Super-Resolution Channel Estimation for Arbitrary Arrays in Hybrid Millimeter-Wave Massive MIMO Systems

Yue Wang, *Member, IEEE*, Yu Zhang, *Student Member, IEEE*, Zhi Tian, *Fellow, IEEE*, Geert Leus, *Fellow, IEEE*, and Gong Zhang, *Member, IEEE*

***Abstract*—This paper develops efficient channel estimation techniques for millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems under practical hardware limitations, including an arbitrary array geometry and a hybrid hardware structure. Taking on an angle-based approach, this work adopts a generalized array manifold separation approach via Jacobi-Anger approximation, which transforms a non-ideal, non-uniform array manifold into a virtual array domain with a desired uniform geometric structure to facilitate super-resolution angle estimation and channel acquisition. Accordingly, structure-based optimization techniques are developed to effectively estimate both the channel covariance and the instantaneous channel state information (CSI) within a short sensing time. The different time-varying scales of channel path angles versus path gains are capitalized to design a two-step CSI estimation scheme that can quickly sense fading channels. Theoretical results are provided on the fundamental limits of the proposed technique in terms of sample efficiency. For computational efficiency, a fast iterative algorithm is developed via the alternating direction method of multipliers. Other related issues such as spurious-peak cancellation in non-uniform linear arrays and extensions to higher-dimensional cases are also discussed. Simulations testify the effectiveness of the proposed approaches in hybrid mmWave massive MIMO systems with arbitrary arrays.**

***Index Terms*—Arbitrary array, gridless compressive sensing, hybrid structure, Jacobi-Anger approximation, mmWave massive MIMO, super-resolution channel estimation, Vandermonde structure.**

Manuscript received February 20, 2019; revised June 6, 2019; accepted July 27, 2019. Date of publication August 26, 2019; date of current version September 20, 2019. This work was supported in part by the U.S. National Science Foundation under Grants 1547364, 1527396, 1546604, and 1730083; in part by the National Science Foundation of China under Grant 61871218; and in part by the frame of the ASPIRE project (project 14926 within the OTP program of NWO-TTW). The guest editor coordinating the review of this article and approving it for publication was Prof. Marius Pesavento. *(Corresponding author: Yue Wang.)*

Y. Wang and Z. Tian are with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030 USA (e-mail: ywang56@gmu.edu; ztian1@gmu.edu).

Y. Zhang is with the Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China, and also with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030 USA (e-mail: skywalker_zy@nuaa.edu.cn).

G. Leus is with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft 2826 CD, The Netherlands (e-mail: g.j.t.leus@tudelft.nl).

G. Zhang is with the Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China (e-mail: gzhang@nuaa.edu.cn).

Digital Object Identifier 10.1109/JSTSP.2019.2937632

I. Introduction

In millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) communications, large antenna gains coupled with the availability of large bandwidths bring many desired benefits such as high throughput, large capacity, and robustness against fading and interference [1], [2], which all hinge on accurate channel knowledge. However, the increase in antennas results in an enlarged channel dimension that gives rise to challenges to traditional channel estimation techniques, in terms of the high signal acquisition cost and the large training overhead [3]–[5]. The hinderance in channel estimation is further aggravated by practical hardware limitations. A hybrid analog-digital architecture is widely suggested for massive MIMO transceivers, which reduces the number of radio frequency (RF) chains by balancing between the analog RF part and the digital baseband part [6], [7]. However, under such a hybrid structure, the channel estimator at baseband can only observe a compressed representation of the channel through a few RF chains. Further, the array geometry of massive MIMO can be shifted from its ideal configuration, or even arbitrary.

To overcome these challenges, compressive sensing (CS) has been advocated for channel estimation in mmWave massive MIMO systems [8]–[14]. These CS-based approaches exploit the channel sparsity that stems from the limited scattering characteristics of mmWave propagation [15]–[18]. Through virtual channel modeling [19], the large-dimensional mmWave massive MIMO channels can be represented by only a small number of parameters, including the angles of departure/arrival (AoDs/AoAs) and the path gains of the sparse scattering paths. Therefore, CS techniques enable channel estimation from a small set of compressively collected training samples. In [8], [9], a sparse multipath channel is formulated as a sparse vector on the angle-delay-Doppler space, and then CS techniques are applied to recover the vectorized sparse channel. In [10], an adaptive CS-based algorithm is proposed to estimate the sparse channel with a hybrid analog-digital hardware architecture. In [11], a hybrid architecture based on phase shifters is proposed to recover the sparse channel via greedy search algorithms. To further reduce the power consumption of phase shifters, a switch-based hybrid architecture is developed in [12], for sparse channel estimation. In [13], to reduce the problem complexity, CS-based channel estimation is divided into angle estimation and path gain estimation subproblems, which are solved sequentially. In [14], the CS-based channel estimation scheme is extended to broadband mmWave MIMO systems.

All the aforementioned techniques aim to estimate the instantaneous channel state information (CSI). Another line of work focuses on estimating the channel statistics, such as the channel covariance [20], [21]. The channel covariance is an important second-order statistics, which remains constant over many channel coherence intervals and therefore can be used for statistics-based design of the precoders, beamformers and linear receivers [22], [23]. To estimate the second-order statistics of the vectorized sparse mmWave MIMO channel, a diagonal-search orthogonal matching pursuit algorithm is developed in [20], which not only utilizes the joint sparsity represented by the available multiple measurement vectors (MMV) but also takes advantage of the Hermitian structure of the channel covariance matrix. In [21], a CS-based channel covariance estimator is proposed by using dynamic sensing schemes and designing dynamic greedy pursuit algorithms for the hybrid architecture.

Existing CS-based channel estimators critically hinge on an on-grid assumption that the values of the AoD/AoA of each propagation path exactly reside on some predefined grid in the angular domain. However, in practice, the AoDs/AoAs of paths are continuous-valued off grid. As a result, CS-based methods suffer from degraded performance due to the power leakage effect around the recovered discrete grid points, a.k.a., the infamous basis mismatch problem [24]. An angle rotation technique is proposed to alleviate this problem, which is developed based on the spatial basis expansion model [25], [26]. It improves the estimation accuracy in the angular domain, but still experiences finite resolution due to some predefined spatial rotation parameters. A continuous basis pursuit technique is proposed for perturbed CS in [27], which is however limited through the series expansion. On the other hand, classical subspace methods, such as MUSIC and ESPRIT, can achieve super-resolution in angle estimation [28], [29]. But, they require a large number of snapshots for collecting sample statistics, which leads to a long sensing time and consumes large training resources. To circumvent the on-grid assumption required by traditional CS and achieve super-resolution at short sensing time, a gridless CS technique is developed via atomic norm minimization (ANM) in the form of semidefinite programming (SDP) [30]–[32]. As a structure-based optimization technique, gridless CS is applied for super-resolution channel estimation in mmWave massive MIMO systems [33]–[35], which utilizes not only the sparsity of channels but also the structure of antenna arrays.

By capitalizing on the critical Vandermonde structure, gridless CS implicitly assumes the use of an ideal uniform array geometry, that is, the antennas have to be uniformly placed with exactly the same separation distance. However, in practical applications, arbitrary arrays arise in several cases, instead of the perfect uniform arrays. For example, the antenna separation distance is measured in the millimeter range over the mmWave frequency bands. Thus, an ideal uniform array geometry is hard to guarantee due to calibration errors introduced in the manufacturing process and/or antenna installation. Another case of arbitrary arrays appears due to sub-array selection. For example, for the purpose of energy saving in the switch-based hybrid architecture, only a small number of antennas is switched to link the RF chains [36]. Arbitrary arrays no longer present the well-featured Vandermonde structure explicitly in the array manifold, which then excludes the use of a large number of geometric-based channel estimation techniques. To overcome this problem, array manifold separation techniques have been developed in array signal processing [37]–[39]. However, the structural feature presented via manifold separation is not efficiently utilized in [37], [38], where conventional subspace methods are used based on a large number of samples. When the sample size is limited, [39] employs the well-structured Fourier series approximation for manifold separation to enable ANM, but it suffers from an expensive computational complexity in order to reduce the approximation error in the substantially expanded Fourier domain.

In this paper, addressing all the aforementioned challenges cohesively, we seek to design high-performance, low-cost channel estimation solutions for arbitrary arrays in hybrid mmWave massive MIMO systems. Specifically, we propose a super-resolution channel estimation framework that not only utilizes the special channel features of the sparse mmWave massive MIMO propagation, but also fully considers the non-ideal array geometry and practical hardware limitations. This framework offers several channel estimation solutions and enables to obtain both the channel statistics and the instantaneous CSI, depending on whether the transceiver design is built on channel covariance [22], [23] or the channel itself [40]. We propose two super-resolution solutions for channel covariance estimation (CCE) through efficient structure-based optimization techniques, with samples collected from multiple snapshots. One is the CCE via low-rank structured covariance reconstruction (LRSCR), which provides super-resolution accuracy at a low computational cost. The other is the CCE via Dynamic-ANM, which further allows for a dynamic configuration where the hybrid hardware parameters change over time for better performance. For block fading channels, given the estimate from CCE, the instantaneous CSI can then be estimated in a timely fashion. We recognize that the angles change slowly and can take a long time to acquire accurately from the channel statistics, whereas the path gains vary frequently but are easy to acquire given the estimated angles. Accordingly, the instantaneous CSI estimation is divided by solving two subproblems sequentially, i.e., angle estimation and path gain estimation. In developing these novel super-resolution channel estimation approaches, this work contains the following contributions.

