf{x}[k] = [x_1[k] \dots x_L[k]]^T$  is the vector of unknown source signals, the matrix  $\mathbf{S}(\boldsymbol{\theta}) \in \mathbb{C}^{M \times L}$  is the collection of the steering vectors corresponding to the angles of arrival of the sources  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_L]^T$ , that is,  $\mathbf{S}(\boldsymbol{\theta}) = [\mathbf{s}(\theta_1) \dots \mathbf{s}(\theta_L)]$ , and  $\mathbf{w}[k] \in \mathbb{C}^{M \times 1}$  denotes a zero-mean additive noise, spatially, and temporally white, independent of the sources with covariance matrix  $\sigma_w^2 \mathbf{I}_M$ , being  $\mathbf{I}_M$  the identity matrix of size  $M$ .

Taking into account (5) the spatial covariance matrix can be expressed as:

$$\mathbf{R} = E\{\mathbf{y}[k]\mathbf{y}^H[k]\} = \mathbf{S}(\boldsymbol{\theta})\mathbf{P}\mathbf{S}^H(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}_M \quad (6)$$

being  $\mathbf{P} = E\{\mathbf{x}[k]\mathbf{x}^H[k]\}$ . The classical direction finding problem can be reformulated as a sparse representation problem. With this aim, let us consider an exploration grid of  $G$  equally spaced angles  $\Phi = \{\varphi_1, \dots, \varphi_G\}$  with  $G \gg M$  and  $G \gg L$ . If the set of angles of arrival of the impinging signals  $\boldsymbol{\theta}$  is a subset of  $\Phi$ , the received signal model (5) can be rewritten in terms of an overcomplete matrix  $\mathbf{S}_G$  constructed by the horizontal concatenation of the steering vectors corresponding to all the potential source locations.

$$\mathbf{y}[k] = \mathbf{S}_G \mathbf{x}_G[k] + \mathbf{w}[k] \quad (7)$$

where  $\mathbf{S}_G \in \mathbb{C}^{M \times G}$  contains the steering vectors corresponding to the angles of the grid  $\mathbf{S}_G = [\mathbf{s}_1 \dots \mathbf{s}_G]$ , with  $\mathbf{s}_i = \mathbf{s}(\varphi_i)$ , and  $\mathbf{x}_G[k] \in \mathbb{C}^{G \times 1}$  is a sparse vector. The non-zero entries of  $\mathbf{x}_G[k]$  are the positions that corresponds to the source locations. In other words, the  $n$ th element of  $\mathbf{x}_G[k]$  is different from zero and equal to the  $q$ th component of the vector  $\mathbf{x}[k]$  defined in (5), denoted by  $x_q[k]$ , if and only if  $\varphi_n = \theta_q$ . It is important to point out that the matrix  $\mathbf{S}_G$  is known and does not depend on the source locations.

The assumption that the set of angles of arrival is a subset of  $\Phi$  is only required for the derivation of the algorithm. Obviously, it does not always hold. Actually, this is a common assumption in many exploration methods in the direction finding literature (e.g., Capon, Normalized Capon, MUSIC, etc). In the case that  $\boldsymbol{\theta} \notin \Phi$ , the contribution of the sources leaks into the neighboring elements of the grid.

Bearing in mind (7) and assuming a white noise with covariance matrix  $\sigma_w^2 \mathbf{I}_M$ , the spatial covariance matrix of (5) can be rewritten in terms of  $\mathbf{S}_G$  and takes the form:

$$\mathbf{R} = E\{\mathbf{y}[k]\mathbf{y}^H[k]\} = \mathbf{S}_G \mathbf{D} \mathbf{S}_G^H + \sigma_w^2 \mathbf{I}_M \quad (8)$$

with  $\mathbf{D} = E\{\mathbf{x}_G[k]\mathbf{x}_G^H[k]\}$ . An important remark is that  $\mathbf{D} \in \mathbb{C}^{G \times G}$  is different to the source covariance matrix  $\mathbf{P} \in \mathbb{C}^{L \times L}$  introduced in (6). Actually, since only  $L^2$  entries out of  $G^2$  can differ from zero,  $\mathbf{D}$  is a sparse matrix.

A common assumption in many direction finding problems is that sources are uncorrelated. Under this assumption, the matrix  $\mathbf{D}$  is a diagonal matrix with only  $L$  non-zero entries given by  $\text{diag}(\mathbf{D}) = [p_1 \dots p_G]^T = \mathbf{p}$ , being  $\mathbf{p} \in \mathbb{R}_+^{G \times 1}$ .

Note that  $\mathbf{p}$  is a  $G \times 1$  sparse vector with non-zero entries at positions corresponding to source locations. Furthermore, the elements of  $\mathbf{p}$  are real-valued and non-negative.

To cast the problem into a positive LASSO with real variables let us make some manipulations on (8). Applying vectorization to (8) it yields:

$$\text{vec}(\mathbf{R}) = \mathbf{S}_G^* \otimes \mathbf{S}_G \text{vec}(\mathbf{D}) + \sigma_w^2 \text{vec}(\mathbf{I}_M) \quad (9)$$

where  $\otimes$  and  $\text{vec}\{\cdot\}$  denote the Kronecker product and the vectorization operator. It should be remarked that the result of  $\mathbf{S}_G^* \otimes \mathbf{S}_G \in \mathbb{C}^{M^2 \times G^2}$ .

Since  $\mathbf{D}$  is a diagonal matrix because the sources are uncorrelated, only  $G$  columns of  $\mathbf{S}_G^* \otimes \mathbf{S}_G$  have to be taken into account. Using this fact, the dimensionality of the problem can be reduced. In this way, it is straightforward to rewrite the expression (9) in terms of vector  $\mathbf{p}$  just removing some columns of  $\mathbf{S}_G^* \otimes \mathbf{S}_G$ :

$$\text{vec}(\mathbf{R}) = \tilde{\mathbf{A}} \mathbf{p} + \sigma_w^2 \text{vec}(\mathbf{I}_M) \quad (10)$$

with  $\tilde{\mathbf{A}} = [\mathbf{s}_1^* \otimes \mathbf{s}_1 \quad \mathbf{s}_2^* \otimes \mathbf{s}_2 \quad \dots \quad \mathbf{s}_G^* \otimes \mathbf{s}_G]$ . Note that  $\tilde{\mathbf{A}} \in \mathbb{C}^{M^2 \times G}$ .

