Regular Sparse Array Direction of Arrival Estimation in One Dimension

Ferre Knaepkens, Annie Cuyt, Wen-Shin Lee, Member, IEEE, and Dirk I. L. de Villiers®, Senior Member, IEEE

Abstract—Traditionally, regularly spaced antenna arrays follow the spatial Nyquist criterion to guarantee an unambiguous analysis. We present a novel technique that makes use of two sparse non-Nyquist regularly spaced antenna arrays, where one of the arrays is just a shifted version of the other. The method offers several advantages over the use of traditional dense Nyquist-spaced arrays, while maintaining a comparable algorithmic complexity for the analysis. Among the advantages we mention: an improved resolution for the same number of receivers and reduced mutual coupling effects between the receivers, both due to the increased separation between the antennas. Because of a shared structured linear system of equations between the two arrays, as a consequence of the shift between the two, the analysis of both is automatically paired, thereby avoiding a computationally expensive matching step as is required in the use of so-called co-prime arrays. In addition, an easy validation step allows to automatically detect the precise number of incoming signals, which is usually considered a difficult issue. At the same time, the validation step improves the accuracy of the retrieved results and eliminates unreliable results in the case of noisy data. The performance of the proposed method is illustrated with respect to the influence of noise as well as to the effect of mutual coupling.

Index Terms—Array antennas, direction of arrival (DOA) estimation, sparse arrays.

I. INTRODUCTION

DIRECTION of arrival (DOA) estimation, using array antenna systems, is a topic of increasing interest in a variety of applications including radar, remote sensing, radio frequency interference mitigation, and smart wireless networks [1]–[3]. One of the most well-known limitations in regularly spaced antenna array systems is, arguably, the requirement that the elements should have spacing closer than a half-wavelength (the spatial Nyquist criterion) in order to avoid aliasing resulting in ambiguous arrival angle estimates. Unique results can be obtained for larger spacings if a limited range of near-broadside receiving angles are considered, or if the antenna element patterns exhibit zeros in the endfire directions, but in general, this still limits the allowable spacing to distances close to the Nyquist limit.

A wealth of research is available in the antenna array and signal processing literature on algorithms which reliably estimate the DOA of incoming signals in real world noisy environments. The most popular ones are the MUliple SIgnal Classification (MUSIC) [4], the Estimation Parameter via Rotational Invariance Technique (ESPRIT) [5], and the matrix pencil method [6]. The requirement of close antenna element spacing limits the achievable resolution of an array system with a fixed number of receivers (or sensors) according to the well-known diffraction (or Rayleigh) limit

$$\Delta_\phi \approx \frac{\lambda}{D} \quad \text{(1)}$$

where $\Delta_\phi$ is the resolution in radians, $\lambda$ is the operating wavelength, and $D$ is the length of the array (in the linear array case). A system with $M$ elements spaced at $\lambda/2$ will thus have a resolution limited to $\Delta_\phi \approx 2/(M - 1)$, explicitly stating the inverse relationship between resolution and the number of elements. Several sparse array configurations, with elements spaced at distances larger than the Nyquist limit, have been proposed including minimum redundancy arrays (MRAs) [7], nested arrays [8], and co-prime arrays [9]. All these configurations have the advantage of increased resolution, for a given number of sensors, over that of a uniform linear array (ULA). The main disadvantage, however, is the increased computational burden associated with estimating the correct angles of arrival in noisy systems, which follows from a combinatorial matching problem that arises when resolving the introduced aliasing.

A disadvantage of densely spaced antenna elements is the effect of mutual coupling, which causes uncertainty in the so-called array manifold, that is the collection of received steering vectors from all possible directions. This in turn leads to errors in the estimated arrival angles, which is normally reduced through extensive calibration of the system. Since accurate estimation of the installed mutual coupling matrix is difficult, in addition to often being time variant, many popular calibration techniques exist [10]–[13]. Typically, the computational cost of the mutual coupling estimation can be high, especially for iterative techniques. To overcome this, some methods that use auxiliary sensors have been proposed, but these come at a cost of reduced effective aperture [14], [15]. A comparison of the performance of several sparse-array methods in the presence of mutual coupling is given in [16].
In this article, we present a DOA estimation method that exploits some recent progress in exponential analysis and completely removes the dense Nyquist spacing requirement in ULAs [17]. Removing the Nyquist requirement allows for larger spacing between individual antenna elements, leading naturally to improved angular resolution as well as generally reduced mutual coupling. The larger antenna element spacing implies an increase in the size of the antenna array, which depending on the application, may not pose a problem [9], [18], [19]. The method makes use of two interleaved sparse ULAs, where one of them is just a shifted version of the other. The set of angles and their aliases obtained from the first sparse ULA via exponential analysis, is subsequently intersected with information received from the second ULA via the solution of a structured linear system, to retrieve the correct directions. Due to introducing a second ULA that is a shifted version of the first one, a combinatorial search step that other co-prime array methods require, is avoided. Instead, the results of the two ULAs are directly linked by their computation. In addition to handling the aliasing problem of sparse arrays in an efficient manner, the proposed technique can automatically determine the number of received signals with no prior user input, and allows detection of fully correlated signals from different directions without any special treatment.

Note that any 1-D exponential analysis solver could be used on the first ULA. Examples are MUSIC [4], ESPRIT [5], the matrix pencil method [6], simultaneous QR factorization [20], a generalized overdetermined eigenvalue solver [21], and the approximate Prony method [22]–[24].