- We leverage a generalized array manifold separation approach to extract the useful geometric structure for a practical system with an arbitrary or imperfect array geometry. In particular, we transform the sparse mmWave massive MIMO channel representation from the physical arbitrary antenna domain to a virtual uniform antenna domain via the Jacobi-Anger approximation [41]. Our method enables gridless CS to exploit the useful Vandermonde structure presented in the virtual uniform array manifold.
- This work not only develops super-resolution channel estimation solutions, but also investigates the fundamental

limits of gridless CS based channel estimation under the constraints of arbitrary arrays and hybrid structures. Our theoretical results shed light on the minimum number of RF chains required by super-resolution channel estimation, as well as the lower and upper bounds on the mode order selected for the Jacobi-Anger approximation. This leads to a tradeoff between the hardware cost of sparse channel estimation and the approximation accuracy to combat the imperfect array geometry.

- To reduce the high computational complexity of the channel estimation solutions based on the SDP formulation, we design a fast algorithm through the alternating direction method of multipliers (ADMM) [42]. It provides an efficient iterative implementation with much lower computational complexity than that of the SDP solvers using the interior-point method.
- We tackle several implementation issues. Specifically, we overcome the side effect of the array manifold separation operation, by removing the spurious peaks generated by the Jacobi-Anger approximation for non-uniform linear array (non-ULA) cases. We also extend our work to 2-dimensional (2D) scenarios, where both the BS and the MS are equipped with multiple antennas.

The rest of this paper is organized as follows. Section II presents the system model and problem formulation for sparse channel estimation in hybrid arbitrary arrays, as well as the state-of-the-art of gridless CS techniques for ideal models. Section III proposes a super-resolution channel estimation framework based on the array manifold separation, in which different channel estimation solutions are developed for obtaining the channel covariance and the instantaneous CSI. Specific issues related to the proposed techniques are discussed in Section IV. Simulation results are presented in Section V, followed by conclusions in Section VI.

Notations: a is a scalar, $\boldsymbol{a}$ denotes a vector, $\boldsymbol{A}$ is a matrix, and $\mathcal{A}$ represents a set. $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ are the transpose, conjugate, and conjugate transpose of a vector or matrix, respectively. conv$(\mathcal{A})$ means the convex hull of a set $\mathcal{A}$. Real$(\cdot)$ and Imag$(\cdot)$ compute the real part and the imaginary part of a vector or matrix, respectively. $|a|$ denotes the absolute value of a. $\|\boldsymbol{a}\|_2$ is the ℓ_2 norm of $\boldsymbol{a}$. diag$(\boldsymbol{a})$ and diag$(\boldsymbol{A}_1, \boldsymbol{A}_2)$ denote a diagonal matrix with the diagonal elements constructed from $\boldsymbol{a}$ and a block diagonal matrix with the submatrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, respectively. $\boldsymbol{I}$ is an identity matrix and $\boldsymbol{I}_a$ is an anti-diagonal identity matrix. T$(\boldsymbol{u})$ is a Hermitian Toeplitz matrix with first column being $\boldsymbol{u}$. $\|\boldsymbol{A}\|_F$, $\boldsymbol{A}^\dagger$, and tr$(\boldsymbol{A})$ are the Frobenius norm, the pseudoinverse, and the trace of $\boldsymbol{A}$, respectively. The operation vec$(\cdot)$ stacks all the columns of a matrix into a vector. $\otimes$ is the Kronecker product of matrices or vectors. $\mathbb{E}\{\cdot\}$ denotes expectation.

II. Models and Preliminaries

In this section, we first present the signal model and state the goal of both CCE and CSI estimation. Then, we overview the prior work on relevant super-resolution techniques that are only applicable for ideal uniform arrays, e.g., the uniform linear array (ULA), and under a fixed hybrid hardware structure.

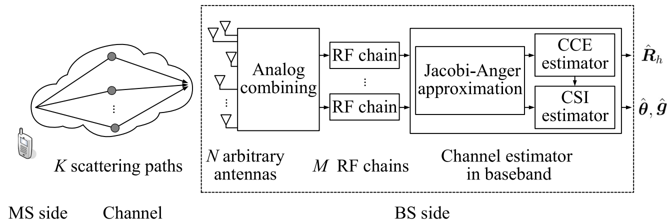


Fig. 1. System model of sparse channel estimation in hybrid mmWave massive MIMO systems with arbitrary arrays.

A. Channel and Signal Models

Consider a narrowband[1] mmWave massive MIMO time division duplex (TDD)[2] system for channel estimation conducted at the base station (BS). As shown in Fig. 1, the BS has a hybrid structure equipped with N arbitrarily deployed antennas and M $(M < N)$ RF chains. For simplicity, we mainly focus on the basic single-antenna case at the mobile station (MS), while we extend to the multiple-antenna MS case in Section IV-D. Noticeably, as shown in Fig. 1, to impose the useful Vandermonde structure in an arbitrary array geometry, a preprocessing block via Jacobi-Anger approximation is added to the channel estimator in the hybrid mmWave massive MIMO system, which will be described in Section III-A.

At the mmWave frequency, the wireless channel experiences limited scattering propagation, which results in a sparse multipath structure [15]–[18], as shown in Fig. 1. In this sense, the channel can be described by a geometric model with K $(K < M < N)$ scatterers[3], in which each path is parameterized by the path angle and the path gain. For simplicity, suppose that each scatterer contributes to one propagation path, which can be straightforwardly extended to cluster scattering where each cluster includes multiple scattering paths [40]. Further, in mmWave channels, the angles of the scattering paths remain constant for a relatively long time, while the channel coefficients change very rapidly [18]. Accordingly, the uplink channel $\boldsymbol{h}_t$ can be expressed as the sum of K paths in the form

$$\boldsymbol{h}_t = \sum_{k=1}^{K} g_{k,t} \boldsymbol{a}(\theta_k), \ \forall t, \tag{1}$$

where $g_{k,t}$ denotes the channel gain for the k-th scattering path at the t-th snapshot, and $\boldsymbol{a}(\theta_k) \in \mathbb{C}^N$ is the array manifold vector corresponding to the k-th channel path.

In this work, we focus on arbitrary arrays, in which the n-th antenna element is placed at a known location $(r_n, \phi_n), n = 1, \ldots, N$ in polar coordinates, where r_n and ϕ_n represent the distance from the polar origin to the n-antenna element and the angle between the polar axis and the direction from the polar

[1]In a wideband case with frequency selectivity, the continuous-valued delays of individual paths of sparse time-dispersive channels can be estimated via gridless CS for super-resolution accuracy in the time domain [43].

[2]This work can be applied to frequency division duplex (FDD) systems as well, given the angle reciprocity between uplink and downlink [26].

[3]We focus on the point scatterer case in this work. The angle spread issue due to the reflecting area of shaped scatterers is out of scope of this paper, whose impact on the proposed methods will be studied in future work.

origin to the n-antenna element, respectively. Define θ_k as the angle between the polar axis and the k-th path, and take the polar origin as the reference point. Then, the n-th component of the array manifold vector for the k-th path can be written as

$$[\boldsymbol{a}(\theta_k)]_n = e^{j2\pi\frac{r_n}{\lambda}\cos(\theta_k - \phi_n)}, \tag{2}$$

where λ denotes the wavelength. In a compact matrix-vector form, the channel $\boldsymbol{h}_t$ in (1) can be rewritten as

$$\boldsymbol{h}_t = \boldsymbol{A}\boldsymbol{g}_t, \tag{3}$$

where $\boldsymbol{g}_t = [g_{1,t}, \ldots, g_{K,t}]^T$ and $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_K)]$.

In uplink channel estimation, the MS sends out training symbols z_t which are also known to the BS. For simplicity, let $|z_t| = 1$ for all snapshots. Then, the received signal at the BS's antennas can be represented as

$$\boldsymbol{x}_t = \boldsymbol{h}_t z_t + \boldsymbol{w}_t = \boldsymbol{A}\boldsymbol{g}_t z_t + \boldsymbol{w}_t, \tag{4}$$

where $\boldsymbol{w}_t$ denotes additive Gaussian noise with complex Gaussian distribution $\mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. According to (4), the covariance matrices for $\boldsymbol{x}_t, \boldsymbol{h}_t$ and $\boldsymbol{g}_t$ have the following linear relationship:

$$\boldsymbol{R}_x = \mathbb{E}\{\boldsymbol{x}_t\boldsymbol{x}_t^H\} = \boldsymbol{R}_h + \sigma^2\boldsymbol{I} = \boldsymbol{A}\boldsymbol{R}_g\boldsymbol{A}^H + \sigma^2\boldsymbol{I}, \tag{5}$$

where $\boldsymbol{R}_h = \mathbb{E}\{\boldsymbol{h}_t\boldsymbol{h}_t^H\}$, and $\boldsymbol{R}_g = \mathbb{E}\{\boldsymbol{g}_t\boldsymbol{g}_t^H\}$.

The BS adopts a hybrid hardware structure in the form $\boldsymbol{W}_t = \boldsymbol{W}_t^{\text{BB}}\boldsymbol{W}_t^{\text{RF}}$, where $\boldsymbol{W}_t^{\text{BB}} \in \mathcal{C}^{M\times M}$ denotes a baseband digital combiner, and $\boldsymbol{W}_t^{\text{RF}} \in \mathcal{C}^{M\times N}$ is an analog combiner. In this paper, we focus on the case where $\boldsymbol{W}_t^{\text{RF}}$ is made of a network of random phase shifters, while this work can be applied to other structures such as the switch-based network as well [12]. To further enhance randomness, $\boldsymbol{W}_t^{\text{BB}}$ can be set as a random Gaussian matrix. After being multiplied with the hybrid combining matrix $\boldsymbol{W}_t \in \mathcal{C}^{M\times N}$ and the known training symbol z_t^*, the received baseband-equivalent signal is transformed into a lower dimension in the form

$$\boldsymbol{y}_t = z_t^*\boldsymbol{W}_t\boldsymbol{x}_t = \boldsymbol{W}_t\boldsymbol{A}\boldsymbol{g}_t + \boldsymbol{W}_t\boldsymbol{n}_t, \tag{6}$$

where $\boldsymbol{n}_t = z_t^*\boldsymbol{w}_t$.

