Sparse Channel Estimation From Discrete-Time Fourier Transform Beam Measurements

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

18-02-2022 / 03-03-2022

CITATION

Kapetanovic, Dzevdan (2022): Sparse Channel Estimation From Discrete-Time Fourier Transform Beam Measurements. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.19189097.v2>

DOI

[10.36227/techrxiv.19189097.v2](https://doi.org/10.36227/techrxiv.19189097.v2)

Sparse Channel Estimation From Discrete-Time Fourier Transform Beam Measurements

Dževdan Kapetanović

Huawei Technologies Sweden AB

Lund Research Center, Sweden

Abstract

In this paper, we study channel estimation at a uniform linear array (ULA) with N antennas, where the channel at the ULA is composed of L paths with different angles of arrival (AoAs). It is assumed that Discrete-Time Fourier Transform (DTFT) beams (also known as *analog* beams) are applied at the ULA to project the incoming signal onto a single (or multiple) RF chain(s), after which the signal is sampled and measured in baseband domain. This measurement procedure arises in various communication systems, such as the receive beam sweeping phase in 5G NR, where DTFT beams are used due to their simple implementation as linear phase shifts on analog antennas. A fundamental question about this procedure is the number of DTFT measurements necessary to recover the L AoAs. Previous work on this problem showed (by applying compressed sensing theory) that $LO(\log(N/L))$ measurements are sufficient for recovering the AoAs, which grows with N . Instead, by using properties of DTFT beam projections, we are able to show that if $N \geq 2L$, then $3L$ arbitrary DTFT measurements are enough; hence, dependency on N is completely removed. The main reason for this drastic improvement is that sparsity is found in a *non-linear* mapping of unknowns to observations, while compressed sensing deals with *linear* mappings of unknowns to observations. Furthermore, if the DTFT beams are centered at harmonic frequencies (resulting in DFT beams) with period N , then $2L$ beam measurements are enough. With these results, an AoA estimation algorithm is formulated which has enormous complexity savings compared to L -dimensional AoA search such as maximum likelihood (ML) estimation. Numerical simulations demonstrate the algorithm's superior performance over a conventional algorithm.

Index Terms

Sparse channel estimation, angle of arrival estimation, direction of arrival estimation, mmWave channel, analog beams, DFT beams.

I. INTRODUCTION

Modern communication systems are utilizing high frequency bands (above 30 GHz) for enhanced data rates. Examples of these systems are IEEE 802.11ay [2] and 5G NR [1]. The main reason for enhanced data rates at these high frequencies is the availability of more bandwidth. However, the propagation channel is not favorable at these frequencies due to high pathloss. In order to obtain power boosting, the typical approach is to employ more antennas at the devices to capture more of the signal. The high frequencies, which result in small signal wavelengths, allow packing more antennas into a small area on the device by following the MIMO rule of thumb of an antenna separation that equals half the signal wavelength [4]; this antenna separation provides maximum number of degrees of freedom in the channel. On the other hand, the high pathloss at these frequencies reduces the degrees of freedom in the channel, which results in a sparse channel that is composed of a relatively small number (L) of dominant spatial directions incoming from different AoAs. This knowledge about the sparsity of the channel allows the devices to customize their hardware and training protocols in order to efficiently estimate the channel. Therefore, at high frequencies, standards such as IEEE 802.11ay and 5G NR, employ a *beam sweeping* phase with beams that have the same structure as the spatial directions to be estimated. These beams go typically under the name of *analog* beams in the communications literature [5], and can be easily implemented as linear phase shifts across the ULA. The beam sweeping phase is composed of transmissions and receptions with analog beams, where each transmit beam results in a channel composed of L paths at the receiver. Moreover, each transmit beam is repeated a number of times so that the receiver can learn the best receive beam for each transmit beam. This is done by projecting each repeated signal at its N antennas with a set of analog beams, where the size of the set depends on the number of RF chains it has access to (one RF chains means it can only use a single analog beam at a time for projection). After the beam sweeping phase, the receiver reports the index of the transmit beam that is best matched to one of its receiving beams. Simultaneously, it can use the projections of the channel that resulted from the reported beam to obtain an estimate of the L AoAs that constitute that channel. The estimate can be used for further refinement of its receiving analog beam so that it matches the channel that arises from the reported transmit beam even better; the estimation of the L AoAs is what the work here is about.

Several works have utilized the sparse channel assumption to estimate L unknown AoAs.

Examples include applications of compressed sensing [6], [7], MUSIC [8], ESPRIT [9], Matrix Pencil Method (MPM) [10], and many others. Most of these methods rely on estimating the covariance matrix of the received signals, which would require many repeated beam transmissions (and thus cause a large delay) as well as the assumption that the underlying MIMO channel is constant during the sweeps (which can be easily violated due high sensitivity to mobility at mmWave frequencies). For these reasons, methods requiring covariance matrix estimation are not always practical in mmWave communications. MPM and the methods in [6], [7], on the other hand, assume no such statistical information and allow estimation of the AoAs from a few sweeps; both of them assume a priori knowledge of L . As will be shown in Section II-C below, for a given transmit beam, applying MPM to the beam sweeping problem would require N sweeps at the receiver to estimate the L AoAs resulting from that transmit beam. Instead, the methods in [6], [7] require at least an order of $\mathcal{O}(L \log(N/L))$ sweeps at the receiver for estimating the same L AoAs. This is a significant improvement compared to MPM, especially when $N \gg L$. The works [6], [7] apply compressed sensing to deduce this lower bound on number of sweeps, which is well known to result in logarithmic dependency on the dimension of the measured signal (which is N in this case). However, a drawback of both MPM and compressed sensing methods are their dependency on N . In mmWave channels, the number of AoAs tends to *decrease* when N *increases* due to the high directivity of large antenna arrays. This gives us the opposite effect of requiring more and more measurements for fewer and fewer AoAs, clearly an undesirable effect. It is the main focus of this work to remove this dependency on N so that we are only left with a dependency on L .

This paper is structured as follows. In Section II we introduce the system model, the problem formulation and discuss the applicability of MPM. Section III frames the noiseless measurement problem under the context of sampling theory, showing that verbatim application of the Shannon-Nyquist sampling theorem to AoA estimation requires at least N samples. Taking the sampling idea further, Section IV reformulates the sampling process in terms of samples of rational functions. This approach exposes the channel sparsity that is inherent in our measurements in the domain of rational functions, from which our main theoretical results are obtained. A direct corollary of the theoretical results is an AoA estimation algorithm for noisy scenarios, which is introduced in Section V. The performance of the algorithm is investigated in Section VI. Finally, concluding remarks and outline of future work is given in Section VII.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

We assume a ULA with N antennas which receives a superposition of L signals incoming from π -normalized angles $\theta_0, \dots, \theta_{L-1}$ (i.e., each within interval $[-1, 1]$) and with complex gains $c_0, \dots, c_{L-1}$. Mathematically, this can be expressed as

$$\begin{aligned} \mathbf{y} &= \sum_{j=0}^{L-1} \mathbf{a}(\theta_j) c_j + \mathbf{n} \\ &= \mathbf{A}(\boldsymbol{\theta}) \mathbf{c} + \mathbf{n} \end{aligned} \quad (1)$$

where $\mathbf{y}$ is the received signal across the ULA, $\mathbf{a}(\theta_j) = [1 \ e^{-i\pi\theta_j} \ \dots \ e^{-i\pi(N-1)\theta_j}]^T$ is a DTFT vector, $\mathbf{n} = [n_0 \ \dots \ n_{N-1}]^T$ is an AWGN vector where each elements has power N_0 , $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_0) \ \dots \ \mathbf{a}(\theta_{L-1})]$ is a matrix with the DTFT vectors as columns and $\mathbf{c} = [c_0 \ \dots \ c_{L-1}]^T$. In (1), one can interpret the signal $y_k = \sum_{j=0}^{L-1} e^{-i\pi k\theta_j} c_j + n_k$ at antenna k as a *sample* of the received signal. This model is commonly used in mmWave communications [3] and is assumed by many classical AoA estimation techniques, such as MUSIC, ESPRIT and MPM. These methods use the samples y_k , $0 \leq k \leq N-1$ to recover θ_l, c_l , $0 \leq l \leq L-1$, where it is noted that the c_l can be estimated by solving a linear system once the AoAs are known. In contrast with those models, herein it is assumed that the receiver further projects $\mathbf{y}$ with unit norm DTFT vectors (beams) centered at angles ϕ_k , $0 \leq k \leq K-1$,

$$\begin{aligned} r_k &= \frac{1}{\sqrt{N}} \mathbf{a}^H(\phi_k) \mathbf{y} \\ &= \sum_{j=0}^{L-1} \mathbf{a}^H(\phi_k) \mathbf{a}(\theta_j) \frac{c_j}{\sqrt{N}} + w_k \end{aligned} \quad (2)$$

where $w_k \triangleq \frac{1}{\sqrt{N}} \mathbf{a}^H(\phi_k) \mathbf{n}$. As with θ_l , we assume that $\phi_k \in [-1, 1]$. Defining $g_j \triangleq \frac{c_j}{\sqrt{N}}$ and expanding $\mathbf{a}^H(\phi_k) \mathbf{a}(\theta_j)$ in (2) further we get

$$\begin{aligned} r_k &= \sum_{j=0}^{L-1} g_j \sum_{l=0}^{N-1} e^{i\pi l(\phi_k - \theta_j)} + w_k, \quad 0 \leq k \leq K-1 \\ &= s_k + w_k, \end{aligned} \quad (3)$$

where we define the noiseless samples

$$s_k \triangleq \sum_{j=0}^{L-1} g_j \sum_{l=0}^{N-1} e^{i\pi l(\phi_k - \theta_j)}, \quad 0 \leq k \leq K-1. \quad (4)$$