This article is structured as follows. First, in Section II, the notation is introduced and the problem statement of DOA estimation is described. Subsequently, in Section III, standard Nyquist-based exponential analysis is discussed and we introduce an approach to determine the number of received signals, called the sparsity in exponential analysis, via density based cluster analysis. Section IV considers sparse ULAs, for which a sub-Nyquist sampling technique is formulated. Finally, we conclude with some examples in Section V.

II. Problem Formulation

Consider an ULA with $M$ antenna elements receiving $n$ ($2n \leq M$) narrowband signals at frequency $\omega$, as illustrated for the $i$th signal in Fig. 1. The antenna elements are omnidirectional and arranged equidistantly along the $x$-axis with a spacing $d < \lambda/2$ to prevent aliasing. The objective is to determine the directions of the received narrowband signals, in other words the angles $\phi_i$, as seen in Fig. 1. The narrowband signal $S_i(t)$ at time $t$ can be expressed as

$$S_i(t) = s_i(t) \exp(j \omega t), \quad s_i(t) = a_i(t) \exp(j p_i(t))$$

where $a_i(t)$ and $p_i(t)$ denote the slowly varying amplitude and phase, respectively. Assuming a plane wave incidence of $S_i(t)$ on the ULA, the time delay between consecutive antennas is given by $\tau_i = d \cos(\phi_i)/c$, where $c$ is the propagation velocity of the signal (speed of light in this case). We denote the output of the ULA at the $m$th antenna element at time $t$ by $f_m(t)$,

$$f_m(t) = \sum_{i=1}^{n} S_i(t - m \tau_i), \quad \tau_i = \frac{d \cos(\phi_i)}{c}$$

because of the time delay with respect to the reference antenna $f_0(t)$. The element $f_0$ receives the sum of the signals $S_i(t)$ at time $t$. The narrowband assumption (the signal response does not change appreciably during the transit of the array) allows us to write

$$s_i(t - m \tau_i) \approx s_i(t)$$

or equivalently

$$S_i(t - m \tau_i) = s_i(t - m \tau_i) \exp(j \omega(t - m \tau_i)) \approx S_i(t) \exp(-j \omega m \tau_i)$$

hence, the output $f_m(t)$ of the ULA, given by (2), can be rewritten as

$$f_m(t) = \sum_{i=1}^{n} S_i(t - m \tau_i) \approx \sum_{i=1}^{n} S_i(t) \exp(-j \omega m \tau_i).$$

Consequently, for every signal the time delay is expressed as a phase shift, depending on the direction $\phi_i$. From these time delays, which are identified by exponential analysis, it is possible to infer the directions $\phi_i$ of the received signals.

III. Standard Exponential Analysis

At a fixed time $t$, the output of the ULA is called a snapshot. This snapshot consists of the samples

$$f_m(t) = \sum_{i=1}^{n} S_i(t) \exp\left(-\frac{j \omega m d \cos(\phi_i)}{c}\right)$$

for $m = 0, \ldots, M - 1$, $M \geq 2n$, and is used to formulate a 1-D exponential analysis problem. We introduce the notations

$$f_m = f_m(t), \quad a_i = S_i(t)$$

$$\psi_i = -\frac{j \omega m d \cos(\phi_i)}{c}, \quad \Psi_i = \exp(\psi_i d).$$

We refer to the $f_m$ as the samples, the $a_i$ as the coefficients, the $\psi_i$ as the exponents, and the $\Psi_i$ as the base terms of the exponential analysis problem. Note that none of the notations...
depend on $t$ since, for the moment, we consider a single snapshot (hence $t$ is fixed). Now it is possible to rewrite (5) as

$$f_m = \sum_{i=1}^{n} a_i \Psi_i^m, \quad m = 0, \ldots, M-1$$

(7)

which can be solved for the $a_i$ and $\Psi_i$ under the assumption that the $\Psi_i$ are mutually distinct. It was shown in 1795 already by de Prony, that it is possible to solve the exponential analysis problem (7) from 2n samples if the sparsity $n$ is given [25]. In its modern version, the method first recovers the values of the $\Psi_i$ as generalized eigenvalues, and subsequently solves a Vandermonde structured linear system to retrieve the coefficients $a_i$, as recalled in Section III-A. The $\psi_i$ are then unambiguously computed as $\log(\Psi_i)/d$, since the distance $d$ between the antennas is smaller than half the wavelength of the signal, or equivalently $|\Im(\psi_i d)| < \pi$, where $\Im(\cdot)$ and $|\cdot|$ denote respectively the imaginary part and the absolute value of a complex number. Finally, the direction $\phi_i$ is found from the computed $\psi_i$, as $\phi_i = \arccos(-\frac{1}{\epsilon} \psi_i/c_0)$. In general, the sparsity $n$, being the number of received signals, is also an unknown. The retrieval of $n$ is based on the singular value decomposition of some Hankel matrices, or density based cluster analysis, as discussed in Section III-B.