In this paper, we assume a block fading channel, where path gains, and hence the CSI, stay constant within a block but vary from block to block. In contrast, path angles vary much slower, and stay unchanged across blocks, until angle re-calibration is needed. Further, angles can be retrieved from the channel covariance, which is the key idea behind the statistical inference methods for angle estimation. This motivates us to design a two-stage channel estimation framework to obtain both the channel covariance and the instantaneous CSI sequentially, as shown in Fig. 2. In the first stage, we apply CCE to obtain the channel covariance over multiple snapshots. Then, in the second stage, considering the difference in time-variation of path angles and path gains, we design a two-step scheme to do instantaneous CSI estimation, which is further divided into two subproblems: angle estimation and path gain estimation.

Remark 1: When the goal of channel estimation is CCE only, the estimator can terminate upon completing the first stage. The CCE by itself is relevant in two cases: one is to simplify either the channel estimation task or the transceiver design, and the other is when path gains experience fast fading that renders the CSI estimates useless for data transmission. In both cases, CCE-based transceiver design can be adopted [22], [23].

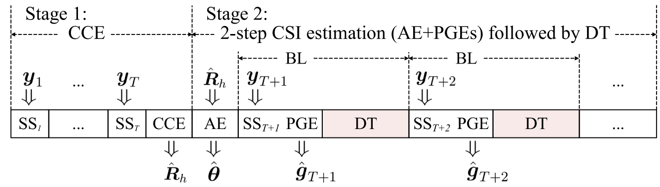


Fig. 2. Illustration of two-stage channel estimation including both channel covariance estimation and instantaneous CSI estimation that is further divided into angle estimation and path gain estimation.
SS: snapshot, BL: block length, DT: data transmission,
CCE: channel covariance estimation, AE: angle estimation, PGE: path gain estimation

Remark 2: Our task of CCE-based angle estimation is also useful during the system calibration stage for fixed wireless applications, in which case the estimated angles can be used to facilitate several system-level tasks such as user grouping and beam-sectoring.

B. Prior Art: Efficient Super-Resolution Techniques in Ideal Models

When the sensing time is not long enough for traditional super-resolution methods to work for massive MIMO with a large N, we have to focus on those super-resolution techniques that exploit the structural feature of the array manifold to improve the sample efficiency. Therefore, in this subsection, we overview existing efficient super-resolution techniques for channel estimation [33], [34], which critically rely on an assumption that the antenna array has to be a uniform array, e.g., a ULA. Then, the array manifold naturally presents an explicit Vandermonde structure. That is, the n-th component of the array manifold vector at the k-th path is a special case of (2) with $\phi_n = 0$, $r_n = d(n-1), \forall n$, which has the form

$$[\boldsymbol{a}(\theta_k)]_n = e^{j2\pi(n-1)\frac{d}{\lambda}\cos(\theta_k)}, \tag{7}$$

where d denotes the same separation distance between any two adjacent antennas placed along the polar axis, which is usually set as $d = \lambda/2$.

Further, assuming $\boldsymbol{W}_t = \boldsymbol{W}, \forall t$ as the fixed hybrid structure for the static channel sensing system, the received signals $\boldsymbol{y}_t$ for $t = 1, \ldots, T$ in (6) can be collected to form a matrix $\boldsymbol{Y}$ with $\boldsymbol{y}_t$ being its columns as

$$\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{A}\boldsymbol{G} + \boldsymbol{W}\boldsymbol{N} = \boldsymbol{W}\boldsymbol{H} + \boldsymbol{W}\boldsymbol{N}, \tag{8}$$

where the matrices $\boldsymbol{H}, \boldsymbol{G}$ and $\boldsymbol{N}$ are similarly defined as $\boldsymbol{Y}$ with $\boldsymbol{h}_t$, $\boldsymbol{g}_t$ and $\boldsymbol{n}_t$ being their columns, respectively. From (6) and (8), the covariance of $\boldsymbol{y}_t$ is given as

$$\boldsymbol{R}_y = \mathbb{E}\{\boldsymbol{y}_t\boldsymbol{y}_t^H\} = \boldsymbol{W}(\boldsymbol{R}_h + \sigma^2\boldsymbol{I})\boldsymbol{W}^H. \tag{9}$$

Then, an atom set in the MMV case is defined as [44], [45]

$$\mathcal{A} = \left\{\boldsymbol{a}(f)\,\boldsymbol{b}^H \,\middle|\, f \in \left(-\tfrac{1}{2}, \tfrac{1}{2}\right], \boldsymbol{b} \in \mathbb{C}^T, \|\boldsymbol{b}\|_2 = 1\right\}, \tag{10}$$

where $\boldsymbol{a}(f) \in \mathbb{C}^N$ with the n-th component being $e^{j2\pi(n-1)f}$. According to the atomic norm theorem [44], [45], the atomic norm of $\boldsymbol{H}$ over the atom set $\mathcal{A}$ is defined as

$$\|\boldsymbol{H}\|_{\mathcal{A}} = \inf\left\{l > 0 : \boldsymbol{H} \in l\,\mathrm{conv}(\mathcal{A})\right\}, \tag{11}$$

which seeks the most concise representation of $\boldsymbol{H}$ by involving the fewest atoms over $\mathcal{A}$.

From the received signals $\boldsymbol{Y}$ in (8), the channel estimation for the instantaneous CSI is conducted by solving the regularized ANM formulation as

$$\hat{\boldsymbol{H}} = \arg\min_{\boldsymbol{H}} \|\boldsymbol{H}\|_{\mathcal{A}} + \frac{\tau}{2}\|\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{H}\|_F^2, \tag{12}$$

where τ denotes the regularization coefficient controlling the tradeoff between the ANM and the residual error tolerance to the observations. According to [46], τ can be set as $\tau = 1/(\sigma + \frac{\sigma}{\log N}\sqrt{N\log N + N\log(4\pi\log N)}\,)$.

Besides the instantaneous CSI itself, the second-order channel statistics in terms of the channel covariance $\boldsymbol{R}_h$ also play an important role in precoding design for mmWave massive MIMO channels [22], [23]. Suppose the channel gains of the sparse paths are uncorrelated with each other. Then, $\boldsymbol{R}_h$ not only presents the low rankness due to the channel sparsity, but also is a well-structured positive semidefinite (PSD) Hermitian Toeplitz matrix thanks to the Vandermonde structure of uniform arrays.

To utilize these useful features in the channel statistics, a structure-based optimization approach named low-rank structured covariance reconstruction (LRSCR) [44], [47], can be applied to do CCE, by imposing the Hermitian Toeplitz structure on $\boldsymbol{R}_h$ as $\boldsymbol{R}_h = \mathrm{T}(\boldsymbol{u}_h)$. Thus, from the sample covariance matrix $\hat{\boldsymbol{R}}_y = \frac{1}{T}\boldsymbol{Y}\boldsymbol{Y}^H$, the CCE can be conducted via LRSCR as,

$$\begin{aligned}\hat{\boldsymbol{R}}_h = \ &\arg\min_{\mathrm{T}(\boldsymbol{u}_h)} \mathrm{tr}\left(\mathrm{T}(\boldsymbol{u}_h)\right) + \frac{\tau}{2}\left\|\hat{\boldsymbol{R}}_y - \boldsymbol{W}\mathrm{T}(\boldsymbol{u}_h)\boldsymbol{W}^H\right\|_F^2 \\ &\text{s.t.} \quad \mathrm{T}(\boldsymbol{u}_h) \succeq 0.\end{aligned} \tag{13}$$

Unfortunately, the uniform array assumption cannot be guaranteed in practice, considering the antenna misalignment and subarray selection issues arisen in hybrid mmWave massive MIMO systems. Moreover, the assumption of the fixed combining matrix over all snapshots is ineffective to find out all potential scattering paths in the dynamic scenario of mmWave channel estimation, where the hybrid hardware structure varies snapshot by snapshot. Regarding these practical situations, two questions arise: 1) can we design new super-resolution channel estimation techniques via LRSCR and ANM for arbitrary arrays? 2) how can we retrieve the desired channel information via the dynamic hybrid structure? In the next section, to fully address these problems, we develop super-resolution and fast channel estimation approaches for obtaining both the channel covariance and the instantaneous CSI.

III. Super-Resolution Channel Estimation for Arbitrary Arrays

In this section, we first leverage the Jacobi-Anger approximation to extract the useful Vandermonde structure from a non-ideal array geometry. Then, two new CCE methods are developed through structure-based optimization techniques for arbitrary arrays. For efficient estimation of the instantaneous CSI, a two-step solution is proposed, which estimates the path angles and path gains sequentially.

A. Jacobi-Anger Approximation

The Jacobi-Anger expansion provides a general infinite terms expansion of exponentials of trigonometric functions in the basis of their harmonics [41]. Specifically, (2) in the Jacobi-Anger expansion form is expressed as

$$[\boldsymbol{a}(\theta_k)]_n = \sum_{i=-\infty}^{+\infty} j^i \mathrm{J}_i\left(2\pi\frac{r_n}{\lambda}\right) e^{-j\phi_n i} e^{j\theta_k i}, \tag{14}$$

where $\mathrm{J}_i(\cdot)$ denotes the i-th order Bessel function of the first kind.

Although (14) indicates a summation of infinite series, the value of $|\mathrm{J}_i(r)|$ decays very rapidly as the value of $|i|$ increases for any $r > 0$, which is a nature of the Bessel function. Thus, the infinite series expansion can be well-approximated by keeping only the terms having large absolute values located around the central range of the series, that is, $|i| \leq I$ in (14). To reach a desired precision, the maximum mode order I for the approximation is chosen as [48]

$$I > \frac{2\pi}{\lambda} r_{\max}, \tag{15}$$

where r_{max} is the maximum r_n, $\forall n$.