Creating the vectors $\mathbf{r} = [r_0 \dots r_{K-1}]^T$, $\mathbf{s} = [s_0 \dots s_{K-1}]^T$, $\mathbf{g} = [g_0 \dots g_{L-1}]^T$, $\boldsymbol{\phi} = [\phi_0 \dots \phi_{K-1}]^T$, $\mathbf{w} = [w_0 \dots w_{K-1}]^T$, equation (3) can be expressed in matrix form as

$$\begin{aligned}\mathbf{r} &= \mathbf{A}(\boldsymbol{\phi})^H \mathbf{A}(\boldsymbol{\theta}) \mathbf{g} + \mathbf{w} \\ &= \mathbf{s} + \mathbf{w}.\end{aligned}\tag{5}$$

Unlike $\mathbf{n}$ in (1), the noise vector $\mathbf{w}$ in (5) is correlated with $\Sigma_{\mathbf{w}} = \frac{N_0}{N} \mathbf{A}(\boldsymbol{\phi})^H \mathbf{A}(\boldsymbol{\phi})$ as the covariance matrix. If the beam matrix $\mathbf{A}(\boldsymbol{\phi})$ is orthogonal, for example when the DTFT beams equal DFT beams, then $\mathbf{w}$ is uncorrelated. Comparing the samples r_k in (3) with the measurements y_k in (1), the main difference is that for measurement k , $e^{-i\pi k \theta_j}$ is changed to $\sum_{l=0}^{N-1} e^{i\pi(\phi_k - \theta_j)}$ for $0 \leq j \leq L-1$. Thus, the measurement model in (3) is fundamentally different from the measurement model in (1).

B. Problem Formulation

The first task in this work is to recover the AoAs θ_l , $0 \leq l \leq L-1$, given the *noiseless* samples s_k in (4). This corresponds to solving a non-linear system of equations in the unknowns θ_l . As mentioned before, once the θ_l are known, by inserting their values into (4) the g_l are easily obtained by solving a linear system of equations. Thus, from now on, we only focus on recovery of the AoAs. The reason for interest in the noiseless case is to answer four fundamental questions:

- 1) How many antennas N are needed to uniquely recover $\theta_0, \dots, \theta_{L-1}$ from (4)?
- 2) How many measurements K are needed to uniquely recover $\theta_0, \dots, \theta_{L-1}$ from (4)?
- 3) At which DTFT angles should these measurements be taken?
- 4) From these measurements, can we reconstruct $\theta_0, \dots, \theta_{L-1}$ without extensive searching in the angular domain $[-1, 1]$ for each AoA?

Questions 1 and 2 are of fundamental importance since they address uniqueness when solving the non-linear system of equations (4) for the AoAs θ_l . If there is no unique solution for the AoAs given the samples s_k in (4), then there is inherent uncertainty in the estimation problem and even the optimal ML algorithm fails in this case (unless there are countably many solutions and we have some prior knowledge about the solution we want). One can suspect that N and K are responsible for guaranteeing the existence of a unique solution. Namely, if the array has a single antenna ($N = 1$), it is well known that even for a single path ($L = 1$) there is no way to distinguish between θ_0 and the phase in g_0 ; hence, in this case, we must have $N \geq 2$. Thus,

there seems to be a lower bound to N , that depends on L , which must be satisfied for there to exist a unique set of AoAs that produces the noiseless samples s_k in (4). Similarly, it is obvious that when L is much larger than K (i.e., few measurements relative to the number of directions in the channel), there could be many realizations of θ_l, g_l that produce the same samples s_k . The only way forward then is to take more measurements, where the number of measurements should somehow be related to L . However, due to the non-linearity of (4), it could happen that there is *never* a unique solution to the AoAs from (4). Fortunately, we will show that there is always a unique solution to (4) once K is above a certain value that only depends on L .

Question 3 is concerned with the DTFT beams used for measurement. Are any (distinct) DTFT angles allowed, or do we need to use particular angles to solve (4)? Is one choice better than other in some aspect? As we will see, it is in general possible to use any distinct set of angles, which is practically relevant since it gives the device more playroom in choosing appropriate angles for measurement based on e.g. prior AoA knowledge. Peculiarly, as will also be shown, a particular set of angles can help in further reducing the number of measurements.

Question 4 is related to complexity. Optimal reconstruction of AoAs is achieved with the ML method, which entails a search across g_l, θ_l ; this becomes of prohibitive complexity even for small L . Moreover, the likelihood function can have many local optima (even if conditions for the uniqueness of AoAs are satisfied), thus resulting in erroneous ML estimation unless starting values close to the global optimum are obtained. Hence, there is a need for a method that can quickly and reliably provide estimates close to the global optimum, after which a refined ML search can be made around the provided estimates.

C. Application of MPM

MPM deals with a sum of complex exponentials with different amplitudes and frequencies as in (1). It shows that if $N \geq 2L$, it is possible to recover the frequencies θ_l and amplitudes c_l , $0 \leq l \leq L-1$, perfectly in absence of noise. Next, we present the essence of MPM, where further details can be found in [10]. Let $\mathbf{D}(\{e^{i\pi(j-k)\theta_l}\}_{0 \leq l \leq L-1})$ denote the diagonal matrix with the elements $e^{i\pi(j-k)\theta_0}, \dots, e^{i\pi(j-k)\theta_{L-1}}$ on the main diagonal and $\mathbf{A}_{j,:}(\boldsymbol{\theta})$ the j :th row of $\mathbf{A}(\boldsymbol{\theta})$. The MPM utilizes the rotational invariance of $\mathbf{A}(\boldsymbol{\theta})\mathbf{c}$ in (1) by noting that $\mathbf{A}_{j,:}(\boldsymbol{\theta})\mathbf{c} = \mathbf{A}_{k,:}(\boldsymbol{\theta})\mathbf{D}(e^{i\pi(j-k)\theta_0}, \dots, e^{i\pi(j-k)\theta_{L-1}})\mathbf{c}$. Clearly, no such rotational invariance exists in (5) in general, which shows that MPM is in general not applicable to (5). However, when $N \geq 2L$ and we take $K = N$ measurements, $\mathbf{A}(\boldsymbol{\phi})$ in (5) becomes a $N \times N$ Vandermonde matrix [11] which is

invertible as long as the DTFT angles are different. In this case, we obtain $\hat{\mathbf{s}} = (\mathbf{A}(\phi)^H)^{-1} \mathbf{s} = \mathbf{A}(\theta)\mathbf{g}$, which is (in absence of noise) equivalent to (1) onto which the MPM is applicable. Note that in presence of noise, if the DTFT angles are close, $\mathbf{A}(\phi)$ becomes ill-conditioned so that the inversion of $\mathbf{A}(\phi)$ results in considerable noise enhancement. Nevertheless, as explained in Section I, we are looking for a method that achieves $\mathcal{O}(L)$ measurements, which excludes application of MPM to (5) since it requires at least N measurements.

III. SAMPLING INTERPRETATION

If we see the DTFT angles as samples of a continuous variable $\phi \in [-1, 1)$, then s_k in (4) are samples of the function $s(\phi)$ at the sampling points $\phi = \phi_k$, $0 \leq k \leq K - 1$, where

$$\begin{aligned} s(\phi) &= \sum_{j=0}^{L-1} g_j p(\phi - \theta_j), \\ p(\phi) &\triangleq \sum_{l=0}^{N-1} e^{i\pi l \phi}. \end{aligned} \quad (6)$$

$p(\phi)$ is the well known *beamforming function* [4], which is the beamforming gain at a ULA for a signal incoming at angle ϕ and a receiving beam centered at angle 0. In the mathematical literature, $\frac{p(\phi)}{2\pi}$ is known as a Dirichlet kernel. Obviously, $p(\phi)$ is periodic with a period of 2, thus so is $s(\phi)$ too. Hence, recovering θ_j, g_j , $0 \leq j \leq L - 1$, is equivalent to recovering $s(\phi)$ from its samples s_k , $0 \leq k \leq K - 1$. This connection makes our problem equivalent to recovering a function from its samples, where in our case the function is a linear combination of shifted Dirichlet kernels. Figure 1 shows an example of how this function can look for a specific case.

