SPARSE SUBSPACE AVERAGING FOR ORDER ESTIMATION

Vaibhav Garg\textsuperscript{1}, David Ramírez\textsuperscript{2,3} and Ignacio Santamaría\textsuperscript{1}

\textsuperscript{1}Dept. of Communications Engineering, Universidad de Cantabria, Spain
\textsuperscript{2} Dept. of Signal Theory and Communications, Universidad Carlos III de Madrid, Spain
\textsuperscript{3} Gregorio Marañón Health Research Institute, Spain

Email: vaibhav.garg@unican.es, david.ramirez@uc3m.es, i.santamaria@unican.es

ABSTRACT

This paper addresses the problem of source enumeration for arbitrary geometry arrays in the presence of spatially correlated noise. The method combines a sparse reconstruction (SR) step with a subspace averaging (SA) approach, and hence it is named sparse subspace averaging (SSA). In the first step, each received snapshot is approximated by a sparse linear combination of the rest of snapshots. The SR problem is regularized by the logarithm-based surrogate of the $\ell_0$-norm and solved using a majorization-minimization approach. Based on the SR solution, a sampling mechanism is proposed in the second step to generate a collection of subspaces, all of which approximately span the same signal subspace. Finally, the dimension of the average of this collection of subspaces provides a robust estimate for the number of sources. Our simulation results show that SSA provides robust order estimates under a variety of noise models.

Index Terms— Array processing, source enumeration, sparse representation, subspace averaging.

1. INTRODUCTION

Detecting the number of signals received by an array of sensors is a crucial problem in many applications, such as wireless communications, radar, and biomedical and geophysical signal processing [1,2]. In array processing, this order estimation problem is known as source enumeration and boils down to the estimation of the array manifold dimension. Most common order estimation approaches are based on information-theoretic criteria [3–6], which minimize functions of the eigenvalues of the sample covariance matrix (SCM) penalized by a term that quantifies the model complexity. Many variants of these methods, based on random matrix theory, have been introduced to address the problem in the small sample regime where the number of observations is on the order of or smaller than the number of sensors [7–9]. These methods, however, are typically designed for spatially white noise and do not perform well under non-white or correlated noise. Source enumeration techniques when the noise samples have an unknown covariance matrix have recently been proposed in [10, 11], but these methods require that the noise be sufficiently weaker than the signal.

Recently, [12, 13] proposed a source enumeration technique for white noise based on subspace averaging (SA). The method was developed for uniform linear arrays (ULAs), as it exploits the shift invariance property (SIP) to generate multiple subspaces. In particular, the SIP allows us to extract submatrices from the SCM that span the same signal subspace in noiseless conditions. Then, averaging these subspaces is a mechanism that enhances the signal subspace, while cancelling out the noise subspace. In the end, the dimension of the average subspace yields an estimate the number of sources. In [14], we extended the technique to account for independent but not identically distributed (non-iid) noise models. However, this method still exploits the SIP, and it is therefore only applicable to ULAs. Moreover it entails a large computational cost.

This paper extends the SA enumeration technique to arrays with arbitrary geometry, referred to here as non-uniform linear arrays (NULAs), and to noise with arbitrary covariance matrix. The method first approximates each snapshot by a sparse linear combination of the rest of the received snapshots. In a noiseless situation, the number of nonzero coefficients of the sparse expansion would directly reveal the signal subspace dimension. In a noisy situation, the sparse expansions can be used to generate a collection of subspaces that, when averaged, will provide the order estimate using the SA method in [13]. This method, which combines sparse reconstruction with subspace averaging, is termed sparse subspace averaging (SSA).

The sparse reconstruction (SR) problem is solved by generalizing to complex-valued signals the well-known logarithm-based surrogate of the $\ell_0$-norm [15]. Using the majorization-minimization (MM) framework [16], we show that the SR problem amounts to solving a reweighted regularized least squares (LS) problem. Based on the SR solution, a sampling mechanism is proposed to generate a collection of random subspaces, all of which approximately span the same signal subspace. Finally, an average of these subspaces provides the order estimate using the SA method in [13], and the dimension of this average provides a robust estimate for the number of sources impinging on the array. Sparse reconstruction approaches have been used before in array signal processing, but mainly for direction-of-arrival (DOA) estimation problems [17–19]. Differently from these works, in this paper we use sparse reconstruction methods as a means to generate multiple approximations to the signal subspace that, when averaged, provide accurate estimates of the underlying signal subspace rank.