Then, given I, (14) can be approximately expressed as

$$\begin{aligned}[\boldsymbol{a}(\theta_k)]_n &\approx \sum_{i=-I}^{I} j^i \mathbf{J}_i\left(2\pi\frac{r_n}{\lambda}\right) e^{-j\phi_n i} e^{j\theta_k i} \\ &= \boldsymbol{c}_n^T \boldsymbol{v}(\theta_k),\end{aligned} \tag{16}$$

where $\boldsymbol{c}_n$ and $\boldsymbol{v}(\theta_k)$ are given by

$$[\boldsymbol{c}_n]_i = j^i \mathbf{J}_i\left(2\pi\frac{r_n}{\lambda}\right) e^{-j\phi_n i}, \ i = -I, \ldots, I; \tag{17a}$$

$$[\boldsymbol{v}(\theta_k)]_i = e^{j\theta_k i}, \ i = -I, \ldots, I. \tag{17b}$$

According to (16), the $N \times 1$ array manifold vector of an arbitrary array can be approximated as

$$\boldsymbol{a}(\theta_k) = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N]^T \boldsymbol{v}(\theta_k) = \boldsymbol{C}\boldsymbol{v}(\theta_k). \tag{18}$$

Noticeably, it is the Jacobi-Anger expansion that enables to separate the unknown channel-related parameter (θ_k) in (17b) from the known array-related configurations (r_n and ϕ_n) in (17a), which are otherwise mingled in the original physical antenna domain of (2). Further, due to the exponential form in (17b), the Vandermonde structure is well-presented in the virtual uniform antenna domain, in terms of the virtual array manifold

$$\boldsymbol{V}(\boldsymbol{\theta}) = [\boldsymbol{v}(\theta_1), \ldots, \boldsymbol{v}(\theta_K)]. \tag{19}$$

Merging (18) and (19), any arbitrary array can be expressed as the multiplication of a Bessel matrix and a Vandermonde matrix, in the form

$$\boldsymbol{A} = \boldsymbol{C}\boldsymbol{V}. \tag{20}$$

To appreciate the Vandermonde structure, we reformulate the channel by substituting (20) into (3) as

$$\boldsymbol{h}_t = \boldsymbol{C}\boldsymbol{V}\boldsymbol{g}_t = \boldsymbol{C}\boldsymbol{d}_t, \tag{21}$$

where $\boldsymbol{d}_t = \boldsymbol{V}\boldsymbol{g}_t$ represents the virtual channel in the virtual uniform antenna domain. From (5) and (20), the channel covariance can be rewritten as

$$\boldsymbol{R}_h = \boldsymbol{C}\boldsymbol{V}\boldsymbol{R}_g\boldsymbol{V}^H\boldsymbol{C}^H = \boldsymbol{C}\boldsymbol{R}_v\boldsymbol{C}^H, \tag{22}$$

where $\boldsymbol{R}_v$ denotes the virtual channel covariance matrix given by

$$\boldsymbol{R}_v = \boldsymbol{V}\boldsymbol{R}_g\boldsymbol{V}^H. \tag{23}$$

Accordingly, after taking the hybrid structure and arbitrary array into consideration, the received signal in (6) can be expressed as

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{W}_t\boldsymbol{C}\boldsymbol{V}\boldsymbol{g}_t + \boldsymbol{W}_t\boldsymbol{n}_t\\ &= \boldsymbol{\Phi}_t\boldsymbol{d}_t + \boldsymbol{W}_t\boldsymbol{n}_t,\end{aligned} \tag{24}$$

where $\boldsymbol{\Phi}_t = \boldsymbol{W}_t\boldsymbol{C}$ denotes the equivalent sensing matrix.

As the Bessel matrix $\boldsymbol{C}$ depends solely on the known array geometry, $\boldsymbol{R}_h$ in (22) can be estimated as long as $\boldsymbol{R}_v$ is retrieved. Next, we need to figure out how to estimate $\boldsymbol{R}_v$ from the collected $\{\boldsymbol{y}_t\}_t$.

B. Channel Covariance Estimation

In this subsection, to estimate the virtual channel covariance $\boldsymbol{R}_v$, and hence also the actual channel covariance $\boldsymbol{R}_h$, we develop two super-resolution CCE techniques for the MMV case with multiple snapshots. One is the CCE via LRSCR, and the other is the CCE via ANM. The former CCE method is designed for the static channel sensing system only, whereas the latter CCE method is applicable to both the static and the dynamic hybrid hardware structures.

1) CCE via LRSCR: Suppose the path gains of the fading channels are uncorrelated with each other. Then, $\boldsymbol{R}_v$ in (23) not only presents the low rankness because of sparse scattering environments, but also is a well-structured PSD Hermitian Toeplitz matrix due to the Vandermonde structure of the virtual uniform array. In this sense, the LRSCR technique (13) can be used to implement these features of $\boldsymbol{R}_v$ in the virtual antenna domain.

Assuming now a static channel sensing system with $\boldsymbol{W}_t = \boldsymbol{W}$, and $\boldsymbol{\Phi}_t = \boldsymbol{W}\boldsymbol{C} = \boldsymbol{\Phi}, \forall t$, and substituting (22) into (9), $\boldsymbol{R}_y$ can be rewritten as

$$\begin{aligned}\boldsymbol{R}_y &= \boldsymbol{W}\boldsymbol{C}\boldsymbol{R}_v\boldsymbol{C}^H\boldsymbol{W}^H + \sigma^2\boldsymbol{W}\boldsymbol{W}^H\\ &= \boldsymbol{\Phi}\boldsymbol{R}_v\boldsymbol{\Phi}^H + \sigma^2\boldsymbol{W}\boldsymbol{W}^H,\end{aligned} \tag{25}$$

which is a function of $\boldsymbol{R}_v$.

Accordingly, by imposing the Hermitian Toeplitz structure on $\boldsymbol{R}_v$ in terms of $\boldsymbol{R}_v = \mathrm{T}(\boldsymbol{u}_v)$, we can describe the LRSCR-based formulation for the estimation of $\boldsymbol{R}_v$ from the sample covariance $\hat{\boldsymbol{R}}_y$ as

$$\begin{aligned}\hat{\boldsymbol{R}}_v = \ &\arg\min_{\mathrm{T}(\boldsymbol{u}_v)} \mathrm{tr}\left(\mathrm{T}(\boldsymbol{u}_v)\right) + \frac{\tau}{2}\left\|\hat{\boldsymbol{R}}_y - \boldsymbol{\Phi}\mathrm{T}(\boldsymbol{u}_v)\boldsymbol{\Phi}^H\right\|_F^2\\ \text{s.t.} \quad &\mathrm{T}(\boldsymbol{u}_v) \succeq 0.\end{aligned} \tag{26}$$

Then, given $\hat{\boldsymbol{R}}_v$ estimated from (26), we can finally obtain $\hat{\boldsymbol{R}}_h$ via (22), with known $\boldsymbol{C}$.

Further, to solve (26) with lower computational complexity compared with using off-the-shelf SDP solvers such as [49], we will develop a fast algorithm via ADMM later in Section IV-B.

2) CCE via Dynamic-ANM: Alternatively, let $\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_T]$ collect the virtual channels defined in (21) from different time slots. Then, an atom set can be defined in the virtual antenna domain as

$$\mathcal{A}' = \left\{\boldsymbol{v}(f)\boldsymbol{q}^H \,\middle|\, f \in (-\tfrac{1}{2}, \tfrac{1}{2}], \boldsymbol{q} \in \mathbb{C}^T, \|\boldsymbol{q}\|_2 = 1\right\}, \tag{27}$$

where $\boldsymbol{v}(f) \in \mathbb{C}^{2I+1}$ with its components being $e^{j2\pi i f}$, $i = -I, \ldots, I$. Obviously, $\boldsymbol{D}$ is a linear combination of the atoms from the set $\mathcal{A}'$. In this sense, the ANM technique can be used to exploit the low-rank and Vandermonde features of $\boldsymbol{D}$.

Then, with the received signals $\{\boldsymbol{y}_t\}_t$ in the matrix-form $\boldsymbol{Y}$, we produce the following ANM formulation for the static scenario:

$$\hat{\boldsymbol{D}} = \arg\min_{\boldsymbol{D}} \|\boldsymbol{D}\|_{\mathcal{A}'} + \frac{\tau}{2}\|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{D}\|_F^2. \tag{28}$$

Next, we consider a time-varying $\boldsymbol{\Phi}_t$ in the dynamic hybrid structure case. Since the ANM term in the objective function can be maintained, we now only have to formulate the residual error tolerance snapshot by snapshot. Hence, the dynamic version of (28) can be rewritten as

$$\hat{\boldsymbol{D}} = \arg\min_{\boldsymbol{D}} \|\boldsymbol{D}\|_{\mathcal{A}'} + \frac{\tau}{2}\sum_{t=1}^{T}\|\boldsymbol{y}_t - \boldsymbol{\Phi}_t\boldsymbol{d}_t\|_F^2. \tag{29}$$

Further, replacing the atomic norm in (29) by its SDP formulation [44], [45], we have

$$\begin{aligned}&(\hat{\boldsymbol{D}}, \mathrm{T}(\hat{\boldsymbol{u}}_d), \hat{\boldsymbol{Q}}) =\\ &\arg\min_{\boldsymbol{D},\mathrm{T}(\boldsymbol{u}_d),\boldsymbol{Q}} \frac{1}{2\sqrt{2I+1}}\left(\mathrm{tr}\left(\mathrm{T}\left(\boldsymbol{u}_d\right)\right) + \mathrm{tr}\left(\boldsymbol{Q}\right)\right)\\ &\quad + \frac{\tau}{2}\sum_{t=1}^{T}\|\boldsymbol{y}_t - \boldsymbol{\Phi}_t\boldsymbol{d}_t\|_F^2, \text{s.t.} \begin{bmatrix}\mathrm{T}(\boldsymbol{u}_d) & \boldsymbol{D}\\ \boldsymbol{D}^H & \boldsymbol{Q}\end{bmatrix} \succeq 0.\end{aligned} \tag{30}$$

For solving the problem (30), an ADMM-based fast algorithm can also be designed and iteratively implemented in a similar way as for (26).