The Fourier transforms $S(f), P(f)$ of $s(\phi), p(\phi)$, respectively, are

$$\begin{aligned} S(f) &= P(f) \sum_{j=0}^{L-1} g_j e^{-i2\pi\theta_j f} \\ P(f) &= \sum_{j=0}^{N-1} \delta\left(f - \frac{j}{2}\right) \end{aligned} \quad (7)$$

where $\delta(x)$ is the Dirac delta function. Clearly, $P(f)$ is a Dirac train with Dirac delta functions located at frequencies $k/2$, $0 \leq k \leq N - 1$, and thus it has a bandwidth of $(N - 1)/2$. From (7), it follows that $S(f)$ also has a bandwidth of $(N - 1)/2$. According to the Shannon-Nyquist sampling theorem, a sampling rate T_s above $2/(N - 1)$ is necessary in order to avoid aliasing of $S(f)$. Because $s(\phi)$ is periodic with period of 2, this amounts to taking more than $2/T_s = N - 1$

Fig. 1. Answering questions 2 and 3 in Section II-B is equivalent to answering how many samples of $s(\phi)$ are required, as well as where the corresponding sampling points should be located in $[-1, 1]$, for perfect recovery of $s(\phi)$. The figure shows $s(\phi)$ for a specific case where $N = 8$, $L = 3$, $s(\phi) = 1.7p(\phi) + p(\phi - 0.3) + 2.3p(\phi + 0.6)$.

samples s_k within the period, i.e., at least N samples. The DFT of these samples produces the values $S(k/2)$, $0 \leq k \leq N - 1$, which equal a linear combination of exponentials as in (1) onto which MPM can be applied for recovering the AoAs. However, as seen, we need $\mathcal{O}(N)$ samples to recover $S(f)$ perfectly with the Shannon-Nyquist sampling theorem. This is similar to the conclusion in Section II-C, where N samples are needed to invert $A(\phi)$, except for a subtle, but important, difference. Namely, the application of the Shannon-Nyquist theorem to $s(\phi)$ requires uniform samples, meaning that in this case, the sampling points are $\phi_0 + 2k/N$, $0 \leq k \leq (N - 1)$, for some starting value ϕ_0 . On the other hand, the method in Section II-C allows using *arbitrary* sampling points in $[-1, 1]$ because the DTFT beams that constitute $A(\phi)$, whose angles ϕ_k are the sampling points, can be arbitrary. Thus, this is a stronger result than direct application of the Shannon-Nyquist sampling theorem to $s(\phi)$, since the different sampling points do not need to be uniform. However, both methods still require N samples for recovering the AoAs. Lowering the number of samples below N will inevitably result in aliasing of $S(f)$ according to the Shannon-Nyquist theorem, corresponding to *sub-Nyquist sampling*. Next, we will make a connection between (6) and *rational functions*, which will allow us to recover the AoAs with sub-Nyquist sampling.

IV. CONNECTION WITH RATIONAL FUNCTIONS

If we make the substitutions $z = e^{i\pi\phi}$, $\phi = \text{Im}\{\log(z)/\pi\}$, where $\text{Im}\{\cdot\}$ denotes the imaginary part of a complex quantity, we can see the samples s_k in (4) as samples of the $N - 1$ degree polynomial $s_z(z)$

$$s_z(z) = s(\text{Im}\{\log(z)/\pi\}) = \sum_{j=0}^{N-1} z^j \sum_{k=0}^{L-1} g_k e^{-i\pi j\theta_k} \quad (8)$$

at the sampling points $z = z_k = e^{i\pi\phi_k}$, $0 \leq k \leq K - 1$. It is seen that the N polynomial coefficients (also referred to as *parameters*) that define $s_z(z)$ uniquely, equal the exponential sum observed at the different antennas in (1) for which MPM can be applied. To the author's knowledge, $N \geq 2L$ is a *sufficient* condition for recovery with MPM but not *necessary*. Hence, the following result can then be stated

Theorem 1. *If $N \geq 2L$, there is a unique set of g_l, θ_l , $0 \leq l \leq L - 1$, for a given $s_z(z)$.*

Later in this work, for the method developed herein, we will prove that $N \geq 2L$ is a necessary and sufficient condition for unique recovery.

It is enough to obtain $2L$ coefficients from (8) for the application of MPM. However, from the theory of polynomial interpolation, it follows that the N polynomial coefficients of $s_z(z)$ are obtained uniquely by evaluating $s_z(z)$ at N different values of z ; this is analogous to the Shannon-Nyquist sampling theorem, expressed here instead as polynomial interpolation. Hence, as before, N DTFT measurements are again required for unique estimation of the polynomial coefficients. To obtain $\mathcal{O}(L)$ samples, we will next provide an equivalent representation of $s_z(z)$ that has $\mathcal{O}(L)$ parameters defining it.

A. Representation as a rational function

If we apply the geometric series identity $\sum_{j=0}^{N-1} z^j e^{-i\pi j\theta_k} = \frac{(ze^{-i\pi\theta_k})^N - 1}{ze^{-i\pi\theta_k} - 1}$ to (8), it readily follows that $s_z(z)$ can be expressed as a *rational function* in z

$$s_z(z) = \frac{u^{\text{rat}}(z)}{v^{\text{rat}}(z)} \quad (9)$$

where

$$u^{\text{rat}}(z) = \sum_{j=0}^{L-1} (z^N e^{-i\pi N \theta_j} - 1) q_j(z) g_j \quad (10)$$

$$v^{\text{rat}}(z) = \prod_{k=0}^{L-1} (z e^{-i\pi \theta_k} - 1) \quad (11)$$

$$q_j(z) \triangleq \prod_{k=0, k \neq j}^{L-1} (z e^{-i\pi \theta_k} - 1) \quad (12)$$

The polynomial $v^{\text{rat}}(z)$ is of degree L , while each polynomial $q_j(z)$ is of degree $L - 1$ which makes $u^{\text{rat}}(z)$ a polynomial of degree $N + L - 1$. $v^{\text{rat}}(z)$ is, up to a constant, uniquely defined by its roots which equal $e^{i\pi \theta_l}$, $0 \leq l \leq L - 1$; hence, there is a one-to-one correspondence between $v^{\text{rat}}(z)$ and the AoAs. Although $u^{\text{rat}}(z)$ is of degree $N + L - 1$, it can be seen that its coefficients that correspond to the $N - L$ exponents $z^L, z^{L+1}, \dots, z^{N-1}$ are zero. Thus, effectively, $u^{\text{rat}}(z)$ has only $2L$ non-zero coefficients. Since $u^{\text{rat}}(z) = 0$ and $v^{\text{rat}}(z) = 0$ when z equals $e^{i\pi \theta_l}$, $0 \leq l \leq L - 1$, they have these and *only* these L roots in common because $v^{\text{rat}}(z)$ is of degree L . Clearly, we now have two different expressions of $s_z(z)$: the polynomial expression in (8) and the rational function expression in (9). The AoAs can be obtained from both of them; in the former case, by applying MPM to $2L$ polynomial coefficients and in the latter case by finding roots of $v^{\text{rat}}(z)$. An important remark here is that the polynomial expression can also be expressed as a rational function, where the numerator is the polynomial in (8) and the denominator is 1; still, in order to avoid confusion, we will refer to (8) as the polynomial expression of $s_z(z)$.

The polynomial expression is determined by N parameters, while the rational function expression is determined by $3L + 1$ parameters ($2L$ from $u^{\text{rat}}(z)$ and $L + 1$ from $v^{\text{rat}}(z)$). Hence, when $N > 3L + 1$, $s_z(z)$ has a *sparser* representation as a rational function than the polynomial expression. If we are able to recover these $3L + 1$ parameters with $\mathcal{O}(L)$ measurements, our goal is achieved.

B. Rational function recovery

Inserting the sampling points z_k into (9), we get a system of equations in the *coefficients* of $u^{\text{rat}}(z), v^{\text{rat}}(z)$

$$\frac{u^{\text{rat}}(z_k)}{v^{\text{rat}}(z_k)} = s_z(z_k) = s_k. \quad (13)$$

From (13), it follows that these coefficients *necessarily* satisfy the linear equations

$$v(z_k)s_k - u(z_k) = 0 \quad (14)$$

where $u(z)$ and $v(z)$ are polynomials of degrees $N + L - 1$ and L , respectively (they do not need to be of the form as in (10), (11)). Denote by $\mathbf{u} = [u_0 \dots u_{N+L-1}]^T$, $\mathbf{v} = [v_0 \ v_1 \dots v_L]^T$ the coefficients of $u(z), v(z)$, respectively, where the sub-index of the coefficient equals the polynomial exponent it multiplies. Further, let $\mathbf{z} = [z_0 \ z_1 \dots z_{K-1}]^T$. The linear system of coefficients in (14) can be expressed in matrix form as

$$\mathbf{A}(\mathbf{z})\mathbf{x} = \mathbf{0}$$

where

$$\mathbf{A}(\mathbf{z}) = \begin{bmatrix} s_0 z_0^L & \dots & s_0 & -z_0^{N+L-1} & \dots & -1 \\ & & \vdots & & & \vdots \\ s_{K-1} z_{K-1}^L & \dots & s_{K-1} & -z_{K-1}^{N+L-1} & \dots & -1 \end{bmatrix}$$

$$\mathbf{x} = [\mathbf{v}^T \ \mathbf{u}^T]^T. \quad (15)$$

Matrix $\mathbf{A}(\mathbf{z})$ is of dimensions $K \times (N + 2L + 1)$. Invoking [12, Th. 2.2.1.4], it follows that if $K \geq N + 2L$, then there is a *unique* rational function (up to a scalar multiple of numerator and denominator) that solves the equations in (13); hence, in our case, this means that $s_z(z)$ in (9) is the only solution to (13). However, as we saw above, $s_z(z)$ can be *expressed* in different ways as a rational function: the polynomial expression in (8) and the rational function expression in (9). All different expressions of $s_z(z)$ as a rational function necessarily satisfy the linear system in (15). As an example, the polynomial expression of $s_z(z)$ in (8) is obtained by letting

$$u_j = (-1)^L \sum_{k=0}^{L-1} g_k e^{-i\pi j \theta_k}, \quad 0 \leq j \leq N-1$$