### A. Generalized Eigenvalue Approach

With the samples $f_m$, we fill the Hankel matrices

$$H_n^{(r)} := \begin{bmatrix} f_r & f_{r+1} & \ldots & f_{r+n-1} \\ f_{r+1} & f_{r+2} & \ldots & f_{r+n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{r+n-1} & f_{r+n} & \ldots & f_{r+2n-2} \end{bmatrix}$$

(8)

where $r$ is an integer in the interval $[0, M-2n+1]$. From the definition of these Hankel matrices and the samples $f_m$, we immediately find the factorization [6]

$$H_n^{(r)} = V_n D_n D_n^T V_n^T$$

(9)

where $V_n = (\Psi_i^{1})_{i=1}^{n}$ is the Vandermonde matrix built with the distinct values $\Psi_i$, $i = 1, \ldots, n$, and both matrices $D_n = \text{diag}(\alpha_1, \ldots, \alpha_n)$, $D_n^T = \text{diag}(\Psi_1, \ldots, \Psi_n)$ are diagonal. Using the factorization (9), we find

$$H_n^{(1)} - \lambda H_n^{(0)} = V_n D_n (D_n - \lambda I_n) V_n^T.$$ 

Hence, the $\Psi_i$ are the generalized eigenvalues of the generalized eigenvalue problem

$$H_n^{(1)} v_i = \lambda_i H_n^{(0)} v_i.$$ 

(10)

The coefficients $a_i = S_i(t)$ are subsequently computed from the Vandermonde structured linear system

$$\begin{bmatrix} 1 & \ldots & 1 \\ \Psi_1 & \ldots & \Psi_n \\ \vdots & \vdots & \vdots \\ \Psi_1^{M-1} & \ldots & \Psi_n^{M-1} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{M-1} \end{bmatrix}.$$ 

(11)

Note that in the noise-free case, only $n$ of these $M$ equations are linearly independent. In the presence of noise, the linear system (11) is solved in the least squares sense for the $a_i$. We can also solve the generalized eigenvalue problem (10) in a least squares sense [21], where the Hankel matrices defined by (8) are enlarged to dimension $(M-n) \times n$ to utilize all the available samples $f_0, \ldots, f_{M-1}$. In the sequel, we denote these rectangular Hankel matrices by $H_n^{(r)}$. Note that the Hankel matrices $H_n^{(r)}$ are equal to the square Hankel matrices $H_n^{(r)}$.

### B. Determining the Sparsity $n$

Until now, we have assumed that the sparsity, or equivalently the number of received signals, $n$ is known. However, in practice this is not always the case. One possibility is to compute the numerical rank of Hankel matrices with increasing dimensions, since it is known that [26, 27, p. 603] 1) det $H_n^{(0)} = 0$ only accidentally for $N < n$; 2) det $H_n^{(0)} \neq 0$; and 3) det $H_n^{(0)} = 0$ for $N > n$.

However, this method is not always reliable. In particular in the presence of noise, the method requires large values for $N$, meaning a large number of samples or antenna elements to estimate the sparsity $n$ correctly [28].

Another more reliable approach is based on the connection with Padé approximation theory [29]–[31]. Let us use the samples $f_m$ to construct a formal power series

$$R(z) = \sum_{m=0}^{\infty} f_m z^m.$$ 

(12)

Using the definition of the samples $f_m$, we find

$$R(z) = \sum_{m=0}^{\infty} \left( \sum_{i=1}^{n} \alpha_i \Psi_i^m \right) z^m = \sum_{i=1}^{n} \frac{\alpha_i}{1 - \Psi_i z}.$$ 

Hence, the formal power series $R(z)$ is that of a rational function of degree $n-1$ in the numerator and degree $n$ in the denominator. The consistency property of Padé approximation theory guarantees that in the noise-free case, this rational function is recovered exactly by its $[n-1, n]_R$ Padé approximant of degree $n-1$ in the numerator and $n$ in the denominator for the formal power series $R(z)$ given by (12). But how to proceed in the noisy case?

Since $R(z)$ represents a rational function, the theorem of Nuttall-Pommerenke [32], [33] and the theory of Froissart doublets [34]–[38] may help us to determine the sparsity $n$ [28]. When adding a white Gaussian noise term $\epsilon_m$ to the output $f_m$ of the ULA, the power series $R(z)$ becomes

$$R(z) + \epsilon(z) = \sum_{m=0}^{\infty} (f_m + \epsilon_m) z^m.$$ 

(13)

The theorem of Nuttall-Pommerenke states that if the noisy power series $R(z) + \epsilon(z)$ is analytic throughout the complex plane except for a countable number of poles [32] and essential singularities [33], then the paradigmatic sequence of Padé approximants $[[N-1, N], \epsilon_+: N \in \mathbb{N}]$ converges to $R(z) + \epsilon(z)$ in measure on compact sets. This means that the measure of the set in the complex plane where the
Padé approximants \([N - 1, N]_{R + \epsilon}\) do not converge, goes to zero as \(N\) tends to infinity. Note that this theorem does not say anything about pointwise or uniform convergence of the sequence, since the pointwise convergence is disrupted by \(N - n\) unwanted pole-zero combinations of the Padé approximant \([N - 1, N]_{R + \epsilon}\): for every additional pole introduced by increasing \(N\) beyond the sparsity \(n\), there is a corresponding zero introduced in the numerator, almost canceling the pole.

Increasing \(N\) thus identifies as the stable ones in successive Padé approximants. Combining the results from multiple snapshots, the true noise.

Doublets offer a way to separate the received signals from the noise. As they are randomly scattered for different Padé approximants or different realizations of the noise, hence they are distinguished by their instability [34], [35]. This means that these Froissart doublets offer a way to separate the received signals from the noise.