In addition, according to the atomic norm definition and its SDP formulation, the reconstructed $\mathrm{T}(\hat{\boldsymbol{u}}_d)$ from (30) can be expressed by a Vandermonde decomposition as

$$\mathrm{T}(\hat{\boldsymbol{u}}_d) = \hat{\boldsymbol{V}}\left(T\hat{\boldsymbol{R}}_g\right)^{\frac{1}{2}}\hat{\boldsymbol{V}}^H. \tag{31}$$

This Vandermonde decomposition can be computed efficiently via root finding or by solving a generalized eigenvalue problem [50].

Finally, using the estimated $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{R}}_g$ from (31), we can obtain $\hat{\boldsymbol{R}}_v$ via (23) and then $\hat{\boldsymbol{R}}_h$ via (22) accordingly.

Remark 3: The ANM-based method achieves better estimation performance than the LRSCR-based method especially in the small number of snapshots, which will be evaluated and discussed later in Section V-A. On the other hand, the LRSCR formulation (26) has the merit of keeping the fixed problem size regardless of the number of snapshots. However, in the ANM formulation, the size of the PSD constraint in (30) grows as the number of snapshots increases. In addition, it is worth noting that CCE via LRSCR can be done in blind mode from data symbols, since it can work as long as $\hat{\boldsymbol{R}}_y$ is available to (26).

C. Two-Step Instantaneous CSI Estimation

As has been mentioned in Section II-A, the spatially sparse mmWave channel $\boldsymbol{h}$ is fully determined by parameters: $\boldsymbol{\theta}$ and $\boldsymbol{g}$. Since the path angles depend only on the relative positions of the BS, the MS, and the scatterers, $\boldsymbol{\theta}$ varies slowly, but can take a long time to acquire accurately from $\hat{\boldsymbol{R}}_h$ according to statistical inference techniques for angle estimation [28], [29], [50]. In contrast, the path gains $\boldsymbol{g}$ are easy to acquire given $\hat{\boldsymbol{\theta}}$, but vary frequently. Therefore, given the estimated $\hat{\boldsymbol{R}}_h$, in the second stage, we design a two-step CSI estimation scheme for block transmission. Angle estimates directly result from the CCE output, which stay unchanged for multiple blocks until they change. Given $\hat{\boldsymbol{\theta}}$, path gains are estimated at each block, followed by CSI-based data transmission.

1) Angle Estimation: Thanks to the specific matrix structures presented by the second-order statistics of mmWave channels, such as the low rankness due to the sparse scattering propagation and the PSD Toeplitz structure imposed by the SDP formulation through either (26) or (30), the Vandermonde decomposition can be applied to estimate the angles.

- If the LRSCR technique is used for CCE in Section III-B1, the recovered $\hat{\boldsymbol{R}}_v$ from (26) is a low-rank PSD Toeplitz matrix. According to the Vandermonde decomposition lemma [51], $\hat{\boldsymbol{R}}_v$ can be uniquely expressed as

$$\hat{\boldsymbol{R}}_v = \hat{\boldsymbol{V}}\hat{\boldsymbol{R}}_g\hat{\boldsymbol{V}}^H. \quad (32)$$

  Since the virtual uniform array geometry is solely decided by the angles as in (19), $\hat{\boldsymbol{\theta}}$ can be directly extracted from $\hat{\boldsymbol{V}}$ according to (17b).
- For the CCE based on the Dynamic-ANM technique as developed in Section III-B2, the Vandermonde decomposition can be carried out as in (31). Although the definition of the atom set in (27) leads to a diagonal matrix in the form of $(T\hat{\boldsymbol{R}}_g)^{\frac{1}{2}}$ in (31) that is different from $\hat{\boldsymbol{R}}_g$ in (32), the common Vandermonde structure of $\hat{\boldsymbol{V}}$ still leads to the same estimation results for $\hat{\boldsymbol{\theta}}$ via (17b).

2) Path Gain Estimation: Given the obtained angular information $\hat{\boldsymbol{\theta}}$ from Section III-C1, next we need to estimate the path gains $\boldsymbol{g}$ in a timely fashion as shown in Fig. 2. To this end, we first form the array matrix $\boldsymbol{A}$ via (2). Then, following the principle of a matched filter, we tune $\boldsymbol{W} = \boldsymbol{A}^H$ for beamforming. Noteworthily, with the obtained $\hat{\boldsymbol{\theta}}$ from angle estimation, the $K \times N$ matched filter based $\boldsymbol{W}$ applied now for path gain estimation is different from the $M \times N$ random phase shifter based $\boldsymbol{W}$ used earlier for CCE when $\hat{\boldsymbol{\theta}}$ is unknown. As a result, the estimation of $\boldsymbol{g}$ can be expressed as a least squares formulation

$$\hat{\boldsymbol{g}} = \arg\min_{\boldsymbol{g}} \|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{A}\boldsymbol{g}\|_2 = (\boldsymbol{W}\boldsymbol{A})^{\dagger}\boldsymbol{y}. \quad (33)$$

IV. Discussion of Related Issues

In this section, we provide detailed discussions on some specific issues related to the proposed solutions. We first provide the theoretical results in terms of fundamental limits for the proposed super-resolution channel estimation in hybrid mmWave massive MIMO with arbitrary arrays. Then, we design a fast algorithm via ADMM to rapidly implement the super-resolution estimation in lieu of invoking the high-computational SDP. In addition, we study the spurious-peak issue of the Jacobi-Anger approximation as a side effect specific to non-ULAs, and provide an effective way to solve this problem. Finally, we extend the work to the multiple-antenna MS case, which is developed based on an efficient 2D gridless CS approach.

A. Analysis of Fundamental Limits

To get an approximation with certain precision, (15) offers a lower bound on the choice of the maximum mode order I of the Jacobi-Anger expansion. From the view of the Jacobi-Anger approximation, the larger I is, the higher the accuracy the approximation can achieve. On the other hand, from the view of sparse channel estimation, we need to retrieve the sparse virtual channel $\{\boldsymbol{d}_t\}_t$ from the compressed measurements $\{\boldsymbol{y}_t\}_t$ with a high probability. For a given hybrid mmWave massive MIMO system, the dimension of $\boldsymbol{y}_t$ is fixed and known as M. To ensure the proposed gridless CS based methods are feasible and effective, it is necessary to study the minimum number of RF chains M_{min} required by our approaches given the approximation and hybrid structure. According to Theorem III.4 in [52], to guarantee successful reconstruction of channel covariance with high probability, M_{min} can be expressed as a function of I and K as

$$M_{\min} = \alpha K \log(2I+1), \quad (34)$$

where α is a numerical constant. Note that although the sensing matrix $\boldsymbol{\Phi}_t$ in [52] is assumed to be an i.i.d. random Gaussian matrix, it is reasonable to relax (34) to accommodate the analysis in this paper when $\boldsymbol{\Phi}_t$ is modeled as a random phase shifter based $\boldsymbol{W}_t$ multiplied by a Bessel matrix $\boldsymbol{C}$. Then, replacing M_{min} in (34) by M with $M > M_{min}$, we can obtain an upper bound on I. Combining with (15), we have

$$\frac{2\pi}{\lambda} r_{\max} < I < \frac{1}{2}\left(e^{\frac{M}{\alpha K}} - 1\right). \quad (35)$$

It is worth noting that (35) actually reflects the tradeoff between the approximation accuracy to combat the imperfection of the array geometry and the hardware cost required for sparse channel estimation, which thus sheds light on the choices of I and M in practice.

B. Fast Implementation via ADMM

To avoid the high computational complexity of the SDP-based solutions, we develop a fast iterative algorithm via ADMM. Next, we mainly discuss the solution for the LRSCR formulation

(26) and omit that for the ANM case (30).[4] To apply ADMM [42], we reformulate (26) as

$$\hat{\boldsymbol{R}}_v = \arg\min_{\mathrm{T}(\boldsymbol{u}_v)} \mathrm{tr}\left(\mathrm{T}(\boldsymbol{u}_v)\right) + \frac{\tau}{2}\left\|\hat{\boldsymbol{R}}_y - \boldsymbol{\Phi}\mathrm{T}(\boldsymbol{u}_v)\boldsymbol{\Phi}^H\right\|_F^2$$
$$\text{s.t.} \quad \boldsymbol{U} = \mathrm{T}(\boldsymbol{u}_v), \boldsymbol{U} \succeq 0, \tag{36}$$

whose augmented Lagrangian can be expressed as

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{u}_v, \boldsymbol{U}, \boldsymbol{\Lambda})\\
&= \mathrm{tr}\left(\mathrm{T}(\boldsymbol{u}_v)\right) + \frac{\tau}{2}\left\|\hat{\boldsymbol{R}}_y - \boldsymbol{\Phi}\mathrm{T}(\boldsymbol{u}_v)\boldsymbol{\Phi}^H\right\|_F^2\\
&\quad + \langle \boldsymbol{\Lambda}, \boldsymbol{U} - \mathrm{T}(\boldsymbol{u}_v)\rangle + \frac{\rho}{2}\left\|\boldsymbol{U} - \mathrm{T}(\boldsymbol{u}_v)\right\|_F^2\\
&= \mathrm{tr}\left(\mathrm{T}(\boldsymbol{u}_v)\right) + \frac{\tau}{2}\left\|\hat{\boldsymbol{R}}_y - \boldsymbol{\Phi}\mathrm{T}(\boldsymbol{u}_v)\boldsymbol{\Phi}^H\right\|_F^2 - \frac{1}{2\rho}\left\|\boldsymbol{\Lambda}\right\|_F^2\\
&\quad + \frac{\rho}{2}\left\|\boldsymbol{U} - \mathrm{T}(\boldsymbol{u}_v) + \rho^{-1}\boldsymbol{\Lambda}\right\|_F^2,
\end{aligned} \tag{37}$$