$$u_j = 0, \quad N \leq j \leq N+L-1$$

$$v_0 = (-1)^L$$

$$v_j = 0, \quad 1 \leq j \leq L.$$

The resulting $\mathbf{x}$ vector in (15), which we denote by $\mathbf{x}^{\text{poly}}$ in this case, solves (15). These choices correspond to the polynomials $u^{\text{poly}}(z) = \sum_{j=0}^{N-1} z^j \sum_{k=0}^{L-1} g_k e^{-i\pi j \theta_k}$ and $v^{\text{poly}}(z) = (-1)^{L+1}$. On the other hand, if we let $\mathbf{u}$ and $\mathbf{v}$ equal the coefficients of $u^{\text{rat}}(z)$ and $v^{\text{rat}}(z)$ in (9) (where

$$\mathbf{B}(\mathbf{z}) = \begin{bmatrix} s_0 z_0^L & s_0 z_0^{L-1} & \dots & s_0 & -z_0^{N+L-1} & -z_0^{N+L-2} & \dots & -z_0^N & -z_0^{L-1} & -z_0^{L-2} & \dots & -1 \\ & \vdots & & & & \vdots & & & & & \vdots & \\ s_{K-1} z_{K-1}^L & s_{K-1} z_{K-1}^{L-1} & \dots & s_{K-1} & -z_{K-1}^{N+L-1} & -z_{K-1}^{N+L-2} & \dots & -z_{K-1}^N & -z_{K-1}^{L-1} & -z_{K-1}^{L-2} & \dots & -1 \end{bmatrix} \quad (16)$$

again $v_0^{\text{rat}} = (-1)^{L+1}$), and denote the resulting $\mathbf{u}, \mathbf{v}, \mathbf{x}$ as $\mathbf{u}^{\text{rat}}, \mathbf{v}^{\text{rat}}, \mathbf{x}^{\text{rat}}$, respectively, then $\mathbf{x}^{\text{rat}}$ also solves (15). Since $\mathbf{x}^{\text{poly}} \neq c\mathbf{x}^{\text{rat}}$ for any complex number c , and they are both non-zero, it implies that $\mathbf{A}(\mathbf{z})$ in (15) always has a null space of at least two dimensions; thus, there is an infinite number of solutions to (15). What [12, Th. 2.2.1.4] states is that all these solutions to (15) represent the *same* unique function $s_z(z)$ when $K \geq N + 2L$. Obtaining these solutions enables finding the solution that corresponds to $\mathbf{x}^{\text{rat}}$, from which we get $\mathbf{v}^{\text{rat}}$ and its associated polynomial $v^{\text{rat}}(z)$ that encodes the AoAs into its roots.

The question is whether we can obtain the $3L + 1$ elements of $\mathbf{x}^{\text{rat}}$ after $\mathcal{O}(L)$ measurements. Direct application of [12, Th. 2.2.1.4] results in $K \geq N + 2L$ measurements for obtaining $s_z(z)$ (and thus $\mathbf{x}^{\text{rat}}$), which is even more measurements than the N measurements required for finding the polynomial expression in (8). The reason for this increase is that [12, Th. 2.2.1.4] deals with recovering a rational function where the only prior knowledge about it are the degrees of the two polynomials defining it. In our case, we have additional knowledge of these polynomials, such as the fact that some exponents in $u^{\text{rat}}(z)$ are absent as well as some knowledge about the roots. Utilizing this extra knowledge, we will show that it is possible to obtain unique recovery of $s_z(z)$ with only $\mathcal{O}(L)$ measurements.

C. Recovering rational expression of $s_z(z)$

Start by removing the columns from $\mathbf{A}$ that correspond to elements in $\mathbf{x}^{\text{rat}}$ that equal to 0; this results in the $K \times (3L + 1)$ matrix $\mathbf{B}(\mathbf{z})$ in (16). Hence, $\mathbf{A}(\mathbf{z})\mathbf{x} = \mathbf{0}$ in (15) becomes equivalent to $\mathbf{B}(\mathbf{z})\mathbf{t} = \mathbf{0}$, with $\mathbf{t} = [\mathbf{v}^T \ u_{0:L-1}^T \ u_{N:N+L-1}^T]^T$, where we let $u_{j:k}$ denote the vector from $\mathbf{u}$ obtained from the elements in $\mathbf{u}$ on positions j to k ; vector $\mathbf{x}^{\text{rat}}$ now corresponds to $\mathbf{t}^{\text{rat}}$ where the 0s in $\mathbf{x}^{\text{rat}}$ are removed. Clearly, $\mathbf{B}(\mathbf{z})\mathbf{t}^{\text{rat}} = \mathbf{0}$ for any $\mathbf{z}$. Note that if $\mathbf{t}$ is a solution to $\mathbf{B}(\mathbf{z})\mathbf{t} = \mathbf{0}$, then all scalar multiples $c\mathbf{t}$ of $\mathbf{t}$ are obviously also a solution to $\mathbf{B}(\mathbf{z})\mathbf{t} = \mathbf{0}$.

Solving $\mathbf{B}(z)\mathbf{t} = \mathbf{0}$ gives all polynomials $u(z), v(z)$ such that $s_k = s_z(z_k) = \frac{u(z_k)}{v(z_k)}$, $0 \leq k \leq K-1$, where $u(z)$ has degree $N+L-1$ and lacks the exponents $z^L, \dots, z^N$ while $v(z)$ has degree L . These polynomials are unique, up to a constant c , if and only if $\mathbf{B}(z)$ has a one-dimensional null space which therefore implies that $u^{\text{rat}}(z) = cu(z)$ and $v^{\text{rat}}(z) = cv(z)$ for all z . Hence, in this case, there is a unique function $s_z(z)$ that produces the observed samples s_k , $0 \leq k \leq K-1$ and it corresponds to *unique* set of AoAs, namely the ones contained in $v(z)$. Next, we identify a necessary condition on N such that this uniqueness holds.

Theorem 2. *If $N < 2L$, then for any given $s_z(z)$ of degree $N-1$, there are infinitely many polynomials $u(z)$ of degree $N+L-1$ without the exponents $z^L, z^{L+1}, \dots, z^{N-1}$, and infinitely many polynomials $v(z)$ of degree L with different roots, satisfying $s_z(z) = \frac{u(z)}{v(z)}$.*

Proof. Assume $s_z(z) = \frac{u(z)}{v(z)}$ for a given $s_z(z)$ and some $u(z), v(z)$. The equality is equivalent to $v(z)s_z(z) - u(z) = 0$ so that $s_z(z)v(z) = u(z)$. Recall from (8) that $s_z(z)$ is a polynomial of degree $N-1$. Thus, because $u(z)$ does not have the $N-L$ exponents $z^L, z^{L+1}, \dots, z^{N-1}$, $s_z(z)v(z) = u(z)$ is a system of $N-L$ linear equations in the coefficients of $v(z)$. The assumption $N < 2L$ gives at most $L-1$ such equations, while there are $L+1$ coefficients in $v(z)$ since it has degree L . Hence, the coefficients of $v(z)$ are inside a null space of at least two dimensions, which corresponds to infinitely many $v(z)$ with different roots. For each such $v(z)$, a $u(z)$ is obtained for which $s_z(z) = \frac{u(z)}{v(z)}$ holds. $\square$

Theorem 2 shows that $N \geq 2L$ is necessary for $\mathbf{B}(z)$ to have a one-dimensional null space. From Theorem 1 we already know that $N \geq 2L$ is sufficient for the existence of a single set of AoAs that gives rise to the observed samples, but Theorem 2 also shows that this is a necessary condition for $s_z(z)$ to have a unique representation as a rational function of the form in (9). An important remark here is that Theorem 2 does not imply that $N \geq 2L$ is necessary for there to exist a single set of AoAs that produces the observed samples. Namely, even though $\mathbf{B}(z)$ has a null space of at least two dimensions, it could still be the case that there is a unique $v(z)$ in that null space that has all its roots on the unit circle. However, by making sure that $\mathbf{B}(z)$ has a one-dimensional null space, uniqueness of AoAs and $s_z(z)$ is directly implied and we avoid these uncertainties.

Although $N \geq 2L$ is necessary for a one-dimensional null space of $\mathbf{B}(z)$, and thus a unique representation of $s_z(z)$ as a rational function, it is unclear whether it is sufficient. In other words,

given $N \geq 2L$, can we always find sampling points z_k , $0 \leq k \leq K-1$, so that $\mathbf{B}(\mathbf{z})$ has a one-dimensional null space? This is answered by the next theorem.

Theorem 3. Assume $N \geq 2L$ and $K \geq 3L$. For $0 \leq k \leq 2L-1$, choose z_k such that $z_k^N = e^{i\gamma}$ for some real number γ , while for $2L \leq k \leq K-1$ choose z_k as any arbitrary sampling points. Then, with probability 1, all solutions to $\mathbf{B}(\mathbf{z})\mathbf{t} = 0$ satisfy $\mathbf{t} = a\mathbf{t}^{\text{rat}}$, where a is any complex number.