Our simulation results show that SSA provides robust order estimates under a variety of noise models, ranging from uncorrelated noise with different variances to noise with arbitrary covariance matrices. Moreover, unlike the method in [13], SSA can be applied to any array geometry, not necessarily a ULA.
2. PROBLEM STATEMENT

Let us consider $K$ narrowband signals impinging on an array of arbitrary geometry composed of $M$ antennas. The received signal is

$$x[n] = Hs[n] + e[n],$$

(1)

where $H$ is the $M \times K$ unknown multiple input multiple output (MIMO) channel. The signals are assumed to be uncorrelated and are modelled as $s[n] \sim \mathcal{C}\mathcal{N}_K(0, \Psi)$, where $\Psi$ is a diagonal matrix and the noise is $e[n] \sim \mathcal{C}\mathcal{N}_M(0, R_n)$, where $R_n$ is the noise covariance matrix. In this paper, we do not assume any particular structure (scale identity, diagonal) for the noise covariance matrix $R_n$.

From (1), the covariance matrix of the received signal is

$$R = \mathbb{E}[x[n]x^H[n]] = R_s + R_n,$$

where $R_s = HH^H$ is the signal covariance matrix. Note that since the MIMO channel $H$ is unknown, we can assume without loss of generality that $\Psi = I$.

If $N$ snapshots are collected in the data matrix $X = [x[1], \ldots, x[N]]$, then the sample covariance matrix is

$$\hat{R} = \frac{1}{N} \sum_{n=1}^{N} x[n]x^H[n] = \frac{1}{N} XX^H,$$

and the source enumeration or order estimation problem consists in estimating $K$ from $X$ or $\hat{R}$.

3. SPARSE SUBSPACE AVERAGING

The key idea of subspace averaging for order estimation is to average subspaces that contain (or ideally identical) versions of the signal subspace, but uncorrelated portions of the noise subspace. For ULAs, these subspaces can be obtained from consecutive subarrays by exploiting the shift invariance property [12,13,20]. When the MIMO channel $H$ has no particular structure, however, a different approach to generate these subspaces is needed. The SSA method generates the subspaces from the sparse reconstructions of different approach to generate these subspaces is needed. The SSA method generates the subspaces from the sparse reconstructions of different antennas by exploiting the shift invariance property [12,13,20]. Hence, the solution at the $t$th snapshot. However, the optimization problem in (2) is NP-hard. One well-known solution to this problem is to replace the $\ell_0$-norm with the $\ell_1$-norm [22], which can be solved by the least absolute shrinkage and selection operator (LASSO) optimization algorithm [23]. Nevertheless, the $\ell_1$-norm surrogate function can sometimes perform poorly and other proxies for the $\ell_0$-norm are preferred, such as the logarithm of the absolute value [15]. Moreover, the equality constraint in (2) is not adequate for low signal-to-noise-ratio (SNR) scenarios. Hence, in this paper the SR will be obtained as the solution to

$$\min_{\alpha_n} \|x[n] - X_n\alpha_n\|^2 + \lambda \sum_{i=1}^{N-1} \log(|\alpha_{n,i}| + \delta)$$

(3)

which is the Lagrangian form of (2) where we have substituted the $\ell_0$-norm by the log-surrogate and the equality by an inequality. We propose to solve (3) by using an MM-based approach [16]. The majorizer of the cost function in (3) is based on a first-order Taylor series of the logarithm, that is,

$$\sum_{i=1}^{N-1} \log(|\alpha_{n,i}| + \delta) \leq \sum_{i=1}^{N-1} \log(\sqrt{|\alpha_{n,i}|^2 + \delta})$$

$$\leq \sum_{i=1}^{N-1} \frac{1}{2\|\alpha_{n,i}^{(t)}\|^2 + 2\delta\|\alpha_{n,i}^{(t)}\|} \left(\|\alpha_{n,i}^{(t)}\|^2 - |\alpha_{n,i}^{(t)}|\right)$$