where $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ are Hermitian matrices. Then the implementation of ADMM involves the following iterative updates:

$$\boldsymbol{u}_v^{l+1} = \arg\min_{\boldsymbol{u}_v} \mathcal{L}(\boldsymbol{u}_v, \boldsymbol{U}^l, \boldsymbol{\Lambda}^l); \tag{38}$$

$$\boldsymbol{U}^{l+1} = \arg\min_{\boldsymbol{U}\succeq 0} \mathcal{L}(\boldsymbol{u}_v^{l+1}, \boldsymbol{U}, \boldsymbol{\Lambda}^l); \tag{39}$$

$$\boldsymbol{\Lambda}^{l+1} = \boldsymbol{\Lambda}^l + \rho(\boldsymbol{U}^{l+1} - \mathrm{T}(\boldsymbol{u}_v^{l+1})), \tag{40}$$

where the superscript l denotes the l-th iteration update. In order to implement (38), we take the partial derivative of (37) with respect to $\boldsymbol{u}_v^*$ at the $(l+1)$-th iteration and force it equal to zero. After taking a series of derivations on $\frac{\partial}{\partial \boldsymbol{u}_v^*}\mathcal{L}(\boldsymbol{u}_v, \boldsymbol{U}^l, \boldsymbol{\Lambda}^l)\Big|_{\boldsymbol{u}_v = \boldsymbol{u}_v^{l+1}} = 0$, we obtain

$$\begin{aligned}
&\tau\mathcal{G}(\boldsymbol{\Phi}^H\boldsymbol{\Phi}\mathrm{T}(\boldsymbol{u}_v^{l+1})\boldsymbol{\Phi}^H\boldsymbol{\Phi}) + \rho\mathcal{G}(\mathrm{T}(\boldsymbol{u}_v^{l+1}))\\
&\quad = \tau\mathcal{G}(\boldsymbol{\Phi}^H\hat{\boldsymbol{R}}_y\boldsymbol{\Phi}) + \rho\mathcal{G}(\boldsymbol{U}^l + \rho^{-1}\boldsymbol{\Lambda}^l) - N_I\boldsymbol{e}_1,
\end{aligned} \tag{41}$$

where $N_I = 2I+1$ is the column (row) size of $\mathrm{T}(\boldsymbol{u}_v^{l+1})$, $\boldsymbol{e}_1$ is the length-N_I vector with only the first element being one, and $\boldsymbol{b} = \mathcal{G}(\boldsymbol{B}) \in \mathbb{C}^{N_I}$ is a mapping from a matrix to a vector where the n_I-th element of $\boldsymbol{b}$ is the sum of all the elements $\boldsymbol{B}_{i,j}$ in $\boldsymbol{B}$ satisfying $i - j + 1 = n_I$. Moreover, denote $\boldsymbol{M}$ as the matrix which satisfies $\boldsymbol{b} = \mathcal{G}(\boldsymbol{B}) = \boldsymbol{M}\mathrm{vec}(\boldsymbol{B})$ and $\boldsymbol{\beta}^l = \tau\mathcal{G}(\boldsymbol{\Phi}^H\hat{\boldsymbol{R}}_y\boldsymbol{\Phi}) + \rho\mathcal{G}(\boldsymbol{U}^l + \rho^{-1}\boldsymbol{\Lambda}^l)$, respectively.

Accordingly, we rewrite (41) as

$$\begin{aligned}
&\tau\boldsymbol{M}\mathrm{vec}(\boldsymbol{\Phi}^H\boldsymbol{\Phi}\mathrm{T}(\boldsymbol{u}_v^{l+1})\boldsymbol{\Phi}^H\boldsymbol{\Phi}) + \rho\boldsymbol{M}\mathrm{vec}(\mathrm{T}(\boldsymbol{u}_v^{l+1}))\\
&\quad = \boldsymbol{\beta}^l - N_I\boldsymbol{e}_1\\
&\quad \Leftrightarrow \left(\tau\boldsymbol{M}\left[(\boldsymbol{\Phi}^H\boldsymbol{\Phi})^T \otimes (\boldsymbol{\Phi}^H\boldsymbol{\Phi})\right] + \rho\boldsymbol{M}\right)\mathrm{vec}(\mathrm{T}(\boldsymbol{u}_v^{l+1}))\\
&\quad = \boldsymbol{\beta}^l - N_I\boldsymbol{e}_1\\
&\quad \Leftrightarrow \boldsymbol{\Pi}\begin{bmatrix}\boldsymbol{u}_{\mathrm{R}}^{l+1}\\ \boldsymbol{u}_{\mathrm{I}}^{l+1}\end{bmatrix} = \boldsymbol{\beta}^l - N_I\boldsymbol{e}_1,
\end{aligned} \tag{42}$$

[4]The design of an ADMM-based fast algorithm for (30) can be developed in a similar way as for (26). The differences are the two additional variables $\boldsymbol{D}$ and $\boldsymbol{Q}$ in (30), which can be easily updated by gradient descent in each iteration. Moreover, the algorithm implementation for (30) is actually simpler than that for (26). The reason is that the Toeplitz structured matrix $\mathrm{T}(\boldsymbol{u})$ is not included in the least squares term in the objective function in (30), which then simplifies the calculation of the partial derivative of the augmented Lagrangian $\mathcal{L}$ with respect to $\boldsymbol{u}^*$. As a result, compared with the LRSCR case, the update of $\boldsymbol{u}$ in each iteration becomes easier in the ANM case.

where $\boldsymbol{u}_v^{l+1} = \boldsymbol{u}_{\mathrm{R}}^{l+1} + j[0, (\boldsymbol{u}_{\mathrm{I}}^{l+1})^T]^T$. Since $\mathrm{T}(\boldsymbol{u}_v^{l+1})$ is only determined by the real and imaginary parts of $\boldsymbol{u}_v^{l+1}$ as $\boldsymbol{u}_{\mathrm{R}}^{l+1}$ and $\boldsymbol{u}_{\mathrm{I}}^{l+1}$, respectively, there exists a fixed matrix $\boldsymbol{\Gamma}$ satisfying $\mathrm{vec}(\mathrm{T}(\boldsymbol{u}_v^{l+1})) = \boldsymbol{\Gamma}[(\boldsymbol{u}_{\mathrm{R}}^{l+1})^T, (\boldsymbol{u}_{\mathrm{I}}^{l+1})^T]^T$. Moreover, $\boldsymbol{\Pi} = \left(\tau\boldsymbol{M}\left[(\boldsymbol{\Phi}^H\boldsymbol{\Phi})^T \otimes (\boldsymbol{\Phi}^H\boldsymbol{\Phi})\right] + \rho\boldsymbol{M}\right)\boldsymbol{\Gamma}$. Since $[(\boldsymbol{u}_{\mathrm{R}}^{l+1})^T, (\boldsymbol{u}_{\mathrm{I}}^{l+1})^T]^T \in \mathbb{R}^{2N_I-1}$, we can rewrite the N_I complex equations of (42) into $2N_I$ real equations as

$$\begin{bmatrix}\mathrm{Real}\{\boldsymbol{\Pi}\}\\ \mathrm{Imag}\{\boldsymbol{\Pi}\}\end{bmatrix}\begin{bmatrix}\boldsymbol{u}_{\mathrm{R}}^{l+1}\\ \boldsymbol{u}_{\mathrm{I}}^{l+1}\end{bmatrix} = \begin{bmatrix}\mathrm{Real}\{\boldsymbol{\beta}^l\} - N_I\boldsymbol{e}_1\\ \mathrm{Imag}\{\boldsymbol{\beta}^l\}\end{bmatrix}. \tag{43}$$

Hence, the update rule for $\boldsymbol{u}_v$ is given by

$$\begin{aligned}
\boldsymbol{u}_v^{l+1} &= \boldsymbol{u}_{\mathrm{R}}^{l+1} + j\begin{bmatrix}0\\ \boldsymbol{u}_{\mathrm{I}}^{l+1}\end{bmatrix}\\
\begin{bmatrix}\boldsymbol{u}_{\mathrm{R}}^{l+1}\\ \boldsymbol{u}_{\mathrm{I}}^{l+1}\end{bmatrix} &= \begin{bmatrix}\mathrm{Real}\{\boldsymbol{\Pi}\}\\ \mathrm{Imag}\{\boldsymbol{\Pi}\}\end{bmatrix}^{\dagger}\begin{bmatrix}\mathrm{Real}\{\boldsymbol{\beta}^l\} - N_I\boldsymbol{e}_1\\ \mathrm{Imag}\{\boldsymbol{\beta}^l\}\end{bmatrix}.
\end{aligned} \tag{44}$$

Let $\Xi^l = \mathrm{T}(\boldsymbol{u}_v^{l+1}) - \rho^{-1}\boldsymbol{\Lambda}^l = \boldsymbol{E}^l\boldsymbol{\Sigma}^l\boldsymbol{E}^{l^H}$ be its eigenvalue decomposition, then based on (37) and (39), we have the update of $\boldsymbol{U}$ at the $(l+1)$-th iteration as

$$\boldsymbol{U}^{l+1} = \boldsymbol{E}^l\boldsymbol{\Sigma}_+^l\boldsymbol{E}^{l^H}, \tag{45}$$

where $\boldsymbol{\Sigma}_+^l$ is obtained by letting all negative eigenvalues of $\boldsymbol{\Sigma}^l$ be zero.

The iterative algorithm will stop until both primal and dual residuals satisfy the pre-set tolerance level [42].

C. Specific Instance of Non-ULAs

Non-ULAs yield one case of arbitrary arrays. However, unlike general random distributed antennas, the antenna elements of non-ULAs are distributed along a line, which results in the Bessel matrix used by the Jacobi-Anger approximation being axial symmetric. This axial symmetric characteristic leads to a special issue in the implementation of the proposed Dynamic-ANM and LRSCR methods. Next, we discuss this specific instance in detail.