Proof. The k :th equation in $\mathbf{B}(\mathbf{z})\mathbf{t} = 0$ equals

$$s_z(z_k) \sum_{j=0}^L v_j z_k^j - \sum_{j=0}^{L-1} u_j z_k^j - \sum_{j=N}^{N+L-1} u_j z_k^j = 0. \quad (17)$$

Since $s_z(z_k) = \frac{u^{\text{rat}}(z_k)}{v^{\text{rat}}(z_k)}$, multiplying both sides of (17) by $v^{\text{rat}}(z_k)$ gives

$$\begin{aligned} u^{\text{rat}}(z_k) \sum_{j=0}^L v_j z_k^j - v^{\text{rat}}(z_k) \sum_{j=0}^{L-1} u_j z_k^j \\ - v^{\text{rat}}(z_k) \sum_{j=N}^{N+L-1} u_j z_k^j = 0. \end{aligned} \quad (18)$$

Multiplication with $v^{\text{rat}}(z_k)$ does not result in $0 = 0$ with probability 1 since z_k , $0 \leq k \leq K-1$, do not equal any of $e^{i\pi\theta_0}, \dots, e^{i\pi\theta_{L-1}}$ with probability 1; this is obvious because $e^{i\pi\theta_0}, \dots, e^{i\pi\theta_{L-1}}$ are random complex numbers on the unit circle. Equation (18) can be interpreted as the samples $p(z_k)$ of the polynomial

$$p(z) = u^{\text{rat}}(z)v(z) - u(z)v^{\text{rat}}(z) \quad (19)$$

being 0, where

$$\begin{aligned} u(z) &= \sum_{j=0}^{L-1} u_j z^j + \sum_{j=N}^{N+L-1} u_j z^j \\ v(z) &= \sum_{j=0}^L v_j z^j. \end{aligned} \quad (20)$$

To prove the theorem, it suffices to show that $v(z) = cv^{\text{rat}}(z)$ and $u(z) = cu^{\text{rat}}(z)$ for some non-zero complex number c , which implies that $\mathbf{t} = c\mathbf{t}^{\text{rat}}$. If this holds for one such number c , then it holds for all possible complex numbers a since $\mathbf{B}(\mathbf{z})a\mathbf{t}^{\text{rat}} = 0$.

Recall that $u^{\text{rat}}(z)$ has the same exponents as $u(z)$. Let

$$\begin{aligned}
 u_{L-1}(z) &\triangleq \sum_{j=0}^{L-1} u_j z^j \\
 z^N \tilde{u}_{L-1}(z) &\triangleq \sum_{j=N}^{N+L-1} u_j z^j \\
 u_{L-1}^{\text{rat}}(z) &\triangleq \sum_{j=0}^{L-1} u_j^{\text{rat}} z^j \\
 z^N \tilde{u}_{L-1}^{\text{rat}}(z) &\triangleq \sum_{j=N}^{N+L-1} u_j^{\text{rat}} z^j.
 \end{aligned} \tag{21}$$

The definition of $u^{\text{rat}}(z)$ in (10) and (12) implies that

$$u_{L-1}^{\text{rat}}(z) = - \sum_{j=0}^{L-1} q_j(z) g_j \tag{22}$$

$$\tilde{u}_{L-1}^{\text{rat}}(z) = \sum_{j=0}^{L-1} e^{-i\pi N \theta_j} q_j(z) g_j. \tag{23}$$

The definitions in (21) allow us to express $p(z)$ as

$$\begin{aligned}
 p(z) &= (u_{L-1}^{\text{rat}}(z) + z^N \tilde{u}_{L-1}^{\text{rat}}(z)) v(z) \\
 &\quad - (u_{L-1}(z) + z^N \tilde{u}_{L-1}(z)) v^{\text{rat}}(z).
 \end{aligned} \tag{24}$$

Because the sampling points z_k , $0 \leq k \leq 2L-1$, satisfy $z_k^N = e^{i\gamma}$ (for which $N \geq 2L$ is necessary), $p(z_k)$ equals a polynomial $q(z_k)$ in z_k of degree $2L-1$ at these sampling points. It follows from (24) that $p(z_k) = 0$ is equivalent to

$$\begin{aligned}
 q(z_k) &= (u_{L-1}^{\text{rat}}(z_k) + e^{i\gamma} \tilde{u}_{L-1}^{\text{rat}}(z_k)) v(z_k) \\
 &\quad - (u_{L-1}(z_k) + e^{i\gamma} \tilde{u}_{L-1}(z_k)) v^{\text{rat}}(z_k) = 0, \quad 0 \leq k \leq 2L-1.
 \end{aligned} \tag{25}$$

Since the highest exponent of z_k in $q(z_k)$ is $2L-1$, it follows that the equality in (25) holds for all z

$$\begin{aligned}
 q(z) &= (u_{L-1}^{\text{rat}}(z) + e^{i\gamma} \tilde{u}_{L-1}^{\text{rat}}(z)) v(z) \\
 &\quad - (u_{L-1}(z) + e^{i\gamma} \tilde{u}_{L-1}(z)) v^{\text{rat}}(z) = 0, \quad \forall z.
 \end{aligned} \tag{26}$$

Next, observe that the roots $e^{i\pi \theta_l}$, $0 \leq l \leq L-1$, of $v^{\text{rat}}(z)$ are not roots of $u_{L-1}^{\text{rat}}(z) + e^{i\gamma} \tilde{u}_{L-1}^{\text{rat}}(z)$ in (26). Namely, direct insertion of $e^{i\pi \theta_l}$ gives from (22) that $u_{L-1}^{\text{rat}}(e^{i\pi \theta_l}) + e^{i\gamma} \tilde{u}_{L-1}^{\text{rat}}(e^{i\pi \theta_l}) =$

$g_l q_l(e^{i\pi\theta_l})(e^{i\gamma}e^{-i\pi N\theta_l} - 1) \neq 0$ with probability 1 since $g_l \neq 0$, $u_l(e^{i\pi\theta_l}) \neq 0$ and $e^{i\gamma}e^{-i\pi N\theta_l} \neq 1$ with probability 1. Thus, it must hold that $e^{i\pi\theta_l}$, $0 \leq l \leq L-1$ are roots of $v(z)$, and because $v(z)$ and $v^{\text{rat}}(z)$ are both of degree L , we have $v(z) = cv^{\text{rat}}(z)$ for some complex number c . To finish the proof, it now remains to show that $u(z) = cu^{\text{rat}}(z)$ for the *same* c .

Inserting $v(z) = cv^{\text{rat}}(z)$ into (26) gives

$$e^{i\gamma}(c\tilde{u}_{L-1}^{\text{rat}}(z) - \tilde{u}_{L-1}(z)) = -(cu_{L-1}^{\text{rat}}(z) - u_{L-1}(z)), \quad \forall z. \quad (27)$$

Using the identities $v(z) = cv^{\text{rat}}(z)$ and (27) in (24) we get

$$\begin{aligned} p(z) &= v^{\text{rat}}(z)z^N(c\tilde{u}_{L-1}^{\text{rat}}(z) - \tilde{u}_{L-1}(z)) \\ &\quad + v^{\text{rat}}(z)(cu_{L-1}^{\text{rat}}(z) - u_{L-1}(z)) \\ &= v^{\text{rat}}(z)(c\tilde{u}_{L-1}^{\text{rat}}(z) - \tilde{u}_{L-1}(z))(z^N - e^{i\gamma}), \quad \forall z. \end{aligned} \quad (28)$$

To summarize, after the $2L$ sampling points z_k , $0 \leq k \leq 2L-1$, satisfying $z_k^N = e^{i\gamma}$, we can make the conclusion that $v(z) = cv^{\text{rat}}(z)$ and that the equations $\mathbf{B}(z)\mathbf{t} = \mathbf{0}$ result in the factorization of $p(z)$ as in (28). Note that $c\tilde{u}_{L-1}^{\text{rat}}(z) - \tilde{u}_{L-1}(z)$ has degree $L-1$. Thus, by sampling $p(z)$ further with points z_k , $2L \leq k \leq K$, where $K \geq 3L-1$, $z_k^N \neq e^{i\gamma}$, and the z_k not being equal to any of $e^{i\pi\theta_0}, \dots, e^{i\pi\theta_{L-1}}$, it follows from (28) that $p(z_k) = 0$, $2L \leq k \leq K$, implies $c\tilde{u}_{L-1}^{\text{rat}}(z) = \tilde{u}_{L-1}(z)$ for all z . Inserting this equality in (27) also gives $cu_{L-1}^{\text{rat}}(z) = u_{L-1}(z)$; hence, $u(z) = cu^{\text{rat}}(z)$. This finishes the proof. $\square$

Theorem 3 shows that the sampling points z_k specified in the theorem produce a $\mathbf{B}(z)$ that always has a one-dimensional null space. Out of the K sampling points, $2L$ satisfy $z_k^N = e^{i\gamma}$ for some real value γ (any γ works), which means that they equal $z_k = e^{i(2\pi l_k + \gamma)/N}$, where l_k are distinct integers; we refer to these points as γ -shifted DFT points, where a shift of $\gamma = 0$ would be the most common choice. The other $K - 2L$ sampling points can be arbitrary sampling points.