(4)

where $\alpha_{n,i}^{(t)}$ is the $i$th component of the solution at the $t$th iteration. Hence, the solution at the $(t + 1)$th iteration is computed from

$$\min_{\alpha_n} \|x[n] - X_n\alpha_n\|^2 + \lambda \alpha_n^H D\alpha_n,$$

(5)

where $D = \text{diag}(d_1, \ldots, d_N)$, with

$$d_i = \frac{1}{2\|\alpha_{i}^{(t)}\|^2 + 2\delta\|\alpha_{i}^{(t)}\| + \delta},$$

(6)

where a small value $\delta$ has been added to the denominator to avoid divisions by zero. Now, we can find a closed-form solution to (4) by taking the derivative and setting it to zero, which yields

$$\alpha_{n,i}^{(t+1)} = (X_n^H X_n + \lambda D)^{-1} X_n^H x[n].$$

The sparse reconstruction algorithm is shown in Algorithm 1.

3.2. Random Generation of Subspaces

In this section we develop a method to randomly generate subspaces based on the SR obtained in Section 3.1, which follows the lines
and let $X_\alpha$ coefficients associated to the subspaces contain a large common portion of the signal subspace and independent portions of the noise subspace. This translates into an averaging procedure that enhances signal coordinates in $[13]$. In noisy conditions, it is important for the SA method that the number of $\gamma_n > 1$.

Therefore, the snapshots associated to the higher values of $\alpha_n$ should be chosen more often. For more details, on this interpretation we refer the reader to $[13]$. Finally, the subspaces are the column spaces of the matrices iteratively constructed as shown in Algorithm 2. This algorithm considers two thresholds: 1) $k_{\text{max}}$, which is the maximum dimension of the signal subspace (in practice an overestimate of $K$), and is selected as the number of $|\alpha_{n,i}|$ greater than $\mu \max_{i}(|\alpha_{n,i}|)$; and 2) $\gamma_n$, which is a minimum value for the probabilities $\gamma_n$, and is chosen as $\gamma_n = \frac{1}{2} \sum_{i=1}^{N} \gamma_{n,i}$.

The random generation procedure can be repeated $T$ times to generate $T$ subspaces for each value of $\alpha_n, n = 1, \ldots, N$. Therefore, we get a total of $NT$ orthogonal projection matrices to be used in the SA procedure. The average projection matrix is

$$P = \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} P_{nt}, \quad (8)$$

Once we get $\overline{P}$, its eigenvalues $1 \geq k_1 \geq k_2 \geq \cdots \geq k_{d_{\text{max}}}$ are used to determine number of sources as

$$\bar{K}_{\text{SSA}} = \arg \max_{1 \leq k \leq d_{\text{max}}} k_k - k_{k+1}. \quad (9)$$

This process can be interpreted as sampling from a discrete distribution that we denote as $D\{X_n, \gamma_n\}$, where $\gamma_n = [\gamma_{n,1}, \ldots, \gamma_{n,N-1}]^T$ in (7) are the concentration parameters of the distribution. To shed some light on this distribution, let us explain the experiment in (7) are the concentration parameters of the distribution.

In noisy conditions, it is important for the SA method that the number of $\gamma_n > 1$.

Therefore, the snapshots associated to the higher values of $\alpha_n$ should be chosen more often. For more details, on this interpretation we refer the reader to [13]. Finally, the subspaces are the column spaces of the matrices iteratively constructed as shown in Algorithm 2. This algorithm considers two thresholds: 1) $k_{\text{max}}$, which is the maximum dimension of the signal subspace (in practice an overestimate of $K$), and is selected as the number of $|\alpha_{n,i}|$ greater than $\mu \max_{i}(|\alpha_{n,i}|)$; and 2) $\gamma_n$, which is a minimum value for the probabilities $\gamma_n$, and is chosen as $\gamma_n = \frac{1}{2} \sum_{i=1}^{N} \gamma_{n,i}$.

This process can be interpreted as sampling from a discrete distribution that we denote as $D\{X_n, \gamma_n\}$, where $\gamma_n = [\gamma_{n,1}, \ldots, \gamma_{n,N-1}]^T$ in (7) are the concentration parameters of the distribution. To shed some light on this distribution, let us explain the experiment in (7) are the concentration parameters of the distribution.

In noisy conditions, it is important for the SA method that the number of $\gamma_n > 1$.