Separating real and imaginary parts the above equation takes the form:

$$\begin{bmatrix} \mathbf{r}_r \\ \mathbf{r}_i \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}}_r \\ \tilde{\mathbf{A}}_i \end{bmatrix} \mathbf{p} + \begin{bmatrix} \sigma_w^2 \text{vec}(\mathbf{I}_M) \\ \mathbf{0}_{M^2 \times 1} \end{bmatrix} \quad (11)$$

where

$$\mathbf{r}_r = \text{Re}\{\text{vec}(\mathbf{R})\} \quad \tilde{\mathbf{A}}_r = \text{Re}\{\tilde{\mathbf{A}}\}$$

$$\mathbf{r}_i = \text{Im}\{\text{vec}(\mathbf{R})\} \quad \tilde{\mathbf{A}}_i = \text{Im}\{\tilde{\mathbf{A}}\}$$

In the expression (11),  $\text{vec}\{\mathbf{I}_M\}$  denotes the vectorization of the identity matrix of dimensions  $M \times M$  and  $\mathbf{0}_{M^2 \times 1}$  is a vector of zeros of size  $M^2 \times 1$ . More compactly, the expression (11) can be rewritten as:

$$\mathbf{r} = \mathbf{A}\mathbf{p} + \mathbf{n} \quad (12)$$

with obvious definitions for  $\mathbf{r}$ ,  $\mathbf{A}$ ,  $\mathbf{p}$ , and  $\mathbf{n}$ . Note that  $\mathbf{r}$  and  $\mathbf{n} \in \mathbb{R}^{2M^2 \times 1}$  and  $\mathbf{A} \in \mathbb{R}^{2M^2 \times G}$ .

Unfortunately, the spatial covariance matrix is unknown in practice and is normally replaced by the sample covariance matrix obtained from a set of  $N$  observations  $\hat{\mathbf{R}} = \frac{1}{N} \sum_{k=1}^N \mathbf{y}[k]\mathbf{y}^H[k]$ . A possible method for finding  $\mathbf{p}$  is the following constrained least squares problem:

$$\begin{aligned} & \min_{\mathbf{p}} \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}\|_2^2 \\ & \text{subject to } p_i \geq 0 \quad i = 1, \dots, G \end{aligned} \quad (13)$$

$$\|\mathbf{p}\|_1 = \sum_{j=1}^G p_j \leq \beta \quad \text{with } \beta \geq 0$$

$$\text{Where } \hat{\mathbf{r}} = \begin{bmatrix} \text{Re} \left\{ \text{vec} \left[ \hat{\mathbf{R}} \right] \right\} \\ \text{Im} \left\{ \text{vec} \left[ \hat{\mathbf{R}} \right] \right\} \end{bmatrix}.$$

Note that (13) is positive LASSO problem. The main idea behind (13) is to fit the unknown powers to the model such that the solution is sparse. The method tries to minimize the residual, or in other words, tries to maintain the fidelity of the sparse representation with the received data subject to a non-negative constraint on the powers and  $\sum_{j=1}^G p_j \leq \beta$ . This last constraint promotes sparsity, as it was exposed in (2), but also imposes a bound in the received signal power. Unfortunately, the parameter  $\beta$  is unknown and has to be estimated. Even worse, the solution of (13) is very sensitive to the parameter  $\beta$ , a little error in the estimation of the parameter can lead to a wrong solution vector.

Instead of solving (13) let us consider the next equivalent formulation:

$$\begin{aligned} & \min_{\mathbf{p}} \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}\|_2^2 + \tau \|\mathbf{p}\|_1 \\ & \text{subject to } \tau \geq 0, \quad p_i \geq 0 \quad i = 1, \dots, G \end{aligned} \quad (14)$$

The problems (13) and (14) are equivalent in the sense that the path of solutions of (13) parametrized by a positive  $\beta$  matches with the solution path (14) as  $\tau$  varies. To go from one formulation to the other one we need a proper correspondence between the parameters.

The problem (14) can be solved in an efficient way with the LARS/homotopy algorithm for positive LASSO. The method operates in an iterative fashion computing the critical regularization parameters  $\tau_0 > \tau_1 > \dots > \tau_{\text{stop}} \geq 0$  and the associated minimizers  $\mathbf{p}(\tau_0)$ ,  $\mathbf{p}(\tau_1)$ , ...,  $\mathbf{p}(\tau_{\text{stop}})$ , where an inactive component of  $\mathbf{p}$  becomes positive or an active element becomes equal to zero. Let us remark that there is only one new candidate to enter or leave the active set at each iteration (this is the ‘‘one at a time condition’’ described by Efron et al. [19]).

The algorithm is based on the computation of the so-called vector of residual correlations, or just residual correlation,  $\mathbf{b}(\tau) = \mathbf{A}^T(\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau))$  at each iteration. The method starts with  $\mathbf{p} = \mathbf{0}$  which is the solution of (14) for all the  $\tau \geq \tau_0 = 2 \max_i (\mathbf{A}^T \hat{\mathbf{r}})_i$ , being  $(\mathbf{A}^T \hat{\mathbf{r}})_i$  the  $i$ th component of the vector  $\mathbf{A}^T \hat{\mathbf{r}}$ , and proceeds in an iterative manner solving reduced-order linear systems. The whole algorithm is summarized in Algorithm 1 (see [19,31] for further details). This iterative procedure must be halted when a stopping condition is satisfied. This stopping criterion, which is the main contribution of this article, will be described later in Section 3.

It should be pointed out that the least squares error of the covariance fitting method exposed in (14) decreases at each iteration of the LARS/homotopy algorithm. This result is justified by the next two theorems.

*Theorem 1:* The sum of the powers increases monotonically at each iteration of the algorithm. Given two vectors with non-negative elements  $\mathbf{p}(\tau_{n+1})$  and  $\mathbf{p}(\tau_n)$  that are minimizers of (14) for two breakpoints  $\tau_{n+1}$  and  $\tau_n$ , respectively, with  $\tau_n > \tau_{n+1}$ , it can be stated that  $\|\mathbf{p}(\tau_{n+1})\|_1 \geq \|\mathbf{p}(\tau_n)\|_1$ .

*Proof:* See Appendix 1.