Clearly, there is something peculiar about γ -shifted DFT points. Denote the set of $2L$ γ -shifted DFT points as $\mathbf{z}_{\text{DFT}}$. From the proof of Theorem 3, it is readily seen that after sampling with $\mathbf{z}_{\text{DFT}}$, we can already conclude that $v(z) = cv^{\text{rat}}(z)$ for all z , i.e., $\mathbf{v} = c\mathbf{v}^{\text{rat}}$. Although $\mathbf{v}$ is unique (up to a scalar) after $\mathbf{z}_{\text{DFT}}$, there are infinitely many $\mathbf{t}$ solving $\mathbf{B}(\mathbf{z}_{\text{DFT}})\mathbf{t} = \mathbf{0}$. This follows because $\mathbf{B}(\mathbf{z}_{\text{DFT}})$ is of dimension $2L \times (3L+1)$ and thus has a null space of dimension at least L . Since $\mathbf{v}$ is unique up to a scalar, it is the vector $\mathbf{u}$ that changes due to this null space, where for each such $\mathbf{u}$ we have $\mathbf{t} = [c(\mathbf{v}^{\text{rat}})^T \mathbf{u}^T]^T$ and $\mathbf{B}(\mathbf{z}_{\text{DFT}})\mathbf{t} = \mathbf{0}$. In other words,

$\mathbf{v}$ is contained in a *one*-dimensional space (spanned by $\mathbf{v}^{\text{rat}}$) while $\mathbf{u}$ is inside a space of at least dimension L . Since all the information of the AoAs is contained in $\mathbf{v}^{\text{rat}}$, we can avoid solving for $\mathbf{u}$ in favor of a simpler equation system. Note that for z_k such that $z_k^N = e^{i\gamma}$, the last L columns in $\mathbf{B}(z)$ equal the L columns to the left of them multiplied with $e^{i\gamma}$. Hence, $\mathbf{B}(z_{\text{DFT}})\mathbf{t} = \mathbf{B}_{\text{sub}}(z_{\text{DFT}})\mathbf{t}_{\text{sub}}$, where

$$\mathbf{B}_{\text{sub}}(z) = \begin{bmatrix} s_0 z_0^L & \dots & s_0 & -z_0^{L-1} & \dots & -1 \\ & & \vdots & & & \vdots \\ s_{K-1} z_{K-1}^L & \dots & s_{K-1} & -z_{K-1}^{L-1} & \dots & -1 \end{bmatrix}$$

$$\mathbf{t}_{\text{sub}} = [\mathbf{v}^T (e^{i\gamma} u_{0:L-1}^T + u_{N:N+L-1}^T)]^T. \quad (29)$$

Solving $\mathbf{B}_{\text{sub}}(z_{\text{DFT}})\mathbf{t}_{\text{sub}} = \mathbf{0}$ thus amounts to solving a $2L \times (2L + 1)$ linear system, where the solution $\mathbf{t}_{\text{sub}}$ is such that the first $L + 1$ elements of $\mathbf{t}_{\text{sub}}$ equal $c\mathbf{v}^{\text{rat}}$ for some complex number c . As a result, the complexity of finding $\mathbf{v}^{\text{rat}}$ is reduced from solving a $3L \times (3L + 1)$ linear system to solving a $2L \times (2L + 1)$ linear system. Define $\mathbf{t}_{\text{sub}}^{\text{rat}}$ by replacing $\mathbf{u}, \mathbf{v}$ in (29) with $\mathbf{u}^{\text{rat}}, \mathbf{v}^{\text{rat}}$. Hence, a corollary to Theorem 3 is

Corollary 1. Assume $N \geq 2L$, $K \geq 2L$. For $0 \leq k \leq 2L - 1$, choose z_k such that $z_k^N = e^{i\gamma}$ for some real value γ . Then, with probability 1, all solutions to $\mathbf{B}_{\text{sub}}(z_{\text{DFT}})\mathbf{t}_{\text{sub}} = \mathbf{0}$ satisfy $\mathbf{t}_{\text{sub}} = a\mathbf{t}_{\text{sub}}^{\text{rat}}$ for any complex number a .

Proof. As shown in the proof of Theorem 3, after $2L$ measurements with γ -shifted DFT points, we obtain $\mathbf{v} = c\mathbf{v}^{\text{rat}}$ for some complex number c and the equation system $\mathbf{B}_{\text{sub}}(z_{\text{DFT}})\mathbf{t}_{\text{sub}} = \mathbf{0}$ becomes equivalent to the polynomial identity $e^{i\gamma}(c\tilde{u}_{L-1}^{\text{rat}}(z) - \tilde{u}_{L-1}(z)) = -(cu_{L-1}^{\text{rat}}(z) - u_{L-1}(z))$ in (27) for all z . Rearranging the identity, we have $c(e^{i\gamma}\tilde{u}_{L-1}^{\text{rat}}(z) + u_{L-1}^{\text{rat}}(z)) = e^{i\gamma}\tilde{u}_{L-1}(z) + \tilde{u}_{L-1}(z)$, so that $e^{i\gamma}u_{0:L-1} + u_{N:N+L-1} = e^{i\gamma}u_{0:L-1}^{\text{rat}} + u_{N:N+L-1}^{\text{rat}}$. Hence, $\mathbf{t}_{\text{sub}} = c\mathbf{t}_{\text{sub}}^{\text{rat}}$ and since any scalar multiple of $\mathbf{t}_{\text{sub}}^{\text{rat}}$ is a solution to $\mathbf{B}_{\text{sub}}(z_{\text{DFT}})\mathbf{t}_{\text{sub}} = \mathbf{0}$, the proof is finished. $\square$

Direct implication of Corollary 1 is that the $2L \times (2L + 1)$ matrix $\mathbf{B}_{\text{sub}}(z_{\text{DFT}})$ has rank $2L$.

So far, $2L$ sampling points z_k were constrained to be γ -shifted DFT points. This can be relaxed with probability 1, so that all z_k are arbitrary.

Theorem 4. Assume $N \geq 2L$, $K \geq 3L$ and $\mathbf{z}$ is any sampling vector. With probability 1, all solutions to $\mathbf{B}(\mathbf{z})\mathbf{t} = \mathbf{0}$ satisfy $\mathbf{t} = c\mathbf{t}^{\text{rat}}$, where c is any complex number.

Proof. Theorem 3 shows that for $N \geq 2L$ and $K \geq 3L$, there exists a vector $\tilde{z}_{2L}$ of K sampling points $\tilde{z}_k$, $0 \leq k \leq K - 1$, such that $\mathbf{B}(\tilde{z}_{2L})$ has a one-dimensional null space. $2L$ of these sampling points are γ -shifted DFT points (thus the subindex $2L$ in $\tilde{z}_{2L}$), and the rest $K - 2L$ sampling points are arbitrary. Start by letting these $K - 2L$ arbitrary sampling points in $\tilde{z}_{2L}$ equal some $K - 2L$ points $z_{i_0}, \dots, z_{i_{K-2L-1}}$ from $\mathbf{z}$. Since $\mathbf{B}(\tilde{z}_{2L})$ is of dimension $K \times (3L + 1)$ and $\mathbf{B}(\tilde{z}_{2L})$ has a one-dimensional null space, this implies that $\mathbf{B}(\tilde{z}_{2L})$ has rank $3L$. Thus, there exists a $3L \times 3L$ submatrix in $\mathbf{B}(\tilde{z}_{2L})$ with full rank. Denote this submatrix as $\mathbf{M}(\tilde{z}_{2L})$; apparently, $\mathbf{M}(\tilde{z}_{2L})$ is obtained from $\mathbf{B}(\tilde{z}_{2L})$ by removing a single column in $\mathbf{B}(\tilde{z}_{2L})$. The proof will be completed by showing that the sampling vector $\tilde{z}_{2L}$ can be changed to any given sampling vector $\mathbf{z}$ and still satisfy $\det(\mathbf{M}(\mathbf{z})) \neq 0$ with probability 1. This shows that with probability 1, $\mathbf{B}(\mathbf{z})$ will have rank $3L$ and thus a one-dimensional null space, and since $\mathbf{B}(\mathbf{z})\mathbf{t}^{\text{rat}} = \mathbf{0}$ we know that this null space equals the line that goes through $\mathbf{t}^{\text{rat}}$.

Take a row from $\mathbf{M}(\tilde{z}_{2L})$ that equals one of the $2L$ sampling points from $\tilde{z}_{2L}$ that satisfy $\tilde{z}_k^N = e^{i\gamma}$, say row l , and replace $\tilde{z}_l$ with the variable z . Because $s_z(\tilde{z}_l)$ is a polynomial of degree $N - 1$ in $\tilde{z}_l$, $\det(\mathbf{M}(\mathbf{z}))$ becomes a polynomial $d(z)$ in z of degree $N + L - 1$. We already know from Theorem 3 that $d(\tilde{z}_l) \neq 0$ with probability 1, thus, this polynomial is not identically equal to 0. Note that the other sampling points $\tilde{z}_k$, $k \neq l$, are roots of $d(z)$ since replacing z with any one of those sampling points makes $\mathbf{M}(\mathbf{z})$ a matrix with two identical rows; thus, its determinant becomes 0. Hence, $d(z)$ can be factorized as $d(z) = \hat{d}(z) \prod_{j=0, j \neq l}^{3L-1} (z - \tilde{z}_j)$ where $\hat{d}(z)$ is a polynomial of degree $N - 2L$; thus, $d(z)$ has at most $N - 2L$ additional roots on the complex unit circle (in practice, however, these roots will most probably not be on the unit circle due to the randomness of g_l, θ_l , $0 \leq l \leq L - 1$). Thus, if z is not equal to any of these $N - 2L$ roots, then $\det(\mathbf{M}(\mathbf{z})) \neq 0$. Since the sampling points $\mathbf{z}$ are not equal to these $N - 2L$ roots with probability 1, it follows that replacing z with any of the other $2L$ sampling points $z_{i_L}, z_{i_{L+1}}, \dots, z_{i_{3L-1}}$ left in $\mathbf{z}$, say z_{i_L} without loss of generality, results in $d(z) \neq 0$ with probability 1. This replacement of $\tilde{z}_l$ with z_{i_L} results in a new set of sampling points $\tilde{z}_{2L-1}$ for which it holds that $\det(\mathbf{M}(\tilde{z}_{2L})) \neq 0$ with probability 1. Repeating this replacement process for the other sampling points $z_{i_{L+1}}, \dots, z_{i_{3L-1}}$, we conclude that $\det(\mathbf{M}(\tilde{z}_{2L})) \neq 0$ with probability 1. $\square$

The results in Corollary 1 and Theorem 4 can be used as a basis to obtain an AoA estimation algorithm under noisy scenarios.