We therefore consider multiple snapshots to solve the problem. Combining the results from multiple snapshots, the true poles (corresponding to the directions \(\phi_i\)) are identified as the stable poles, while the noise is represented by the randomly scattered ones. Hence, cluster analysis can be used to identify the stable poles as clusters and discard the remaining poles as noise. A density-based method such as DBSCAN is a possible choice to correctly find the clusters [39]. DBSCAN requires two parameters \(\mu\) and \(\delta\). The first one is the minimal number of points \(\mu\) required in the neighborhood of a point to be a core point and form a cluster. The second one is a distance parameter \(\delta\) which defines the size of a neighborhood. The noise level influences the optimal choice for the parameters \(\mu\) and \(\delta\). Small values for the distance parameter \(\delta\) allow us to detect dense clusters and thus very stable estimates for the correct \(\Psi_i\), which is useful for low noise levels. In the case of high noise levels, a larger distance parameter \(\delta\) is recommended, since this allows to detect wider clusters. When we choose \(\mu\) equal to the number of snapshots, every pole has to be confirmed by all snapshots, resulting in an increased certainty that no incorrect angles \(\phi_i\) are detected. On the other hand, when \(\mu\) is smaller than the number of snapshots, it allows to discard erroneous results or noisy outliers. Also, in case of a low SNR, the number of snapshots can be increased while the value of \(\mu\) can be relaxed. We even found that weak signals buried in noise, so with very low SNR compared with the other terms, can be detected [31] because of their structured character as opposed to the unstructured noise. In such a situation, multiple runs with different \(\mu\), \(\delta\) combinations can be considered.

So, to retrieve \(n\) and the \(\Psi_i\), \(i = 1, \ldots, n\) we solve the generalized eigenvalue problems

\[
H_{N, M}^{(1)} p_i = \lambda_i H_{N, M}^{(0)} p_i
\]

where \(N\) is an integer in the interval \([n, M/2]\), for every snapshot. Note that we have overestimated the sparsity \(n\). Subsequently, cluster analysis is used to retrieve the stable generalized eigenvalues and discard the remaining generalized eigenvalues since they represent the noise. Finally, the \(\Psi_i\) are computed as the centers of mass of the retrieved clusters. It is recommended to take \(N\), and in addition the number of antenna elements \(M\), as large as possible. The new method is compared with standard ESPRIT on multiple snapshots in Section V.

In Section IV, we indicate how to break the Nyquist restriction \(d < \lambda/2\).

IV. SPATIAL SUB-NYQUIST SAMPLING

Let us now consider sparse antenna arrays where the distance \(d\) between the antenna elements does not satisfy the spatial Nyquist condition \(d < \lambda/2\), where \(\lambda\) is the wavelength of the received signal. Using a sparse array in contrast to a dense array does offer some advantages. Most importantly, for sparse arrays, not only the mutual coupling effect is reduced, but also the angular resolution increases for the same number of antenna elements. Similar to the theory of co-prime arrays, we combine the results of two ULAs separately to retrieve the true angles from the aliased ones. However, our proposed method avoids the search step of matching and pairing the results obtained by the two separate sparse ULAs. Instead, they are automatically linked by the new algorithm. Additionally, the proposed method also allows to validate the results and automatically retrieve the sparsity \(n\), or the number of received signals [40]. This is similar to the clustering approach discussed in Section III-B.

A. Sub-Sampled Exponential Analysis

When the spatial Nyquist bound is no longer satisfied, aliasing is introduced, which means that it is no longer possible to uniquely retrieve the \(\psi_i\) from the \(\Psi_i\), defined in Section III by (6), since it is no longer guaranteed that \(|\mathcal{D}(\psi_i, d)| < \pi\). This aliasing problem is solved by using the output generated by a second ULA [17], as explained below.

We view both sparse ULAs as a sub-ULA of a virtual dense ULA, which does satisfy the spatial Nyquist bound. Note that we do not require the full dense ULA, only the antenna elements of the two sparse ULAs. A first ULA uses the output of the antennas \(f_{\sigma, \cdot}\), where the spacing between the antenna elements is \(\sigma d\). A second similar ULA uses the antenna elements shifted over a distance of \(\rho d\) resulting in the output of the antenna elements \(f_{\sigma, \cdot + \rho}\). The scale and shift parameters \(\sigma \in \mathbb{N}_0\) and \(\rho \in \mathbb{Z}_0\) are required to be co-prime, meaning \(\gcd(\sigma, \rho) = 1\), in order to fix the aliasing. In the following, we denote the number of antennas in the first array by \(M_\sigma\) and the number of antennas in the second array by \(M_\rho\). Similar to the standard exponential analysis case, discussed in Section III, the number of antennas \(M_\sigma\) and \(M_\rho\) are restricted by the sparsity \(n\). We require that \(M_\sigma \geq 2n\) and \(M_\rho \geq n\). Note that the second sparse ULA need only consist of half the number of elements of the first sparse ULA. However, additional antenna elements can be used to solve the problem in a least squares sense. The set-up is illustrated in Fig. 2.