*Theorem 2:* The least squares error  $\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau)\|_2^2$  decreases at each iteration of LARS/homotopy algorithm. Given two vectors with non-negative elements  $\mathbf{p}(\tau_n)$  and  $\mathbf{p}(\tau_{n+1})$  that are minimizers of (14) for two consecutive breakpoints  $\tau_n$  and  $\tau_{n+1}$  of the LARS/homotopy, with  $\tau_n > \tau_{n+1}$ , it can be stated that  $\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_{n+1})\|_2^2 \leq \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2$ .

*Proof:* See Appendix 2.

**Algorithm 1** Proposed method

INITIALIZATION:  $\mathbf{p} = \mathbf{0}$ ,  $\tau_0 = 2 \max_i (\mathbf{A}^T \hat{\mathbf{r}})_i$ ,  $n = 0$

$J = \text{active set} = \emptyset$ ,  $I = \text{inactive set} = \mathcal{J}$

**while**  $\neq$  stopping criterion and  $\exists i \in I$  such that  $b_i > 0$   
**do**

- 1) Compute the residual correlation  $\mathbf{b} = \mathbf{A}^T(\hat{\mathbf{r}} - \mathbf{A}\mathbf{p})$
- 2) Determine the maximal components of  $\mathbf{b}$ . These will be the non-zero elements of  $\mathbf{p}(\tau_n)$  (active

components).

$$J = \arg \max\{b_j\}, \quad I = J^c$$

3) Calculate the update direction  $\mathbf{u}$  such that all the active components lead to an uniform decrease of the residual correlation (equiangular direction).

$$\mathbf{u}_J = (\mathbf{A}_J^T \mathbf{A}_J)^{-1} \mathbf{1}_J$$

4) Compute the step size  $\gamma$  such that a new element of the  $\mathbf{b}$  becomes equal to the maximal ones ( $\exists i \in I$  such that  $b_i(\tau_{n+1}) = b_{j \in J}(\tau_{n+1})$ ) or one non-zero component of  $\mathbf{p}$  becomes zero ( $\exists j \in J$  such that  $\mathbf{p}_j(\tau_{n+1}) = 0$ ).

5) Actualize  $\mathbf{p} \rightarrow \mathbf{p} + \gamma \mathbf{u}$ ,  $\tau_{n+1} = \tau_n - 2\gamma$ ,  $n = n + 1$

**end while**

### 3. Stopping criterion: the cumulative spectrum

The definition of an appropriate stopping criterion is of paramount importance because it determines the final regularization parameter  $\tau_{\text{stop}}$  and consequently the number of active positions in the solution vector. In general, larger values of  $\tau$  produce sparser solutions. Nevertheless, this fact cannot be guaranteed: at some breakpoints, some entries may need to be removed from the active set.

Most of the traditional approaches exposed in the literature for choosing the regularization parameter in discrete ill-posed problems are based on the norm of the residual error in one way or another, e.g., discrepancy principle, cross-validation, and the L-curve. Nevertheless, recent publications [32,33] suggest the use of a new parameter-choice method based on the residual spectrum. This technique is based on the evaluation of the shape of the Fourier transform of the residual error. From the best of authors' knowledge, this approach has never been used as a stopping criterion in sparse representation problems. The method exposed herein is inspired in the same idea with some slight modifications. The main difference resides in the fact that no Fourier transform needs to be computed over the residual, as it will be exposed later on, the spatial spectrum of the residual arises in a natural way when the LARS/homotopy is applied to (14). The following result is the key point of the stopping criterion proposed in this article.

*Theorem 3:* When the LARS/homotopy is applied to the problem (14), the residual correlation obtained at

the  $k$ th iteration of the algorithm, expressed as  $\mathbf{b}(\tau_k) = \mathbf{A}^T(\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_k))$ , is equivalent to the Barlett estimator applied to the residual covariance matrix  $\hat{\mathbf{C}}_k = \hat{\mathbf{R}} - \sum_{i=1}^G p_i(\tau_k) \mathbf{s}_i \mathbf{s}_i^H$ . Then, the  $i$ th component of the vector of residual correlations satisfies  $\mathbf{b}_i(\tau_k) = \mathbf{s}_i^H \hat{\mathbf{C}}_k \mathbf{s}_i$ .

*Proof:* See Appendix 3.

This theorem provides an alternative interpretation of the residual correlation at the  $k$ th iteration  $\mathbf{b}(\tau_k)$  which can be seen as a residual spatial spectrum. Bearing in mind this idea and under the assumption that the noise is zero-mean and spatially white the following parameter-choice method is proposed: to stop as soon as the residual correlation resembles white noise.

Under the assumption that the noise is spatially white, the power is distributed uniformly over all the angles of arrival and the spatial spectrum has to be flat. To determine whether the residual correlation corresponds to a white noise spectrum a statistical tool has to be considered. Several tests are available in the literature to test the hypothesis of white noise. Herein the metric that will be considered to see if the residual looks like noise is:

$$c_k(l) = \frac{\sum_{i=1}^l \mathbf{b}_i(\tau_k)}{\sum_{i=1}^G \mathbf{b}_i(\tau_k)} \quad l = 1, \dots, G \quad (15)$$

where the subindex  $k$ , with  $k = 0, \dots, k_{\text{stop}}$ , denotes the  $k$ th iteration of the LARS/homotopy algorithm. The metric  $c_k$  is a slight modification of the conventional normalized cumulative periodogram proposed by Barlett [34] and later by Durbin [35]. Traditionally, the cumulative periodogram has been defined for real-valued time series. In the real case, the spectrum is symmetric and only half of the spectrum needs to be computed. However, it can be easily extended to embrace complex-valued vectors as it is shown in (15). Throughout this entire document  $c_k$  will be referred to as normalized cumulative spectrum (NCS).

For an ideal white noise the plot of the NCS is a straight line and resembles the cumulative distribution of a uniform distribution. Thus, any distributional test, such as the Kolmogorov-Smirnov (K-S) test, can be considered to determine the "goodness of fit" between the cumulative spectrum and the theoretical straight line. In [34], Barlett proposed the use of the K-S test which is based on the largest deviation in absolute value between the cumulative spectrum and the theoretical straight line. The K-S test rejects the hypothesis of white noise whenever the maximum deviation between the

cumulative spectrum and the straight line is too large. On the contrary, the cumulative spectrum is considered white noise if it lies within the K-S limits. The upper and the lower K-S limits, as a function of index  $l$ , are given by

$$\frac{l}{G} \pm \frac{\delta}{\sqrt{MN}} \quad (16)$$

where  $\delta = 1.36$  for the 95% confidence band and  $\delta = 1.63$  for the 99% band.