V. AOA ESTIMATION ALGORITHM

In presence of noise, the samples s_k in (4) are replaced with r_k in (3). This transforms $\mathbf{B}(\mathbf{z})$ in (16) to $\mathbf{C}(\mathbf{z}) = \mathbf{B}(\mathbf{z}) + \mathbf{W}(\mathbf{z})$, where the first $L + 1$ columns in $\mathbf{W}(\mathbf{z})$ equal those of $\mathbf{B}(\mathbf{z})$ with s_k replaced by w_k , while the last $2L$ columns in $\mathbf{W}(\mathbf{z})$ are 0. Hence, under noise, the one-dimensional null space of $\mathbf{B}(\mathbf{z})$ equals the one dimensional space of $\mathbf{C}(\mathbf{z})$ that corresponds to its smallest singular value (which is in general non-zero due to the noise). Let $\mathbf{z}^j = \{z_k^j\}_{0 \leq k \leq K-1}$ be the vector of elements z_k^j , $0 \leq k \leq K - 1$. We can thus write

$$\mathbf{W}(\mathbf{z}) = \begin{bmatrix} \mathbf{D}(\mathbf{z}^L)\mathbf{w} & \mathbf{D}(\mathbf{z}^{L-1})\mathbf{w} & \dots & \mathbf{w} & \mathbf{0} \end{bmatrix} \quad (30)$$

Hence, the non-zero elements in $\mathbf{W}(\mathbf{z})$ are AWGN with same distribution as the elements w_k in $\mathbf{w}$. Multiplying $\mathbf{C}(\mathbf{z})$ from the left by a $K \times K$ full rank matrix $\mathbf{T}^H$, we obtain $\mathbf{T}^H \mathbf{C}(\mathbf{z}) = \mathbf{T}^H \mathbf{B}(\mathbf{z}) + \mathbf{T}^H \mathbf{W}(\mathbf{z})$, where $\tilde{\mathbf{W}}(\mathbf{z}) = \mathbf{T}^H \mathbf{W}(\mathbf{z})$ is a new noise matrix (essentially, we are performing noise filtering). Since $\mathbf{T}^H$ is of full rank, the null space of $\mathbf{T}^H \mathbf{B}(\mathbf{z})$ equals the null space of $\mathbf{B}(\mathbf{z})$. From (30), it follows that the covariance matrix between columns i and j in $\tilde{\mathbf{W}}(\mathbf{z})$, $0 \leq i, j \leq L$, is $\Sigma_{i,j} = E\{\mathbf{T}^H \mathbf{D}(\mathbf{z}^{L-i})\mathbf{w}\mathbf{w}^H \mathbf{D}(\mathbf{z}^{j-L})\mathbf{T}\} = \mathbf{T}^H \mathbf{D}(\mathbf{z}^{L-i})E\{\mathbf{w}\mathbf{w}^H\}\mathbf{D}(\mathbf{z}^{j-L})\mathbf{T} = N_0 \mathbf{T}^H \mathbf{D}(\mathbf{z}^{j-i})\mathbf{T}$. Proper choice of $\mathbf{T}$ can make the elements in $\Sigma_{i,j}$ small, and thus make the impact of noise on the null space of $\mathbf{T}^H \mathbf{B}(\mathbf{z})$ small. The design of $\mathbf{T}$ is left for future work, so in this work we will assume that $\mathbf{T} = \mathbf{I}$.

Same transformation happens to $\mathbf{B}_{\text{sub}}(\mathbf{z}_{\text{DFT}})$ in (29), where it now equals $\mathbf{C}_{\text{sub}}(\mathbf{z}_{\text{DFT}}) = \mathbf{B}_{\text{sub}}(\mathbf{z}_{\text{DFT}}) + \mathbf{W}_{\text{sub}}(\mathbf{z}_{\text{DFT}})$, with the first $L + 1$ columns in $\mathbf{W}_{\text{sub}}(\mathbf{z})$ equal to corresponding columns in $\mathbf{W}(\mathbf{z})$ and the last L columns in $\mathbf{W}_{\text{sub}}(\mathbf{z})$ are 0. Again, the impact of noise filtering with a full rank $K \times K$ matrix $\mathbf{T}$ is left for future work, where we herein assume that $\mathbf{T} = \mathbf{I}$.

So far, we have focused on the null space of $\mathbf{B}(\mathbf{z})$ or $\mathbf{B}_{\text{sub}}(\mathbf{z}_{\text{DFT}})$. If we choose $v_0 = (-1)^L$ and $c = 1$ (i.e., we are specifically looking for the solution where $v(\mathbf{z})$ has $(-1)^L$ as its coefficient), then $\mathbf{B}(\mathbf{z})\mathbf{t} = \mathbf{0}$ equals $\tilde{\mathbf{B}}(\mathbf{z})\tilde{\mathbf{t}} = \mathbf{s}$, where $\tilde{\mathbf{B}}(\mathbf{z})$ equals $\mathbf{B}(\mathbf{z})$ but without its $(L + 1)$:th column $\mathbf{s}$ and $\tilde{\mathbf{t}}$ equals $\mathbf{t}$ without its $(L + 1)$:th element v_0 . Hence, $\tilde{\mathbf{B}}(\mathbf{z})$ has dimensions $K \times 3L$. The results from the previous sections imply that with probability 1, $\tilde{\mathbf{B}}(\mathbf{z})$ is full rank and thus invertible; hence, we obtain the solution $\tilde{\mathbf{t}}$ as $\tilde{\mathbf{t}} = \tilde{\mathbf{B}}^{-1}(\mathbf{z})\mathbf{s}$. In case $\mathbf{z} = \mathbf{z}_{\text{DFT}}$, we use the pseudo-inverse of $\tilde{\mathbf{B}}(\mathbf{z}_{\text{DFT}})$, denoted as $\tilde{\mathbf{B}}^+(\mathbf{z}_{\text{DFT}})$. Thus, in both cases, we can write $\tilde{\mathbf{t}} = \tilde{\mathbf{B}}^+(\mathbf{z})\mathbf{s}$ as our solution. Adding noise to this system, we get $\tilde{\mathbf{C}}(\mathbf{z})\tilde{\mathbf{t}} = \mathbf{s} + \mathbf{w}$, where $\tilde{\mathbf{C}}(\mathbf{z}) = (-1)^{L+1}(\tilde{\mathbf{B}}(\mathbf{z}) + \tilde{\mathbf{W}}(\mathbf{z}))$ and $\tilde{\mathbf{W}}(\mathbf{z})$ equals $\mathbf{W}(\mathbf{z})$ but without its $(L + 1)$:th column $\mathbf{w}$. Hence, under noise, we have $\tilde{\mathbf{t}} = \tilde{\mathbf{C}}^+(\mathbf{z})(\mathbf{s} + \mathbf{w})$. Noise impact analysis due to the inversion is also left as future

work. From this concise representation we can formulate the following estimation algorithm.

AoA Estimation Algorithm (AoA-DTFT)

Inputs: $L, N, \{r_k, z_k = e^{i\pi\phi_k}\}_{0 \leq k \leq K-1}$.

Constraints:

$$K \geq 3L$$

$$z_k, 0 \leq k \leq K-1, \text{ are all different}$$

$$2L \leq N$$

Outputs: AoA and gain estimates $\hat{\theta}_l, \hat{g}_l, 0 \leq l \leq L-1$.

Procedure:

- 1) Construct the $K \times 3L$ matrix $\tilde{\mathbf{C}}(\mathbf{z})$ by removing the $(L+1)$:th column $\mathbf{s}$ in (16), replacing s_k with the observed noisy samples r_k and multiplying the resulting matrix with $(-1)^{L+1}$.
 - 2) Construct $\tilde{\mathbf{t}} = \tilde{\mathbf{C}}^+(\mathbf{z})\mathbf{r}$.
 - 3) Let $v(z) = (-1)^L + \sum_{j=0}^{L-1} \tilde{t}_j z^{L-j}$. The L roots of the polynomial $v(z)$ are the AoA estimates $\hat{\theta}_l, 0 \leq l \leq L-1$.
 - 4) Construct $\hat{\mathbf{g}} = \left(\mathbf{A}(\phi)^H \mathbf{A}(\hat{\boldsymbol{\theta}})\right)^{-1} \mathbf{r}$, which are the gain estimates $\hat{g}_l, 0 \leq l \leq L-1$.
-

Note that the complexity of AoA-DTFT is mainly determined by inverting the $K \times 3L$ matrix $\tilde{\mathbf{C}}(\mathbf{z})$ followed by finding the roots of the degree L polynomial $v(z)$. For $L \leq 4$, it is well known that there are closed form formulas for the roots of $v(z)$ in terms of its coefficients; for larger L , the roots can be found by any popular root searching algorithm. In contrast, ML estimation involves a search across the L -dimensional cube $[-1, 1]^L$ as well as the complex gains $g_l, 0 \leq l \leq L-1$, which is of prohibitive complexity even for very small values of L .