Therefore, the snapshots associated to the higher values of $\alpha_n$ should be chosen more often. For more details, on this interpretation we refer the reader to [13]. Finally, the subspaces are the column spaces of the matrices iteratively constructed as shown in Algorithm 2. This algorithm considers two thresholds: 1) $k_{\text{max}}$, which is the maximum dimension of the signal subspace (in practice an overestimate of $K$), and is selected as the number of $|\alpha_{n,i}|$ greater than $\mu \max_{i}(|\alpha_{n,i}|)$; and 2) $\gamma_n$, which is a minimum value for the probabilities $\gamma_n$, and is chosen as $\gamma_n = \frac{1}{2} \sum_{i=1}^{N} \gamma_{n,i}$.

\textbf{Algorithm 2}: Generation of a Random Subspace

\textbf{Input:} $X_n, \alpha_n, \mu$

\textbf{Initialization:} $X_n = X_n, A_{nt} = 0$, and $L = N - 1$

Compute $\gamma_n$ from (7)

Compute $k_{\text{max}}$ as the number of $|\alpha_{n,i}|$ greater than $\mu \max_{i}(|\alpha_{n,i}|)$

Compute $\gamma_n = \frac{1}{2} \sum_{i=1}^{N} \gamma_{n,i}$

while $\text{rank}(A_{nt}) \leq k_{\text{max}}$ or $\sum_{i=1}^{N} \gamma_{n,i} \leq \gamma_n$ do

| * Sample from $D\{X_n, \gamma_n\}$ * |

$g = [g_1, \ldots, g_L]^T$, with $g_i \sim U(0, 1)$

$I = \{i \mid g_i \leq \gamma_{n,i}\}$ and $\bar{I} = \{i \mid g_i > \gamma_{n,i}\}$

$G = X_n(:, I)$

/* Append new submatrix */

$A_{nt} = [A_{nt} G]$

/* Eliminate selected directions */

$X_n = X_n(:, \bar{I}), \gamma_n = \gamma_n(I)$, and $L = |\bar{I}|$

\textbf{Output:} Projection matrix onto the subspace $P_{nt} = A_{nt}(A_{nt}^H A_{nt})^{-1} A_{nt}^H$

\textbf{Algorithm 3}: Sparse Subspace Averaging

\textbf{Input:} $X, d_{\text{max}}$

\textbf{Output:} Order estimate $\hat{K}_{\text{SSA}}$

for $n = 1, \ldots, N$ do

Find $X_n$ by removing the $n$th column from $X$

Compute $\alpha_n$ using Algorithm 1

Generate $T$ random projection matrices using Algorithm 2

Compute $P$ using (8)

Obtain $\hat{K}_{\text{SSA}}$ using (9)

\textbf{3.3. Subspace Averaging (SA) Criterion}

The random generation procedure can be repeated $T$ times to generate $T$ subspaces for each value of $\alpha_n, n = 1, \ldots, N$. Therefore, we get a total of $NT$ orthogonal projection matrices to be used in the SA procedure. The average projection matrix is

$$P = \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} P_{nt}, \quad (8)$$

Once we get $\overline{P}$, its eigenvalues $1 \geq k_1 \geq k_2 \geq \cdots \geq k_{d_{\text{max}}}$ are used to determine number of sources as

$$\hat{K}_{\text{SSA}} = \arg \max_{1 \leq k \leq d_{\text{max}}} k_k - k_{k+1}. \quad (9)$$

Basically, the criterion in (9) detects the gap between the signal and noise eigenvalues. It is noted here that to avoid numerical issues and to save computational cost, we are considering only the largest $d_{\text{max}}$ eigenvalues of $\overline{P}$, where $d_{\text{max}} \gg K$ is an overestimation of $K$. A summary of the proposed method is shown in Algorithm 3.

\textbf{4. SIMULATION RESULTS}

In this section, the performance of the proposed method is compared with some representative methods for different noise models. As a figure of merit we use the probability of correct detection ($P_D$), which is the probability that $\hat{K}_{\text{SSA}} = K$. For all simulations we assume that $K$ uncorrelated narrowband equal-power signals are impinging on a NULA with $M$ antennas. The signal-to-noise-ratio is defined as $\text{SNR} = 10 \log_{10}(\frac{\text{median}(\bar{R}_k)}{\text{median}(\bar{R}_n)}), and 2000 Monte Carlo simulations are averaged for all the results. We found that $\lambda = \frac{M}{2N}, T = 20, \mu = 0.1,$ and $d_{\text{max}} = \frac{M}{2}$ provide in general good performance over a wide range of scenarios. Since $\delta$ is a small constant that is used to avoid numerical issues, we select a $\delta = 10^{-5}$.