Notice that the NCS does not require an accurate estimation of the noise power at the receiver. Since the cumulative spectrum (15) is normalized with respect to the average power at each  $k$ th iteration, the decision metric only depends on the shape of the spatial spectrum.

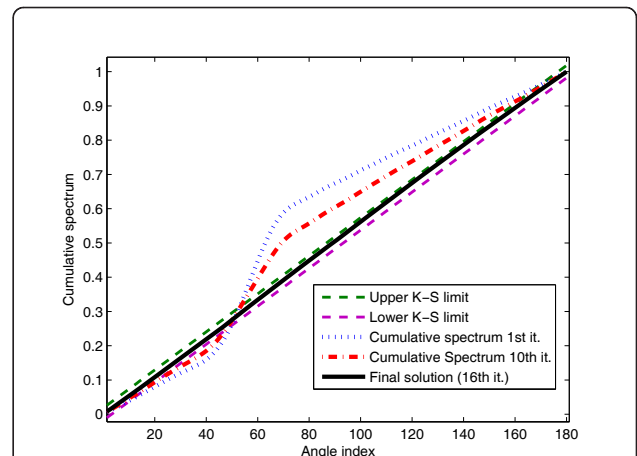
The proposed stopping condition is: to stop as soon as the residual correlation resembles white noise, that is, when the NCS lies within the K-S limits.

#### 4. Numerical results

The aim of this section is to analyze the performance of the covariance fitting method proposed in this article. To carry out this objective, some simulations have been done in Matlab. Throughout the simulations, a uniform grid with  $1^\circ$  of resolution has been considered for all the analyzed techniques. Furthermore, a zero-mean white Gaussian noise with power  $\sigma_w^2 = 1$  has been considered. The generated source signals are uncorrelated and distributed as circularly symmetric i.i.d complex Gaussian variables with zero mean. Since the same power  $P$  will be considered for all the sources, throughout this entire section the signal to noise ratio (SNR) is defined by  $\text{SNR}(\text{dB}) = 10 \log_{10} \left( \frac{P}{\sigma_w^2} \right)$ .

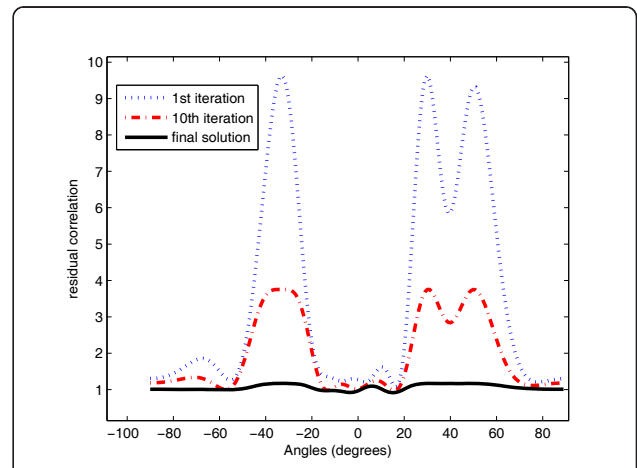
To illustrate the algorithm and the new stopping condition based on the cumulative spectrum, we have considered four uncorrelated sources located at  $-36^\circ$ ,  $-30^\circ$ ,  $30^\circ$ ,  $50^\circ$  that impinge on a ULA with  $M = 10$  sensors separated by half the wavelength. The SNR is set to 0dB and the sample covariance matrix is computed with  $N = 600$  snapshots. Figures 1 and 2 show the evolution of the NCS and the vector of residual correlations, respectively. As it is shown in Figure 1, the algorithm is stopped after 16 iterations when the NCS lies within the K-S limits of the 99% confidence band. The final solution  $\mathbf{p}$  is shown in the Figure 3. Note that the residual spectrum of the final solution in Figure 2 is almost flat and the residual correlation resembles white noise.

Next, the probability of resolution of the covariance fitting method as a function of the SNR is investigated. With this aim we have considered two uncorrelated sources located at  $-36^\circ$  and  $-30^\circ$  that impinge on a ULA

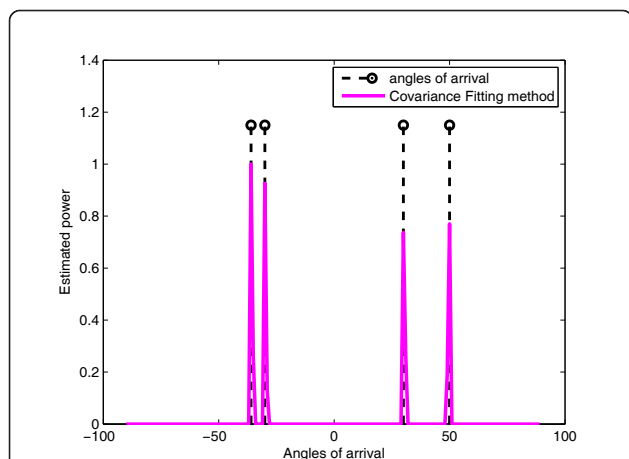


**Figure 1 Cumulative Spectrum as a function of the angle index.** The scenario is composed by four sources located at  $\theta = [-36^\circ, -30^\circ, 30^\circ, 50^\circ]$ ,  $M = 10$  sensors,  $N = 600$  snapshots,  $\text{SNR} = 0\text{dB}$ . The final solution is achieved after 16 iterations of the LARS/homotopy and it is chosen as the first one that lies within the K-S limits of the 99% confidence band.