Suppose the locations of antenna elements of a non-ULA are formed as $(r_n, 0), n = 1, \ldots, N$ in polar coordinates. Then, (16) is rewritten as

$$[\boldsymbol{a}(\theta_k)]_n \approx \sum_{i=-I}^{I} j^i\mathrm{J}_i\left(2\pi\frac{r_n}{\lambda}\right)e^{j\theta_k i} = \boldsymbol{c}_n^T\boldsymbol{v}(\theta). \tag{46}$$

Based on the property of the Bessel function of the first kind that says

$$\mathrm{J}_{-i}(x) = (-1)^n\mathrm{J}_i(x), \quad \text{for} \quad \forall x > 0, \tag{47}$$

we have

$$\begin{aligned}
[\boldsymbol{c}_n]_{-i} &= j^{-i}\mathrm{J}_{-i}\left(2\pi\frac{r_n}{\lambda}\right)\\
&= j^i\mathrm{J}_i\left(2\pi\frac{r_n}{\lambda}\right) = [\boldsymbol{c}_n]_i.
\end{aligned} \tag{48}$$

From (48), the Bessel matrix $\boldsymbol{C}$ explicitly holds

$$\boldsymbol{C}\boldsymbol{I}_a = \boldsymbol{C}. \tag{49}$$

Accordingly, (18) can be expressed as

$$\begin{aligned}\boldsymbol{a}(\theta) &= \frac{1}{2}\left(\boldsymbol{C}+\boldsymbol{C}\right)\boldsymbol{v}(\theta) = \frac{1}{2}\left(\boldsymbol{C}+\boldsymbol{C}\boldsymbol{I}_a\right)\boldsymbol{v}(\theta)\\ &= \frac{1}{2}\boldsymbol{C}\left(\boldsymbol{v}(\theta)+\boldsymbol{I}_a\boldsymbol{v}(\theta)\right) = \frac{1}{2}\boldsymbol{C}\left(\boldsymbol{v}(\theta)+\boldsymbol{v}^*(\theta)\right)\\ &= \boldsymbol{C}\,\mathrm{Real}\left(\boldsymbol{v}(\theta)\right). \end{aligned} \tag{50}$$

Then, given (50), (21) can be rewritten as

$$\begin{aligned}\boldsymbol{h}_t &= \frac{1}{2}\boldsymbol{C}(\boldsymbol{V}+\boldsymbol{V}^*)\,\boldsymbol{g}_t\\ &= \boldsymbol{C}\left[\boldsymbol{V},\boldsymbol{V}^*\right]\left[\frac{\boldsymbol{g}_t^T}{2},\frac{\boldsymbol{g}_t^T}{2}\right]^T, \end{aligned} \tag{51}$$

which indicates

$$\boldsymbol{d}_t = \left[\boldsymbol{V},\boldsymbol{V}^*\right]\left[\frac{\boldsymbol{g}_t^T}{2},\frac{\boldsymbol{g}_t^T}{2}\right]^T. \tag{52}$$

This means that (52) turns out be an alternative possible solution of (30). Accordingly, different from (31), the estimated $\mathrm{T}(\boldsymbol{u}_d)$ leads to another Vandermonde decomposition as

$$\mathrm{T}(\boldsymbol{u}_d) = \frac{1}{2}\left(\left[\boldsymbol{V},\boldsymbol{V}^*\right]\mathrm{diag}\left((T\boldsymbol{R}_g)^{\frac{1}{2}},(T\boldsymbol{R}_g)^{\frac{1}{2}}\right)\left[\boldsymbol{V},\boldsymbol{V}^*\right]^H\right). \tag{53}$$

Further, since $\boldsymbol{R}_g$ is a diagonal matrix with positive elements, substituting (50) into (22), we have

$$\begin{aligned}\boldsymbol{R}_h &= \boldsymbol{C}\,\mathrm{Real}(\boldsymbol{V})\boldsymbol{R}_g\,\mathrm{Real}(\boldsymbol{V})^H\boldsymbol{C}^H\\ &= \boldsymbol{C}\,\mathrm{Real}(\boldsymbol{V}\boldsymbol{R}_g\boldsymbol{V}^H)\,\boldsymbol{C}^H\\ &= \boldsymbol{C}\,\mathrm{Real}(\boldsymbol{R}_v)\,\boldsymbol{C}^H. \end{aligned} \tag{54}$$

Noticeably, instead of (32), $\mathrm{Real}(\boldsymbol{R}_v)$ is decomposed as

$$\begin{aligned}\mathrm{Real}(\boldsymbol{R}_v) &= \mathrm{Real}(\boldsymbol{V}\boldsymbol{R}_g\boldsymbol{V}^H)\\ &= \frac{1}{2}\left(\boldsymbol{V}\boldsymbol{R}_g\boldsymbol{V}^H+\boldsymbol{V}^*\boldsymbol{R}_g(\boldsymbol{V}^H)^*\right)\\ &= \frac{1}{2}\left([\boldsymbol{V},\boldsymbol{V}^*]\,\mathrm{diag}(\boldsymbol{R}_g,\boldsymbol{R}_g)[\boldsymbol{V},\boldsymbol{V}^*]^H\right). \end{aligned} \tag{55}$$

As a result, beside the estimation of the true AoAs as $\{\hat{\theta}_k\}$, spurious results are also generated as $\{-\hat{\theta}_k\}$. According to the above analysis that there exist multiple solutions, we need to use the prior knowledge that the feasible domain of AoAs is $[0,\pi)$. Then we can simply remove the spurious results appearing in $(-\pi,0)$.

D. Extension to the Multiple-Antenna MS Case

In this subsection, we extend the work to the case where the MS also has a hybrid architecture with multiple RF chains and antennas in arbitrary arrays. Now, the 1D uplink channel model for the case of the single-antenna at the MS in (1) is extended to a 2D uplink channel model

$$\boldsymbol{H} = \sum_{k=1}^{K} g_k\boldsymbol{a}_{\mathrm{BS}}\left(\theta_{\mathrm{BS},k}\right)\boldsymbol{a}_{\mathrm{MS}}^H\left(\theta_{\mathrm{MS},k}\right), \tag{56}$$

where $\theta_{\mathrm{MS},k}$ and $\theta_{\mathrm{BS},k}$ denote the continuous-valued AoD and AoA of the k-th path at the MS as transmitter and at the BS as receiver, respectively. Accordingly, the vectors $\boldsymbol{a}_{\mathrm{MS}}\left(\theta_{\mathrm{MS},k}\right)$ and $\boldsymbol{a}_{\mathrm{BS}}\left(\theta_{\mathrm{BS},k}\right)$ represent the array manifold vectors corresponding to the k-th path for the N_{MS}-antenna and N_{BS}-antenna arrays, respectively, which both have components in the form of (2). This 2D channel model is general enough to subsume the multi-user case where each column of $\boldsymbol{H}$ corresponds to one MS (user) with a single antenna.

Given the hybrid structures and the arbitrary arrays employed at both the BS and the MS sides, the received signal can be expressed as

$$\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{A}_{\mathrm{BS}}\,\mathrm{diag}(\boldsymbol{g})\,\boldsymbol{A}_{\mathrm{MS}}^H\boldsymbol{F}+\boldsymbol{W}\boldsymbol{N}, \tag{57}$$

where $\boldsymbol{F}$ denotes the hybrid precoding matrix used at the transmitter side.

Applying the array manifold separation approach described in (20), (57) can be rewritten as

$$\begin{aligned}\boldsymbol{Y} &= \boldsymbol{W}\boldsymbol{C}_{\mathrm{BS}}\boldsymbol{V}_{\mathrm{BS}}\,\mathrm{diag}(\boldsymbol{g})\,\boldsymbol{V}_{\mathrm{MS}}^H\boldsymbol{C}_{\mathrm{MS}}^H\boldsymbol{F}+\boldsymbol{W}\boldsymbol{N},\\ &= \boldsymbol{W}\boldsymbol{C}_{\mathrm{BS}}\,\boldsymbol{\Psi}\,\boldsymbol{C}_{\mathrm{MS}}^H\boldsymbol{F}+\boldsymbol{W}\boldsymbol{N}, \end{aligned} \tag{58}$$

where $\boldsymbol{\Psi} = \boldsymbol{V}_{\mathrm{BS}}\,\mathrm{diag}(\boldsymbol{g})\,\boldsymbol{V}_{\mathrm{MS}}^H$ denotes the virtual 2D channel that presents the 2D Vandermonde structure in the virtual uniform antenna domain.

To estimate the 2D channel through gridless CS, a straightforward way is to vectorize the 2D formulations and then to cast the 2D Vandermonde structure into a vectorized SDP formulation via a two-level Toeplitz structured matrix, a.k.a., vectorization-based ANM (V-ANM) [33], [53], [54]. However, the V-ANM leads to a high computational complexity on the order of $\mathcal{O}(N_{\mathrm{BS}}^{3.5}N_{\mathrm{MS}}^{3.5})$ [55], because of the huge problem scale resulting from the vectorization operation.