In the next section, we will demonstrate the performance of AoA-DTFT. Before doing so, however, we use the results derived so far to answer all four questions posed in Section II-B:

- 1) It is sufficient with $N \geq 2L$ antennas.
- 2) The lowest we get with the results in this work is $2L$ measurements, achieved with $2L$ DFT beams.
- 3) Any DTFT angles are fine. However, using non-DFT measurement beams requires at least $3L$ measurements.

- 4) Yes. The AoA-DTFT algorithm is a matrix inversion method followed by a one-dimensional root search in $[-1, 1]$; it avoids searching across the L -dimensional cube $[-1, 1]^L$.

VI. NUMERICAL RESULTS

Herein, we will show the performance of AoA-DTFT for a specific simulation setup. The simulations in this section assume the model in (2), with the following simulation setup

- $c_j \in \mathcal{CN}(0, 1)$, where $\mathcal{CN}(a, b)$ denotes the complex normal distribution with mean a and variance b .
- AoAs θ_j are uniformly distributed in $[-1, 1]$.
- $w_k \in \mathcal{CN}(0, N_0)$, $0 \leq k \leq K - 1$ and SNR is defined as $1/N_0$.
- The DTFT angles $\phi_k = -1 + 2l_k/N$, $l_k \in [0, N - 1]$, i.e. DFT angles are used for measurement.

After the DTFT measurements, the receiver constructs a beam $\frac{\mathbf{u}}{\sqrt{N}}$ which ideally should be well matched with the channel in (1), i.e., have a large value of the beamforming gain $g = |\mathbf{u}^H \mathbf{h}|^2 / N$ where $\mathbf{h} = \sum_{j=0}^{L-1} \mathbf{a}(\theta_j) c_j$.

Figures 2, 3 and 4 depict beamforming gain g versus SNR for different simulation setups and algorithms. The value of K is referred to as "number of sweeps". When $K < N$, the DFT angles are spaced uniformly in order to cover the whole interval $[-1, 1]$. The title of each figure shows the value of N and the value of L in (2) which is denoted as L_{true} . Following algorithms are compared

- Max SNR beam: choose $\mathbf{u} = \arg \max_{\mathbf{a}(\phi_k)} |\mathbf{u}^H \mathbf{h}|$, i.e., $\mathbf{u}$ is the DTFT beam that produces the largest SNR during the measurements. This is a conventional algorithm commonly used in practical beam sweeping implementations and is our benchmark.
- AoA-DTFT, $L = 1$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$: AoA-DTFT algorithm used with the input $L = 1$, i.e., recovery of LOS assuming that there is only a single path in the channel. The beam $\mathbf{u}$ equals $\mathbf{a}(\hat{\theta}_0)$.
- AoA-DTFT, $L = 2$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$: AoA-DTFT algorithm used with the input $L = 2$. $\mathbf{u}$ equals the complex phases of $\mathbf{h}_{\text{est}} = \sum_{j=0}^{L-1} \mathbf{a}(\hat{\theta}_j) \hat{g}_j$, i.e., $\mathbf{u} = e^{\sqrt{-1} \text{angle}(\hat{\mathbf{h}})}$ where $\text{angle}(\mathbf{v})$ is a vector of complex phases of the elements in $\mathbf{v}$.
- AoA-DTFT, $L = 2$, $\mathbf{u} = \text{strongest AoA}$: AoA-DTFT algorithm used with the input $L = 2$. $\mathbf{u} = \mathbf{a}(\hat{\theta}_0)$ if $|\hat{g}_0| > |\hat{g}_1|$ and $\mathbf{u} = \mathbf{a}(\hat{\theta}_1)$ otherwise.
- Optimal beam: $\mathbf{u} = e^{i \text{angle}(\mathbf{h})}$.

When L_{true} is known in advance, it is clear that "AoA-DTFT, $L = L_{\text{true}}$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$ " converges to "Optimal beam" since in the noiseless case it estimates $\mathbf{h}$ perfectly. In Figure 2, when there are two paths in channel and this is assumed to be known in advance, "AoA-DTFT, $L = L_{\text{true}}$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$ " is able to converge to optimal performance rather quickly; none of the other methods are able to converge to optimal performance. Moreover, it is able to do this even with 4 sweeps, although at significantly higher SNRs as expected. In contrast, "AoA-DTFT, $L = 1$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$ " performs worse than the benchmark, which shows the importance of having some knowledge about the number of paths in the channel in advance.

Figure 3 shows that if $L < L_{\text{true}}$, then "AoA-DTFT, $L = 2$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$ " is unable to converge to the optimal beam, although it still provides the best performance. This shows that it is still able to estimate the two strongest paths in the channel, while the rest of the paths (four in this case) will act as noise that is independent of SNR which thus produces a constant gap to the optimal beam. Finally, Figure 4 shows the performance when we have a larger ULA of 32 antennas. Clearly, even with as little as 8 sweeps, "AoA-DTFT, $L = L_{\text{true}}$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$ " converges to the performance of "Optimal beam". However, in this case, the "Max SNR beam" provides best performance at low SNRs. The reason for this is that a larger N increases the spatial directivity of the channel, which requires more accurate AoA estimation and thus a higher SNR to achieve lower estimation error.

VII. CONCLUSIONS AND FUTURE WORK

This work has studied estimation of L AoAs from DTFT beam (a.k.a. analog beam) measurements at a ULA with N antennas. The main purpose of the work is to investigate the number of measurements necessary to estimate the L AoAs. First, basic questions about uniqueness of the estimation in the noiseless case were formulated in Section II-B. It was shown that $N \geq 2L$ is a sufficient condition for unique recovery of the L AoAs. The estimation problem is then expressed in terms of sampling a rational function, for which there exists a rich theory. However, direct application of this theory to our rational function requires in general at least $N + 2L$ measurements, which is more than what application of MPM, Nyquist sampling or compressed sensing to our problem require, as showed in sections II-C and III. Instead, it was noted that our rational function possesses sparsity which can be utilized to reduce the number of measurements from $N + 2L$ to $3L$ for any DTFT measurement beams. Moreover, if DFT beams are chosen for measurement, with N as the DFT period, $3L$ measurements are further reduced to

Fig. 2. The "AoA-DTFT, $L = L_{\text{true}}$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$ " beam construction converges to optimal performance and provides overall best performance. Convergence to optimum occurs even with four beam sweeps.

Fig. 3. The "AoA-DTFT, $L = 2$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$ " beam construction provides best performance but no longer converges to optimal performance. The reason for this loss is that it uses only two paths from the channel for estimation while the channel has six paths, so the other four paths become noise that is independent of SNR and thus produces a constant gap to optimum performance.

Fig. 4. When $N = 32$, the "AoA-DTFT, $L = L_{\text{true}}$, $\mathbf{u} = \text{angle}(\mathbf{h}_{\text{est}})$ " provides best performance after a certain SNR and converges slower to optimal performance than when $N = 8$. The reason is that a larger N increases the spatial directivity, which requires more accurate AoA estimation and thus a high enough SNR.

$2L$ measurements. In cases when $N \gg L$, such as mmWave, this is a significant improvement to previous work on the problem which requires at least $\mathcal{O}(L \log(N/L))$ samples. These noiseless results are easily applied to the noisy case, resulting in a relatively simple AoA estimation algorithm denoted as AoA-DTFT. Numerical simulation of beamforming gain demonstrate the feasibility of AoA-DTFT and its superior performance compared to the conventional algorithm that chooses the DTFT beam with highest SNR during the measurements. As discussed in Section V, a more accurate analysis of the noise present in this algorithm leaves room for further improvements of it in future work.

REFERENCES

- [1] NR; Physical layer; General description, 38.201, Jan. 2022. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.201/
- [2] Pei Zhou *et al*, "IEEE 802.11ay based mmWave WLANs: Design Challenges and Solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, Mar. 2018.
- [3] A. Alkhateeb, O. E. Ayach, G. Leus, R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, Oct. 2014.
- [4] D. Tse, P. Viswanath, "Fundamentals of Wireless Communication," Cambridge, UK: Cambridge University Press, 2005.

- [5] A. Alkhateeb, G. Leus, R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems", *IEEE Transactions on Wireless Communications (revised)*, Mar. 2015.
- [6] H. Ghauch, T. Kim, C. Fischione, M. Skoglund, "Compressive Sensing with Applications to Millimeter-wave Architectures," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019.
- [7] X. Li, J. Fang, H. Li, P. Wang, "Millimeter Wave Channel Estimation via Exploiting Joint Sparse and Low-Rank Structures," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, Feb. 2018.
- [8] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, Mar. 1986.
- [9] R. Roy, T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, Jul 1989.
- [10] T. K. Sarkar, O. Pereira, "Using the matrix pencil method to estimate the parameters of a sum of complex exponentials," *IEEE Antennas and Propagation Magazine*, vol. 37, no. 1, Feb. 1995.
- [11] R. A. Horn, C. R. Johnson (1991), "Topics in matrix analysis," Cambridge, UK: Cambridge University Press, 1991.
- [12] J. Stoer, R. Bulirsch, "Introduction to Numerical Analysis (Texts in Applied Mathematics, 12)," 3rd ed. Würzburg, München: Springer-Verlag, 2002.