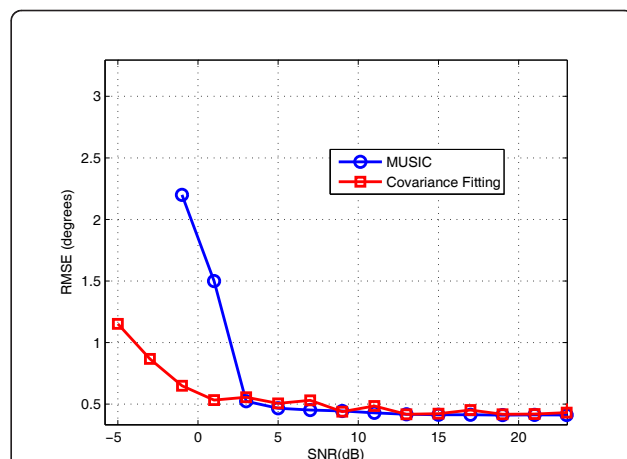
with  $M = 9$  sensors. Both sources transmit with the same power and the sample covariance has been computed with  $N = 1000$  snapshots. Figure 4 shows the results of the covariance fitting method compared to other classical estimators: MUSIC [4], Capon [2] and Normalized Capon [3]. In order to make a fair comparison between the different techniques, the number of sources of the MUSIC algorithm has been estimated



**Figure 2 Evolution of the vector of residual correlations  $\mathbf{b}$  with the iterations of the LARS/homotopy.** The scenario is composed by four sources located at  $\theta = [-36^\circ, -30^\circ, 30^\circ, 50^\circ]$ ,  $M = 10$  sensors,  $N = 600$  snapshots,  $\text{SNR} = 0\text{dB}$ . The final solution is achieved after 16 iterations of the LARS/homotopy and it is chosen as the first one that lies within the K-S limits of the 99%. Note that the residual correlation is almost flat.



**Figure 3** The power estimate of the proposed covariance matrix fitting method. Final solution  $\hat{p}$  obtained by the LARS/homotopy after 16 iterations. The settings are: four sources located at  $\theta = [-36^\circ, -30^\circ, 30^\circ, 50^\circ]$ ,  $M = 10$  sensors,  $N = 600$  snapshots, SNR = 0dB.

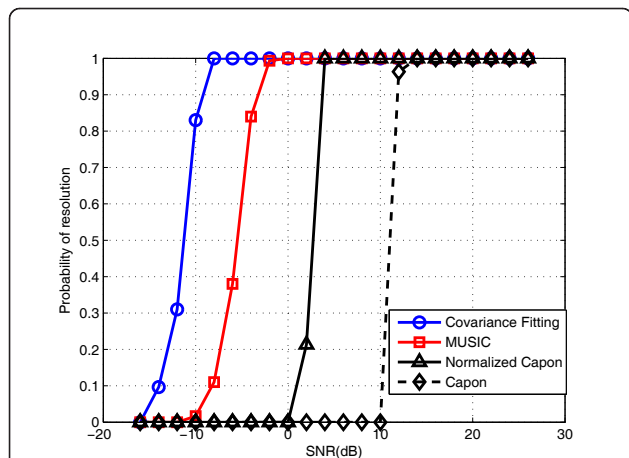


**Figure 5** Root mean square error as a function of the signal-to-noise ratio. Two uncorrelated sources separated by  $\Delta\theta = 6^\circ$ .  $M = 10$  sensors,  $N = 900$  snapshots, SNR = 0dB. The curves were obtained by averaging the results of 300 independent simulation runs.

with the Akaike information criterion (AIC) [7]. The curves in Figure 4 are averaged over 300 independent simulation runs. From this figure, it is clear that the proposed covariance fitting technique outperforms the other classical estimators and it is about 6dB better than the MUSIC algorithm and about 12dB better than the Normalized Capon method.

Next, the performance of the proposed method in terms of root mean square error (RMSE) is analyzed and presented in Figure 5. Two uncorrelated sources separated by  $\Delta\theta = 6^\circ$  that impinge on an array of  $M = 9$  sensors were taken into account in the simulations. In this case, the positions of the sources do not correspond

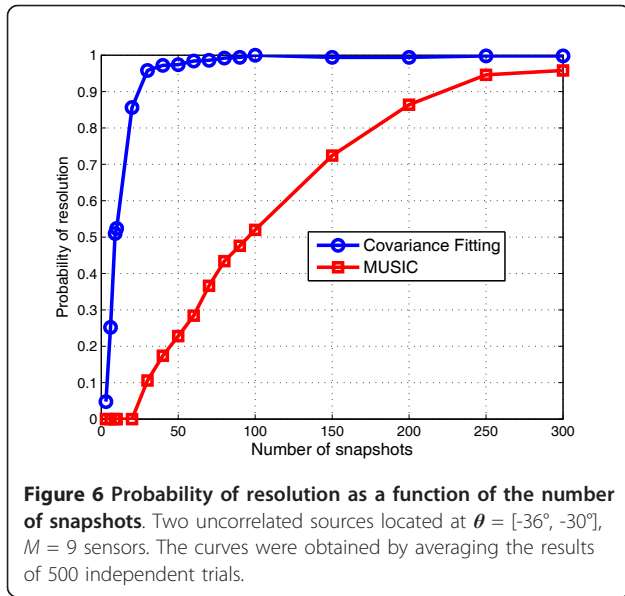
to the angles of the grid. With this aim, the angle of the first source  $\theta_1$  is generated as a random variable following a uniform distribution between  $-80^\circ$  and  $80^\circ$  and the angle of the second source is generated as  $\theta_2 = \theta_1 + \Delta\theta$ . The sample covariance has been computed with 900 snapshots. Figure 5 shows the RMSE of the proposed method and MUSIC as a function of the SNR as long as the two sources are resolved with a probability equal to 1. In the case of MUSIC the determination of the number of signal sources is performed by the AIC. The two curves are based on the average of 300 independent runs. From Figure 5 it can be concluded that at low SNR the proposed method outperforms MUSIC. When the SNR increases both methods tends to exhibit the same performance.



**Figure 4** Probability of resolution against SNR.  $\theta = [-36^\circ, -30^\circ]$ ,  $M = 9$  sensors,  $N = 1000$  snapshots. The curves were obtained by averaging the results of 300 independent simulation runs.