From the samples \(f_0, f_\sigma, \ldots, f_{(M_\sigma - 1)\sigma}\), we first compute the generalized eigenvalues of the model

\[
f(m) = \sum_{i=1}^{n} a_i \Psi_i^{m\sigma} = \sum_{i=1}^{n} a_i (\Psi_i^{\sigma})^m, \quad m = 0, \ldots, M_\sigma - 1 \quad (14)
\]
as described in Section III. Note that any 1-D implementation can be used, each with its own advantages and disadvantages, to retrieve the base terms $\Psi_i^\sigma$ associated with the samples taken at $0, \sigma d, \ldots, (M_\sigma - 1)\sigma d$. From only the generalized eigenvalues $\Psi_i^\sigma$ it is not possible to retrieve the $\Psi_i$, since several possibilities exist, namely all the values in the set
$$\left\{ \exp(\psi_i d + \frac{2\pi j}{\sigma} \ell) : \ell = 0, \ldots, \sigma - 1 \right\}. \quad (15)$$

The second ULA is required to pinpoint the one correct value from this set. The samples collected by the second ULA satisfy
$$f_{m\sigma + \rho} = \sum_{i=1}^n a_i \Psi_i^{m\sigma + \rho} = \sum_{i=1}^n (a_i \Psi_i^\sigma)^m \Psi_i^\rho, \quad m = 0, \ldots, M_\rho - 1. \quad (16)$$

We find the same generalized eigenvalues $\Psi_i^\sigma$ but different coefficients compared with the model of the sparse ULA 1. Hence, both times a Vandermonde structured linear system with the same generators, but a different number of rows, is solved for the left-hand sides in (14) and (16), to obtain both the values for the coefficients $a_i$ in (14) and $a_i \Psi_i^\sigma$ in (16).

Dividing the latter by the coefficients $a_i$ results in the values for $\Psi_i^\rho$. Given the $\Psi_i^\rho$, again a set of possible values remains for each $\Psi_i$, namely
$$\left\{ \exp(\psi_i d + \frac{2\pi j}{\rho} \ell) : \ell = 0, \ldots, \rho - 1 \right\}. \quad (17)$$

Since we choose $\sigma$ and $\rho$ co-prime, the intersection of the sets (15) and (17) only contains the desired $\Psi_i$ [17]. This is illustrated in Fig. 3. Hence, the aliasing problem is solved. Note that we already know which sets (15) and (17) correspond to each other due to the shared Vandermonde system, on the contrary to the theory of co-prime arrays, where a search step is required to match the results of both ULAs. When noise is present, both sets do not intersect exactly. In this case, we search for the two closest points in both sets. Note that while in theory $\sigma$ and $\rho$ may be chosen arbitrarily large, this is not the case in practice since noise can make it hard to point at the one correct value when a large number of points lie closer together in (15) and (17). Practically one would always design a system with as many sensors as possible. Both $\rho$ and $\sigma$ should be selected to be as small as possible for the desired angular resolution (since the resolution is ultimately proportional to $\sigma$), while the practical constraint on the total length must also be considered. Considering the noise effects, numerical experiments indicate that a smaller $\rho$ is more important than a smaller $\sigma$. Also, the $\Psi_i$ values from (10) are usually less affected by noise than the $\Psi_i^\rho$ which are obtained as solution of a Vandermonde structured linear system.

The fact that the method solves (14) and (16) per snapshot and passes the individual results to a cluster detection algorithm, makes it computationally more efficient than traditional methods performing the analysis on much larger matrices containing the measurements of all snapshots simultaneously. In addition, the independence of all the smaller systems involved allows to parallelize the computation.

B. Determining the Sparsity n

In the sub-sampled case, we also use the connection with Padé approximation theory to determine the sparsity $n$. We first solve the generalized eigenvalue problem
$$H_{N,M_\rho}^{(1)} v_i = \lambda_i H_{N,M_\rho}^{(0)} v_i, \quad n \leq N \leq \lfloor M_\sigma/2 \rfloor$$

where the Hankel matrices are filled with the output of the first sparse (not shifted) ULA, so the samples $f_{m\sigma}$. This yields $N$ generalized eigenvalues $\Psi_i^{\sigma}$, of which $n$ values correspond to the true angles and $N - n$ values correspond to the noise. Subsequently, an $M_\sigma \times N$ and an $M_\rho \times N$ Vandermonde structured linear system is solved to retrieve $N$ estimates for the $\Psi_i^\rho$. Once again, $n$ of these values correspond to the true angles, while the remaining $N - n$ model the noise. This process is repeated for all the snapshots. We collect all the results for multiple snapshots in two large sets: the set $A_\sigma$ for the first ULA and the set $A_\rho$ for the shifted ULA. Now cluster analysis identifies the $n$ true $\Psi_i^{\sigma}$ and $\Psi_i^\rho$ in the two sets $A_\sigma$ and $A_\rho$, as discussed in Section III-B.

Because of the shared Vandermonde structured linear system in both (14) and (16), the $\Psi_i^{\sigma}$ and $\Psi_i^\rho$ are matched automatically. Hence, every element in the set $A_\sigma$ is directly connected with an element in the set $A_\rho$. By combining the DBSCAN results for $A_\sigma$ and $A_\rho$, ultimately three scenarios may occur [40] as follows.

1) A cluster $C_\sigma$ is detected in the set $A_\sigma$ and its center of mass is used to estimate the generalized eigenvalue $\Psi_i^{\sigma}$. The points in the set $A_\rho$ connected to the elements in the cluster $C_\sigma$ also form a cluster $C_\rho$ in $A_\rho$ and their center of mass is used as estimate for the corresponding $\Psi_i^\rho$. From both centers of mass we recover the value for $\Psi_i$. 

![Fig. 2. Original dense array versus two sparse arrays.](image)

![Fig. 3. Intersection of the sets (15) and (17) for $\sigma = 7, \rho = 5, \omega = 1.5$ GHz, $d = 0.48\phi_i$, and the angle $\phi_i$ of the incoming signal being 105°. No noise was added.](image)
as discussed in Section IV-A, and subsequently retrieve the angle $\phi_i$. This scenario is the standard one.