In the first experiment, we assume that the noise is uncorrelated across antennas with different variances at each sensor. Therefore, the noise covariance matrix is $R_n = \text{diag} \{\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2\}$, where $\sigma_n^2$ is the noise variance at the $n$th sensor. The noise variances are modeled as uniformly distributed independent random variables: $\sigma_n^2 \sim U[\sigma^2(1-\epsilon), \sigma^2(1+\epsilon)]$, where $\sigma^2$ is a common noise variance and $0 \leq \epsilon \leq 1$ allows us to control the spatial non-whiteness of the noise. Note that noise would be white for $\epsilon = 0$. As methods for comparison, we select LSMDL [8] and BIC [9], which are designed for white noise, the MDL criterion adapted to uncorrelated noises [14, 24–26] (labeled in the figures as MDL-unc), and the linear-regression method designed for colored noise in [10]. Fig. 1 shows the $P_D$ vs. $\epsilon$ for $M = 70, N = 100, K = 5$, and SNR
Fig. 1: $P_D$ vs. $\epsilon$ for diagonal noise in a scenario with $M = 70$, $N = 100$, $K = 5$, and SNR = $-2$ dB.

Fig. 2: $P_D$ vs. $\rho$ for exponentially correlated noise in a scenario with $M = 20$, $N = 80$, $K = 3$, and SNR = 10 dB.

Fig. 3: $P_D$ vs. $N$ for an arbitrary noise model in a scenario with $M = 100$, $K = 4$, and SNR = 15 dB.

In this example SSA provides robust and accurate detection results over the entire range of variation of $\epsilon$. Since LSMDL and BIC are originally designed for white noise, their performances degrade for higher values of $\epsilon$. The performance of the linear regression method in [10] also degrades for higher values of $\epsilon$. Finally, MDL-unc provides good results for higher values of $\epsilon$, but its performance degrades when the noise is nearly spatially white.

In the second experiment, we consider an exponentially correlated noise model whose noise covariance matrix is defined as $[R_{\nu}]_{i,j} = \rho^{i-j}$, where $0 \leq \rho \leq 1$ is the correlation coefficient. Fig. 2 shows $P_D$ vs. $\rho$ for an array with $M = 20$ antennas, $N = 80$ snapshots, $K = 3$ sources, and SNR = 10 dB. For this noise model, we are comparing the results of MDL-unc, the method in [10] and the method in [27], which applies a rank-reduction preprocessing step followed by an order estimation criterion based on canonical correlation analysis (CCA). As Fig. 2 suggests, the proposed SSA method again provides a robust solution for the whole range of $\rho$.

In this work we have presented an order estimation technique for arbitrary unknown covariance matrix, that is, the noise covariance matrix is drawn from a complex Wishart distribution with $M$ degrees of freedom and scale matrix proportional to $I$. The scenario has $M = 100$ antennas, $K = 4$ sources, and the number of snapshots $N$ varies up to 100, so that we are in the small sample regime. The SNR for this experiment is chosen as $\text{SNR} = 15$ dB, because when the noise covariance matrix is arbitrary, the SNR must be sufficiently high for any method to work. Since CCA and method in [10] do not usually provide good results with few snapshots, we are considering $N \geq 15$ for these methods, however, lower values of $N$ are considered for SSA. As it is shown in Fig. 3, CCA does not perform well for this noise model in the small sample regime, while the method in [10] provides good results for sufficiently large $N$. Finally, SSA performs satisfactorily even when $N/M$ is around 0.1.

5. CONCLUSIONS

All other methods fail when $\rho$ increases.

In the last experiment, we consider Gaussian noise with arbitrary unknown covariance matrix, that is, the noise covariance matrix is arbitrary, unknown spatial correlation is around 0. In this case SSA provides robust and accurate detection results over the entire range of variation of $\epsilon$. Since LSMDL and BIC are originally designed for white noise, their performances degrade for higher values of $\epsilon$. The performance of the linear regression method in [10] also degrades for higher values of $\epsilon$. Finally, MDL-unc provides good results for higher values of $\epsilon$, but its performance degrades when the noise is nearly spatially white.
6. REFERENCES