Finally, the resolution capability of the method as a function of the number of snapshots is investigated. The scenario considered for this purpose is the following: two sources located at  $\theta_1 = -36^\circ$  and  $\theta_2 = -30^\circ$  that impinge on a ULA with  $M = 9$  sensors. In this case, the transmitted signals have constant modulus, which is a common situation in communications applications,  $s_1(t) = e^{j\varphi_1(t)}$  and  $s_2(t) = e^{j\varphi_2(t)}$ . The signal phases  $\{\varphi_k(t)\}_{k=1}^2$  are independent and follow a uniform distribution in  $[0, 2\pi]$ . Figure 6 shows the probability of resolution of the proposed method and MUSIC as a function of the number of snapshots  $N$ . In this case the signal to noise ratio is fixed to 1 dB. As in the previous cases, in order to make a fair comparison between the two techniques, the number of sources of the MUSIC algorithm has been determined using AIC. The curves were obtained by averaging the results of 500 independent trials. Note that the covariance fitting method clearly outperforms





MUSIC and is able to resolve the two sources with a probability greater than 95% if  $N \geq 30$ .

## 5. Conclusions

A new method for finding the DoA of multiple sources that impinge on a ULA has been presented in this article. The proposed technique is based on sparse signal representation and outperforms classical direction finding algorithms, even subspace methods, in terms of RMSE and probability of resolution. The proposed technique assumes white noise and uncorrelated point sources. Furthermore, it does not require either the knowledge of the number of sources or a previous initialization.

### Appendix 1: proof of Theorem 1

The LARS/homotopy provides all the breakpoints  $\tau_0 > \tau_1 > \dots > \tau_{\text{stop}} \geq 0$  and the associated solutions  $\mathbf{p}(\tau_0)$ ,  $\mathbf{p}(\tau_1)$ , ...,  $\mathbf{p}(\tau_{\text{stop}})$  where a new component enter or leaves the support (the set of active elements) of  $\mathbf{p}(\tau)$ . It can be proved that the sum of powers increases monotonically at each iteration of the algorithm. Suppose two non-negative vectors  $\mathbf{p}(\tau_n)$  and  $\mathbf{p}(\tau_{n+1})$  that are minimizers of (14) for the regularization parameters  $\tau_n$  and  $\tau_{n+1}$ , respectively, with  $\tau_n > \tau_{n+1} \geq 0$ . The following inequality holds for the breakpoint  $\tau_n$ :

$$\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 + \tau_n \|\mathbf{p}(\tau_n)\|_1 \leq \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_{n+1})\|_2^2 + \tau_n \|\mathbf{p}(\tau_{n+1})\|_1 \quad (17)$$

Note that the regularization parameter  $\tau_n$  is the same on both sides of the inequality. The expression on the right-hand side of the inequality (17) is equal to  $\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_{n+1})\|_2^2 + \tau_{n+1} \|\mathbf{p}(\tau_{n+1})\|_1 + (\tau_n - \tau_{n+1}) \|\mathbf{p}(\tau_{n+1})\|_1$ . Therefore, the expression (17) can be rewritten as:

$$\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 + \tau_n \|\mathbf{p}(\tau_n)\|_1 \leq \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_{n+1})\|_2^2 + \tau_{n+1} \|\mathbf{p}(\tau_{n+1})\|_1 + (\tau_n - \tau_{n+1}) \|\mathbf{p}(\tau_{n+1})\|_1 \quad (18)$$

By using minimization properties, if  $\mathbf{p}(\tau_{n+1})$  is the minimizer of (14) for the regularization parameter  $\tau_{n+1}$ . Then, next inequality holds:

$$\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_{n+1})\|_2^2 + \tau_{n+1} \|\mathbf{p}(\tau_{n+1})\|_1 \leq \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 + \tau_{n+1} \|\mathbf{p}(\tau_n)\|_1 \quad (19)$$

Note that the regularization parameter  $\tau_{n+1}$  is the same on both sides of the inequality. Bearing in mind (19) and (18), it is straightforward to obtain

$$\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 + \tau_n \|\mathbf{p}(\tau_n)\|_1 \leq \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 + \tau_{n+1} \|\mathbf{p}(\tau_n)\|_1 + (\tau_n - \tau_{n+1}) \|\mathbf{p}(\tau_n)\|_1 \quad (20)$$

If the term  $\tau_n \|\mathbf{p}(\tau_n)\|_1$  is added and subtracted from expression on the right-hand side of the inequality (20), the next expression is obtained

$$\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 + \tau_n \|\mathbf{p}(\tau_n)\|_1 \leq \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 + \tau_{n+1} \|\mathbf{p}(\tau_n)\|_1 + (\tau_n - \tau_{n+1}) (\|\mathbf{p}(\tau_{n+1})\|_1 - \|\mathbf{p}(\tau_n)\|_1) \quad (21)$$

From this expression we can conclude that  $(\tau_n - \tau_{n+1}) (\|\mathbf{p}(\tau_{n+1})\|_1 - \|\mathbf{p}(\tau_n)\|_1) \geq 0$ . As  $\tau_n > \tau_{n+1} \geq 0$ , then  $\|\mathbf{p}(\tau_{n+1})\|_1 - \|\mathbf{p}(\tau_n)\|_1 \geq 0$ . Finally, we obtain  $\|\mathbf{p}(\tau_{n+1})\|_1 \geq \|\mathbf{p}(\tau_n)\|_1$ .

### Appendix 2: proof of Theorem 2

If  $\mathbf{p}(\tau_{n+1})$  is a vector with non-negative components that minimizes the problem (14) for  $\tau_{n+1} > 0$ , then the following inequality is fulfilled:

$$\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_{n+1})\|_2^2 + \tau_{n+1} \|\mathbf{p}(\tau_{n+1})\|_1 \leq \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 + \tau_{n+1} \|\mathbf{p}(\tau_n)\|_1 \quad (22)$$

which can be rewritten as:

$$\tau_{n+1} (\|\mathbf{p}(\tau_{n+1})\|_1 - \|\mathbf{p}(\tau_n)\|_1) \leq \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 - \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_{n+1})\|_2^2 \quad (23)$$

Since  $\tau_{n+1} > 0$  and  $\|\mathbf{p}(\tau_{n+1})\|_1 - \|\mathbf{p}(\tau_n)\|_1 \geq 0$ , as it was proved in Theorem 1, the following in-equality is fulfilled  $\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 - \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_{n+1})\|_2^2 \geq 0$ . Finally, we obtain  $\|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_n)\|_2^2 \geq \|\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau_{n+1})\|_2^2$ .

### Appendix 3: an alternative interpretation of the residual

The residual correlation  $\mathbf{b}$  that appears when the LARS/homotopy algorithm is applied to the problem (14), has a clear physical interpretation.