2) A cluster $C_\sigma$ is detected in the set $A_\sigma$. However, the points in the set $A_\rho$ connected to the cluster $C_\sigma$ do not form a cluster, hence the cluster $C_\sigma$ is discarded and not used in the computation for the values $\Psi_i$. Note that this affects the estimate for the sparsity $n$ since a cluster is removed.

3) A cluster $C_\sigma$ is detected in the set $A_\sigma$ and its center of mass is used to estimate the generalized eigenvalue $\Psi_i^\sigma$. The points in the set $A_\rho$ connected to the cluster $C_\sigma$ not all belong to a cluster $C_\rho$, since some of them are outliers. In this case, the center of mass of only the clustered points is used as estimate for the corresponding $\Psi_i^\rho$. From both centers of mass we recover the value for $\Psi_i$, as discussed in Section IV-A, and subsequently retrieve the angle $\phi_i$.

Usually multiple DBSCAN runs are performed, starting with a high validation rate and relaxing it gradually (by decreasing $\mu$ and increasing $\delta$), until the clusters detected in $A_\sigma$ are not validated anymore by a cluster in $A_\rho$.

The computational effort required is thus the solution of $K$ (number of snapshots) Hankel matrices of size $(M_\rho-N) \times N$, and $2K$ Vandermonde structured linear systems of size $M_\rho \times N$. This process can be easily parallelized since all snapshots are uncoupled. In a last step, DBSCAN is performed on $KN$ different points to retrieve $n \leq N$ distinct clusters.

V. Examples

We illustrate the performance of our proposed method under the influence of noise and mutual coupling. In the first examples we only look at the effect of noise, hence no mutual coupling effects are considered. Subsequently, we consider some simulation examples which include full coupling under low noise levels, to isolate the effects of coupling only.

A. Effects of Noise

First, we demonstrate the behavior under influence of noise, mainly to illustrate the effect of the validation step. Therefore, we consider $n = 10$ received signals, as given in Table I. We use both standard ESPRIT and the new method with ESPRIT underlying to retrieve the ten corresponding angles for increasing levels of noise. The additive noise is expressed in terms of SNR, which is defined by $20 \log_{10}(||f||_2/||\epsilon||_2)$, where $||f||_2$ and $||\epsilon||_2$ denote the 2-norm of respectively the sample and additive noise vector.

For the standard ESPRIT approach, a ULA of 60 antennas with a distance of 0.48\lambda between the elements is considered. We also tell ESPRIT the correct number of signals, i.e., $n = 10$. We solve the DOA problem using ESPRIT on 256 snapshots. In Fig. 4 at the top, the results of all 256 snapshots are shown together. We observe that ESPRIT works well for a high SNR, however, for higher noise levels (approaching 10 dB), the standard ESPRIT approach delivers unreliable results.

At the same time, our new approach (with ESPRIT as underlying method) also uses 60 antennas in total: a first sparse ULA of 30 antennas and a shifted ULA of 30 antennas, with a scale and shift parameter of, respectively, $\sigma = 25$ and $\rho = 14$. The distance between the virtual dense array elements is also chosen as $0.48\lambda$ resulting in a total array size of more than 350\lambda. This might be an unrealistically large system for many applications, but the example serves to illustrate the efficacy of the proposed method even under such demanding conditions where both $\sigma$ and $\rho$ are large. As stated before, increasing $\sigma$ results in a more difficult root intersection (validation) problem. Since the method detects the number of signals $n$ automatically, $n$ need not be passed to the algorithm. For the clusters $C_\sigma$, we choose the DBSCAN parameters $\mu = 0.8 \times 256 = 205$ with increasing equidistant $\delta$ values, namely $\delta = 0.01, 0.0825, 0.155, 0.2275, 0.3$. For the $C_\rho$ clusters, we take $\mu = 0.6 \times 256 = 154$ with $\delta = 0.5$ because they are usually less accurate, as already pointed out.

At the bottom of Fig. 4, we find the results from 256 snapshots. For the lower noise levels, we clearly see that our method performs comparably. When the signals are perturbed with a lot of noise, we observe that the new method does not return all the angles, however, it also does not return unreliable results such as the stand-alone ESPRIT method. It detects that the signal is heavily perturbed, since the results are not

<table>
<thead>
<tr>
<th>Angle</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude</td>
<td>0.3</td>
<td>0.2</td>
<td>0.4</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Phase</td>
<td>0.9\pi</td>
<td>1.2\pi</td>
<td>0.8\pi</td>
<td>0.7\pi</td>
<td>1.1\pi</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Angle</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude</td>
<td>0.4</td>
<td>0.7</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Phase</td>
<td>0.7\pi</td>
<td>1.3\pi</td>
<td>1.2\pi</td>
<td>1.0\pi</td>
<td>1.1\pi</td>
</tr>
</tbody>
</table>

Fig. 4. Solution of the DOA problem given by Table I for increasing noise levels of both standard ESPRIT (top) and the new method (bottom).
ULAs, we require 70% validation with $\mu$ factor arrays. In our case, we have one array sub-sampled by a factor $\sigma$, while their method just combines the results of two sparse arrays, which are respectively sub-sampled by factors $\sigma_1$ and $\sigma_2$. Then, a matching step is performed to be able to link the results of both sparse arrays. This matching is based on a projection of a 2-D point on a 1-D line segment that corresponds to the entire angular domain due to the co-primeness of $\sigma_1$ and $\sigma_2$.