Bearing in mind (11), the residual correlation  $\mathbf{b}$  when LARS/homotopy is applied to (14) takes the form

$$\mathbf{b}(\tau) = \mathbf{A}^T (\hat{\mathbf{r}} - \mathbf{A}\mathbf{p}(\tau)) = \begin{bmatrix} \tilde{\mathbf{A}}_r^T & \tilde{\mathbf{A}}_i^T \end{bmatrix} \left( \begin{bmatrix} \hat{\mathbf{r}}_r \\ \hat{\mathbf{r}}_i \end{bmatrix} - \begin{bmatrix} \tilde{\mathbf{A}}_r \\ \tilde{\mathbf{A}}_i \end{bmatrix} \mathbf{p}(\tau) \right) \quad (24)$$

which can be rewritten in terms of complex matrices  $\tilde{\mathbf{A}}$  exposed in (10) and the sample covariance  $\hat{\mathbf{R}}$ .

$$\mathbf{b}(\tau) = \text{Re} \left\{ \tilde{\mathbf{A}}^H \left( \text{vec} \left[ \hat{\mathbf{R}} \right] - \tilde{\mathbf{A}}\mathbf{p}(\tau) \right) \right\} \quad (25)$$

The term  $\tilde{\mathbf{A}}\mathbf{p}(\tau)$  can be expressed as

$$\tilde{\mathbf{A}}\mathbf{p}(\tau) = \begin{bmatrix} \mathbf{s}_1^* \otimes \mathbf{s}_1 & \mathbf{s}_2^* \otimes \mathbf{s}_2 & \cdots & \mathbf{s}_G^* \otimes \mathbf{s}_G \end{bmatrix} \begin{bmatrix} p_1(\tau) \\ p_2(\tau) \\ \vdots \\ p_G(\tau) \end{bmatrix} = \sum_{i=1}^G p_i(\tau) \mathbf{s}_i^* \otimes \mathbf{s}_i \quad (26)$$

Since  $\mathbf{s}_i^* \otimes \mathbf{s}_i = \text{vec}(\mathbf{s}_i \mathbf{s}_i^H)$ , then

$$\tilde{\mathbf{A}}\mathbf{p}(\tau) = \text{vec} \left\{ \sum_{i=1}^G p_i(\tau) \mathbf{s}_i \mathbf{s}_i^H \right\}$$

Applying from (26) to (25) the residual correlation at breakpoint  $\tau$  yields:

$$\mathbf{b}(\tau) = \text{Re} \left\{ \tilde{\mathbf{A}}^H \left( \text{vec}[\hat{\mathbf{R}}] - \text{vec} \left[ \sum_{i=1}^G p_i(\tau) \mathbf{s}_i \mathbf{s}_i^H \right] \right) \right\} = \text{Re} \left\{ \tilde{\mathbf{A}}^H \text{vec} \left( \hat{\mathbf{R}} - \sum_{i=1}^G p_i(\tau) \mathbf{s}_i \mathbf{s}_i^H \right) \right\} \quad (27)$$

Bearing in mind the matrix  $\tilde{\mathbf{A}}$  presented in (10), the last expression can be rewritten as:

$$\mathbf{b}(\tau) = \text{Re} \left\{ \begin{bmatrix} \mathbf{s}_1^* \otimes \mathbf{s}_1^H \\ \mathbf{s}_2^* \otimes \mathbf{s}_2^H \\ \vdots \\ \mathbf{s}_G^* \otimes \mathbf{s}_G^H \end{bmatrix} \text{vec}[\hat{\mathbf{C}}_\tau] \right\} = \text{Re} \left\{ \begin{bmatrix} (\mathbf{s}_1^* \otimes \mathbf{s}_1^H) \text{vec}[\hat{\mathbf{C}}_\tau] \\ (\mathbf{s}_2^* \otimes \mathbf{s}_2^H) \text{vec}[\hat{\mathbf{C}}_\tau] \\ \vdots \\ (\mathbf{s}_G^* \otimes \mathbf{s}_G^H) \text{vec}[\hat{\mathbf{C}}_\tau] \end{bmatrix} \right\} = \text{Re} \left\{ \begin{bmatrix} \mathbf{s}_1^H \hat{\mathbf{C}}_\tau \mathbf{s}_1 \\ \mathbf{s}_2^H \hat{\mathbf{C}}_\tau \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_G^H \hat{\mathbf{C}}_\tau \mathbf{s}_G \end{bmatrix} \right\} \quad (28)$$

being  $\hat{\mathbf{C}}_\tau = \hat{\mathbf{R}} - \sum_{i=1}^G p_i(\tau) \mathbf{s}_i \mathbf{s}_i^H$ .

The  $i$ th component of  $\mathbf{b}(\tau)$  is real because it fulfills  $\mathbf{s}_1^H \hat{\mathbf{C}}_\tau \mathbf{s}_1 = \mathbf{s}_1^H \hat{\mathbf{C}}_\tau^H \mathbf{s}_1$ . Therefore, the residual correlation yields:

$$\mathbf{b}(\tau) = \begin{bmatrix} \mathbf{s}_1^H \hat{\mathbf{C}}_\tau \mathbf{s}_1 & \mathbf{s}_2^H \hat{\mathbf{C}}_\tau \mathbf{s}_2 & \cdots & \mathbf{s}_G^H \hat{\mathbf{C}}_\tau \mathbf{s}_G \end{bmatrix}^T \quad (29)$$

This result provides an alternative interpretation of the residual correlation. At each breakpoint  $\tau$ , the corresponding residual  $\mathbf{b}(\tau)$  can be seen as the Barlett estimator applied to the residual covariance matrix

$$\hat{\mathbf{C}}_\tau = \hat{\mathbf{R}} - \sum_{i=1}^G p_i(\tau) \mathbf{s}_i \mathbf{s}_i^H.$$

#### Acknowledgements

This study was partially supported by the Spanish Ministry of Economy and Competitiveness under projects TEC2011-29006-C03-01 (GRE3N-PHY) and TEC2011-29006-C03-02 (GRE3N-LINK-MAC) and by the Catalan Government under grant 2009 SGR 891.

#### Author details

<sup>1</sup>Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Castelldefels, Barcelona, Spain <sup>2</sup>Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Barcelona, Spain

#### Competing interests

The authors declare that they have no competing interests.