Similar to the first experiment, we use both methods to retrieve the $n = 6$ angles given in Table II, for increasing noise levels. For this experiment, we only consider six instead of ten signals, since the matching in [41] becomes considerably worse for larger $n$-values. For both methods, 256 snapshots are collected by two sparse ULAs of 20 elements each, where the distance between the elements of the virtual dense array is $0.48\lambda$, $\sigma = \sigma_1 = 10$, and $\rho = \sigma_2 = 3$. We also pass the number of signals $n = 6$ to the co-prime array method, while our method is able to detect this number of impeding signals automatically. For the cluster analysis, we use DBSCAN on ULA1 with $\mu = 218$ for 85% validation and $\delta = 0.01, 0.02, 0.04, 0.08, 0.16, 0.32$, while for ULA2, we require 70% validation with $\mu = 179$ and $\delta = 0.6$. The results are shown in Fig. 5, where for every noise level we plot the joint output of 100 different noise realizations to observe the effect of noise on both methods.

We see that for low noise levels, both methods perform comparably, however, for higher noise levels, the difference becomes clear. The co-prime arrays have trouble matching the results of both ULAs, hence, especially for high noise levels, we observe that mismatches are not uncommon and thus result in erroneously retrieved angles. On the other hand, our method may not validate all six angles in case of high noise levels, however, it also does not yield any erroneous results. Note that it is possible to obtain all six angles for a low SNR with the proposed method when we are less strict on the validation part. However, this can lead to angles which are slightly less accurate than the ones returned in this experiment.

Both ESPRIT and the sparse array method in [41] require solutions of matrices of the order $M \times Kn$, which can be significantly slower than the series of smaller systems required for the new method.

<table>
<thead>
<tr>
<th>Angle</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude</td>
<td>35.0</td>
<td>62.5</td>
<td>90</td>
<td>96.5</td>
<td>123.5</td>
<td>151</td>
</tr>
<tr>
<td>Phase</td>
<td>0.3</td>
<td>0.2</td>
<td>0.4</td>
<td>0.5</td>
<td>0.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>

![Fig. 5. Solution of the DOA problem given by Table II for increasing noise levels of both co-prime arrays (top) and the new proposed method (bottom).](image)

**TABLE II**

<table>
<thead>
<tr>
<th>Angle</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude</td>
<td>0.9π</td>
<td>1.2π</td>
<td>0.8π</td>
<td>0.7π</td>
<td>1.1π</td>
<td>0.7π</td>
</tr>
</tbody>
</table>

**TABLE OF THE SIX RECEIVED SIGNALS USED TO GENERATE THE RESULTS IN FIG. 5**

**B. Effects of Mutual Coupling**

To illustrate the improved resolution of the algorithm over standard dense ULAs, as well as the effects of mutual coupling on the performance thereof, a 12 element dipole array is considered. The antennas have length $\lambda/2$ and radius of $\lambda/1000$ (thin wires), are oriented along the $z$-axis of a standard Cartesian coordinate system, and are arranged linearly along the $x$-axis. Spacing is defined as in Fig. 2, with $\sigma = 11$, $\rho = 5$, and $M_\sigma = M_\rho = 6$, with the central feed points as reference positions. All elements are loaded with $50\Omega$ sources, and full-wave method of moment (MoM) simulations using FEKO [42] are used to characterize the array system. For this experiment, we fix $\text{SNR} = 30\,\text{dB}$, use 256 snapshots and investigate a range of values for $d \in [0.09\lambda, 0.48\lambda]$. Smaller values of $d$ correspond to denser arrays, while larger values result in sparse arrays with finer resolution and less mutual coupling.

The resolution of the system is evaluated by exciting the array with two monochromatic plane waves. The incoming directions of the $z$-polarized sources are both fixed at $\theta = 90^\circ$, while the first is also fixed at $\phi_1 = 90^\circ$ and the second varied in $\phi_2 \in [75^\circ, 90^\circ]$. Standard spherical coordinate system azimuth and polar angles are used for $\phi$ and $\theta$.

Three different spacing cases are investigated with $d = 0.09\lambda, 0.3\lambda, 0.48\lambda$. For all cases, 100 Monte-Carlo trial runs are performed, and the accuracy of the estimated angles is presented as the RMS value, over these trials, of the absolute difference between the estimated and the actual DOA (for each of the signals). Since $n = 2$ here, both the success rate as well as RMS accuracy of the DOA estimation algorithm is computed. The success rate is just the number of successful trials divided by the total number of trials as a percentage. If fewer than two angles are returned, the trial is counted as unsuccessful.

For comparison, results obtained by MUSIC on the same geometry are also included. To this end, the second signal is excited in quadrature with the first to remove all correlation...
between the signals. Results for the method presented here
are identical to the uncorrelated case when correlated signals
are used, while MUSIC fails in this case. To avoid any ambi-
guity issues for the sparse cases, the array manifold search
space in the algorithm is restricted to the region $\left[ \min(\phi_2) - \Delta \phi, \phi_1 + \Delta \phi \right]$, where the resolution $\Delta \phi$ is calculated using (1).

Note that, for a standard $\lambda/2$-spaced ULA with 12 elements,
the resolution is expected to be around $10.5^\circ$, and resolutions
for the other examples are indicated as vertical dotted lines on
the result plots in Figs. 6–8.

The results for the $d = 0.09\lambda$ case are shown in Fig. 6. For this configuration ($\rho = 5$), the spacing results in a
$0.45\lambda$ separation between the closest elements. The worst
case mutual coupling is around $-15.3$ dB, very similar to
the value obtained from a classical $\lambda/2$-spaced ULA. Since
the maximum separation here is only $0.54\lambda$, standard dense
array techniques can be used near broadside without
the risk of ambiguity. Clearly, the new method suggested in this
article, performs significantly better than MUSIC in terms of
resolution. Both incoming signals can be resolved reliably
from about half the Rayleigh limit.

Results for $d = 0.3\lambda$ and $d = 0.48\lambda$ are shown
in Figs. 7 and 8, respectively. Similar results as before are
obtained, where the present method not only displays better
resolution than MUSIC, but also lower error values. It must
be stressed here that MUSIC cannot really be used in these
cases, since ambiguous results are found when searching over
the full array manifold (as is normally done). The ambiguity is
artificially removed through prior knowledge of the incoming
signal directions for MUSIC, while the present method returns
unique results. Compared with the narrower spaced results,
increasing $d$ results in reduced mutual coupling and thus
improved accuracy in angle estimation for both methods. Note
again that the resolution for the $d = 0.48\lambda$ case is about a
factor 5 better than that achievable with a standard ULA, and
this value can be further increased by increasing the scale and
shift parameters $\sigma$ and $\rho$ in the array.

VI. Conclusion

A new method is presented for DOA estimation in sparse
regularly spaced array systems. The method can be used
on top of any exponential analysis method and completely
removes the dense Nyquist spacing requirement. Instead, it
only requires that the elements are arranged in two uniform
arrays with the one an endfire translated version of the
other. By enforcing a co-prime relation between the element
spacing and the translation distance, aliasing due to the spatial sparsity of the elements is removed. The larger spacing allows improved resolution for a fixed number of sensors and reduced mutual coupling between the antenna elements. Furthermore, the method provides an accurate estimate of the number of signals impinging on the system without any prior knowledge, while there is no requirement that the respective signals be uncorrelated. Distinct, fully correlated signals, such as those that occur in a multipath environment, can thus accurately be distinguished using this new method. Several examples are presented investigating both the performance in increasing noise, as well as increasing mutual coupling scenarios. In all of these, the new method performs better than standard techniques on a variety of metrics. Work is currently ongoing toward expansion of the method to the planar case, which will be reported in a subsequent article. A prototype implementation is also under development.

REFERENCES

Ferre Knaepkens received the M.Sc. degree (Hons.) in mathematics from the University of Antwerp, Antwerp, Belgium, in 2017, where he is currently pursuing the Ph.D. degree with the Department of Mathematics and Computer Science.

In view of his strong interest in numerical and applied mathematics, he subsequently started working on a research project with the Computational Mathematics (CMA) Group, University of Antwerp.

Annie Cuyt received the Doctor Scientiae degree (summa cum laude) from the University of Antwerp, Antwerp, Belgium, in 1982, with the felicitations of the jury.

She was a Research Fellow of the Alexander von Humboldt Foundation, Germany. She is currently a Full Professor with the Department of Mathematics and Computer Science, University of Antwerp. She is the author of more than 200 peer-reviewed publications in international journals and conference proceedings, the author or editor of several books, and the organizer of several larger and smaller international events. Her current research interests are in numerical approximation theory and its numerous applications in scientific computing. The latter varies from different engineering applications (telecommunication, radar imaging, EM modeling, exponential analysis), to bioinformatics, computational finance, and computer graphics. In view of her expertise, she has served on many national and international science foundation boards.

Dr. Cuyt has been a lifetime member of the Royal Flemish Academy of Belgium for the Sciences and Arts since 2014. She obtained the Habilitation in 1986, and was honoured with a Masuda Research Grant (Japan).

Wen-Shin Lee (Member, IEEE) received the Ph.D. degree in computational mathematics from North Carolina State University, Raleigh, NC, USA, in 2001.

After then, she was a Post-Doctoral Fellow with the University of Waterloo, Waterloo, ON, Canada, and INRIA Sophia Antipolis, Biot, France. In 2005, she moved to Belgium and became a member of the Computational Mathematics Research Group, University of Antwerp. She is also Lecturer with the Division of Computing Science and Mathematics, University of Stirling, Stirling, U.K. Her research background is from computer algebra and symbolic-numeric computation. Her current focus is on exponential analysis, especially its connection to sparse interpolation and applications in signal processing.

Dr. Lee is an Editor of the ACM Communications in Computer Algebra.

Dirk I. L. de Villiers (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in electrical and electronic engineering from the University of Stellenbosch, Stellenbosch, South Africa, in 2004 and 2007, respectively.

From 2008 to 2009, he was a Post-Doctoral Fellow with the University of Stellenbosch, where he was working on antenna feeds for the South African SKA Program. He is currently a Professor with Stellenbosch University, where he holds the SARChI Research Chair in antenna systems for SKA. He has spent several months as a Visiting Researcher with Antwerp University, Antwerp, Belgium, and the Chalmers University of Technology, Gothenburg, Sweden. Since 2010, he has been regularly working on contract for EMSS Antennas (Pty) Ltd., Stellenbosch, South Africa, on the design of the reflector antenna systems for the MeerKAT, SKA, and ngVLA radio telescopes. His main research interests include reflector antennas for radio astronomy and surrogate-based optimization of microwave components.

Dr. de Villiers is also a Senior Member of URSI. He served as the Chair for the South African IEEE Joint AP/MTT/EMC Chapter from 2017 to 2018.