Received: 3 October 2011 Accepted: 17 May 2012

Published: 17 May 2012

#### References

1. HLV Trees, *Detection, Estimation, and Modulation Theory, Part IV: Optimum Array Processing*, (John Wiley & Sons, New York, USA, 2002)

2. J Capon, High-resolution frequency-wavenumber spectrum analysis. *Proc IEEE*. **57**(8), 1408–1418 (1969)
3. MA Lagunas, A Gasull, An improved maximum likelihood method for power spectral density estimation. *IEEE Trans Acoustics Speech Signal Process.* **ASSP-32**(1), 170–173 (1984)
4. R Schmidt, Multiple emitter location and signal parameter estimation. *IEEE Trans Antennas Propag.* **34**(3), 276–280 (1986). doi:10.1109/TAP.1986.1143830
5. P Stoica, A Nehorai, Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Trans Acoustics Speech Signal Process.* **38**(10), 1783–1795 (1990). doi:10.1109/29.60109
6. J Högbom, Aperture synthesis with a non-regular distribution of interferometer baselines. *Astron Astrophys.* **15**, 417–426 (1974)
7. P Stoica, R Moses, *Spectral Analysis of Signals*, (Prentice Hall, NJ, USA, 2005)
8. TE Tuncer, B Friedlander, *Classical and Modern Direction-of-Arrival Estimation*, (Elsevier Academic Press, Burlington, USA, 2009)
9. P Charbonnier, L Blanc-Feraud, G Aubert, M Barlaud, Deterministic edge-preserving regularization in computed imaging. *IEEE Trans Image Process.* **6**, 298–311 (1997). doi:10.1109/83.551699
10. H Zou, T Hastie, Regularization and variable selection via the elastic net. *J Royal Stat Soc Ser B.* **67**(2), 301–320 (2005). doi:10.1111/j.1467-9868.2005.00503.x
11. DL Donoho, Compressed sensing. *IEEE Trans Inf Theory.* **52**(4), 1289–1306 (2006)
12. S Boyd, L Vandenberghe, *Convex Optimization*, (Cambridge University Press, Cambridge, 2004)
13. DL Donoho, M Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via  $l_1$  minimization. *Proc Natl Aca Sci.* **100**(5), 2197–2202 (2003). doi:10.1073/pnas.0437847100
14. DL Donoho, For most large underdetermined systems of equations, the minimal  $l_1$ -norm near-solution approximates the sparsest near-solution. *Commun Pure Appl Math.* **59**(7), 907–934 (2006). doi:10.1002/cpa.20131
15. R Tibshirani, Regression shrinkage and selection via the lasso. *J Royal Stat Soc Ser B.* **58**, 267–288 (1996)
16. SS Chen, DL Donoho, MA Saunders, Atomic decomposition by basis pursuit. *SIAM J Sci Comput.* **20**, 33–61 (1998). doi:10.1137/S1064827596304010
17. DL Donoho, Y Tsaig, Fast solution of  $l_1$ -norm minimization problems when the solution may be sparse. *IEEE Trans Inf Theory.* **54**, 4789–4812 (2008)
18. MR Osborne, B Presnell, BA Turlach, A new approach to variable selection in least squares problems. *IMA J Numer Anal.* **20**(3), 389–403 (2000). doi:10.1093/imanum/20.3.389
19. B Efron, T Hastie, I Johnstone, R Tibshirani, Least angle regression. *Annals Stat.* **32**, 407–499 (2004). doi:10.1214/009053604000000067
20. IF Gorodnitsky, BD Rao, Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm. *IEEE Trans Signal Process.* **45**(3), 600–616 (1997). doi:10.1109/78.558475
21. J Fuchs, Linear programming in spectral estimation: application to array processing, in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Atlanta, GA, **6**, 3161–3164 (1996)
22. SF Cotter, BD Rao, KE, K Kreutz-delgado, Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans Signal Process.* **53**(7), 2477–2488 (2005)
23. T Yardibi, J Li, P Stoica, M Xue, AB Baggerboer, Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares. *IEEE Trans Aerospace Electron Syst.* **46**(1), 425–443 (2010)
24. DM Malioutov, M Çetin, AS Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans Signal Process.* **53**, 3010–3022 (2005)
25. YC Eldar, H Rauhut, Saverage case analysis of multichannel sparse recovery using convex relaxation. *IEEE Trans Inf Theory.* **56**(1), 505–519 (2010)
26. T Yardibi, J Li, P Stoica, LN Cattafesta, Sparsity constrained deconvolution approaches for acoustic source mapping. *J Acoust Soc Am.* **123**(5), 2631–2642 (2008). doi:10.1121/1.2896754
27. JS Picard, AJ Weiss, Direction finding of multiple emitters by spatial sparsity and linear programming, *International conference on Communications and information technologies (ISCIIT)*, Incheon, Korea, 1258–1262 (2009)
28. P Stoica, P Babu, J Li, Spice: A sparse covariance-based estimation method for array processing. *IEEE Trans Signal Process.* **59**(2), 629–638 (2011)

29. Z Liu, Z Huang, Y Zhou, Direction-of-arrival estimation of wideband signals via covariance matrix sparse representation. *IEEE Trans Signal Process.* **59**(9), 4256–4270 (2011)
30. P Stoica, P Babu, J Li, A sparse covariance-based method for direction of arrival estimation, *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2844–2847 (2011)
31. M Mørup, KH Madsen, LK Hansen, Approximate  $L_0$  constrained non-negative matrix and tensor factorization, *International Conference on Circuits and Systems (ISCAS)*, Seattle, WA, 1328–1331 (2008)
32. BW Rust, DP O'Leary, Residual periodograms for choosing regularization parameters for ill-posed problems. *Inverse Probl.* **24**(3), 1–30 (2008)
33. P Hansen, M Kilmer, R Kjeltdsen, Exploiting residual information in the parameter choice for discrete ill-posed problems. *BIT Numer Math.* **46**(1), 41–59 (2006). doi:10.1007/s10543-006-0042-7
34. M Barlett, *An introduction to Stochastic Processes*, (Cambridge University Press, Cambridge, 1966)
35. J Durbin, Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. *Biometrika.* **56**(1), 1–15 (1969). doi:10.1093/biomet/56.1.1

doi:10.1186/1687-6180-2012-111

**Cite this article as:** Blanco and Nájár: Sparse covariance fitting for direction of arrival estimation. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:111.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---