To solve this problem of V-ANM, we develop an efficient 2D channel estimation at much lower computational cost, by using a decoupled-ANM (D-ANM) technique [55], [56]. Different from the V-ANM, we introduce a matrix-form atom set $\mathcal{A}_{\mathrm{M}}$ as

$$\mathcal{A}_{\mathrm{M}} = \left\{\boldsymbol{v}_{\mathrm{BS}}(f_{\mathrm{BS}})\,\boldsymbol{v}_{\mathrm{MS}}^H(f_{\mathrm{MS}})\,\middle|\, f_{\mathrm{BS}}\in\left(-\tfrac{1}{2},\tfrac{1}{2}\right],\, f_{\mathrm{MS}}\in\left(-\tfrac{1}{2},\tfrac{1}{2}\right]\right\}, \tag{59}$$

which naturally results in a matrix-form atomic norm as

$$\|\boldsymbol{\Psi}\|_{\mathcal{A}_{\mathrm{M}}} = \inf\left\{\sum_l |g_l|\,\middle|\,\boldsymbol{\Psi} = \sum_l g_l\,\boldsymbol{v}_{\mathrm{BS}}(f_{\mathrm{BS},l})\,\boldsymbol{v}_{\mathrm{MS}}^H(f_{\mathrm{MS},l})\right\}. \tag{60}$$

Then, given $\boldsymbol{Y}$ from (58), the virtual 2D channel $\boldsymbol{\Psi}$ can be reconstructed via the following decoupled SDP formulation

$$\begin{aligned}&(\hat{\boldsymbol{\Psi}},\mathrm{T}(\hat{\boldsymbol{u}}_{\mathrm{BS}}),\mathrm{T}(\hat{\boldsymbol{u}}_{\mathrm{MS}})) =\\ &\quad\arg\min_{\hat{\boldsymbol{\Psi}},\mathrm{T}(\boldsymbol{u}_{\mathrm{BS}}),\mathrm{T}(\boldsymbol{u}_{\mathrm{MS}})}\frac{1}{2\sqrt{N_{\mathrm{BS}}N_{\mathrm{MS}}}}\left(\mathrm{tr}(\mathrm{T}(\boldsymbol{u}_{\mathrm{BS}}))+\mathrm{tr}(\mathrm{T}(\boldsymbol{u}_{\mathrm{MS}}))\right)\\ &\quad+\frac{\tau}{2}\left\|\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{C}_{\mathrm{BS}}\boldsymbol{\Psi}\boldsymbol{C}_{\mathrm{MS}}^H\boldsymbol{F}\right\|_F^2,\ \mathrm{s.t.}\begin{bmatrix}\mathrm{T}(\boldsymbol{u}_{\mathrm{BS}}) & \boldsymbol{\Psi}\\ \boldsymbol{\Psi}^H & \mathrm{T}(\boldsymbol{u}_{\mathrm{MS}})\end{bmatrix}\succeq 0. \end{aligned} \tag{61}$$

Noticeably, since the PSD constraint in (61) is of size $(N_{\mathrm{BS}}+N_{\mathrm{MS}})\times(N_{\mathrm{BS}}+N_{\mathrm{MS}})$, the D-ANM allows a reduced computational complexity on the order of $\mathcal{O}\left((N_{\mathrm{BS}}+N_{\mathrm{MS}})^{3.5}\right)$

[55], which is much smaller than that of the V-ANM with large arrays.

V. Numerical Results

This section presents numerical results to evaluate the channel estimation performance achieved by the proposed methods for arbitrary arrays and a hybrid precoding structure. In each Monte Carlo simulation, the random path angles are generated uniformly from $[0^\circ, 180^\circ)$. The existing channel covariance estimation methods via covariance orthogonal matching pursuit (COMP) and Dynamic-COMP (DCOMP) and the existing instantaneous CSI estimation methods via simultaneous orthogonal matching pursuit (SOMP) and Dynamic-SOMP (DSOMP) are also simulated as benchmarks for performance comparison [21], where a predefined grid of size 360 is employed for the grid-based CS technique that leads to an angle-resolution of 0.5°.

A. Channel Estimation Performance

First, we test different channel estimation approaches for an arbitrary planar array, where the antennas are randomly placed in a semicircular area with a radius of $r_{max} = 4\lambda$. The channel estimation performance[5] is evaluated in terms of the normalized mean squared error (NMSE) for the CCE as $\mathbb{E}\{\|\boldsymbol{R}_h - \hat{\boldsymbol{R}}_h\|_F^2\}/\mathbb{E}\{\|\boldsymbol{R}_h\|_F^2\}$ and the NMSE for the CSI estimation as $\mathbb{E}\{\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|_2^2\}/\mathbb{E}\{\|\boldsymbol{h}\|_2^2\}$, respectively. In simulations, our LRSCR-based methods as described in Section III-B1 and the COMP-based and SOMP-based methods in [21] are tested on the fixed combining matrix $\boldsymbol{W}$, while our Dynamic-ANM-based methods as developed in Section III-B2 and the DCOMP-based and the DSOMP-based methods in [21] are applied with the dynamic combining matrix $\boldsymbol{W}_t$. In our proposed two-step CSI estimation scheme, we employ a Vandermonde decomposition in the form of either (31) or (32) to retrieve the path angles in the first step and to estimate the path gains via least squares in (33) in the second step.

Fig. 3 and Fig. 4 present the NMSE of CCE and the NMSE of CSI versus the number of snapshots, respectively. The comparison of the curves indicates that our proposed methods based on LRSCR and Dynamic-ANM outperform the existing methods based on grid-based CS. While the dynamic configurations usually provide a higher sensing accuracy than the fixed counterparts, our LRSCR can even work better than DSOMP as the number of snapshots increases. This is because our methods utilize not only the sparsity of mmWave channels but also the structural feature of the array geometry. Further, our Dynamic-ANM method always achieves the best performance especially given a small number of snapshots, because ANM can efficiently utilize such structures directly from the collected samples. On the other hand, since LRSCR is a statistics-based design, it requires a sufficient number of snapshots for computing an accurate sample covariance. When the number of snapshots becomes small e.g. less than 20, the finite-sample effect ruins the Toeplitz structure presented in the ideal covariance matrix, which thus degrades the performance of LRSCR. Fig. 5 and Fig. 6 present the NMSE performance of these approaches for different numbers of RF chains, which show the same trends as in Fig. 3 and Fig. 4.

[5]In this work, we focus on evaluating the estimation accuracy achieved at the training stage, while the evaluation on the achievable spectral efficiency subject to channel estimation errors can be found in [22], [33].

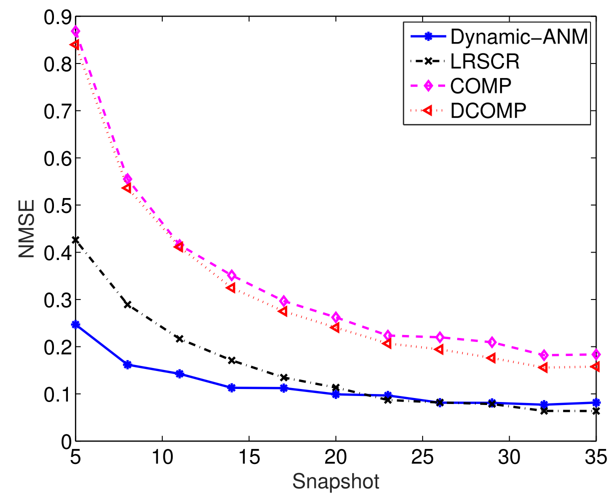


Fig. 3. NMSE of CCE versus T for LRSCR, Dynamic-ANM, COMP and DCOMP, when $N = 64, M = 16, I = 35, K = 4$, SNR $= 10$ dB.

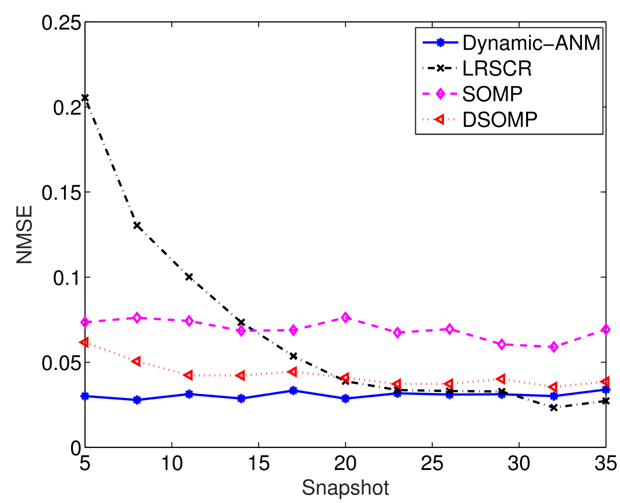


Fig. 4. NMSE of CSI estimation versus T for LRSCR, Dynamic-ANM, SOMP and DSOMP, when $N = 64, M = 16, I = 35, K = 4$, SNR $= 10$ dB.

B. Non-ULAs

Next, we study the non-ULA as a special case of arbitrary arrays. Fig. 7 shows that the NMSE of the instantaneous CSI estimation via our Dynamic-ANM method can be much smaller than that of the existing ANM without array manifold separation [33], which demonstrates the necessity of the Jacobi-Anger approximation for an imperfect array geometry. Besides, it also indicates that our method proposed for arbitrary arrays can approach the performance of the ideal case of a same-size ULA

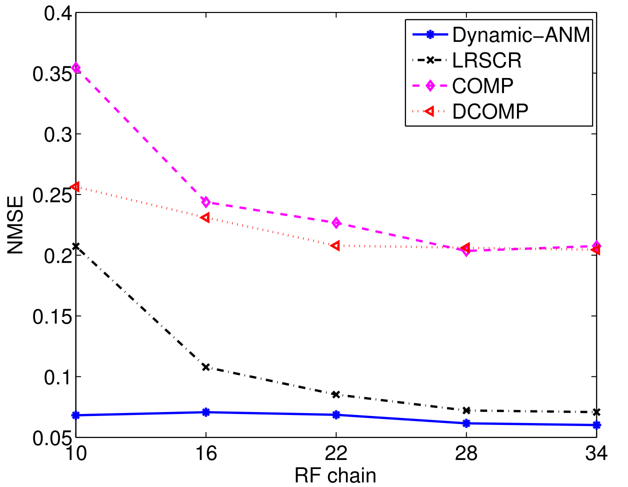


Fig. 5. NMSE of CCE versus M for LRSCR, Dynamic-ANM, COMP and DCOMP, when $N = 64, T = 20, I = 35, K = 4$, SNR = 10 dB.

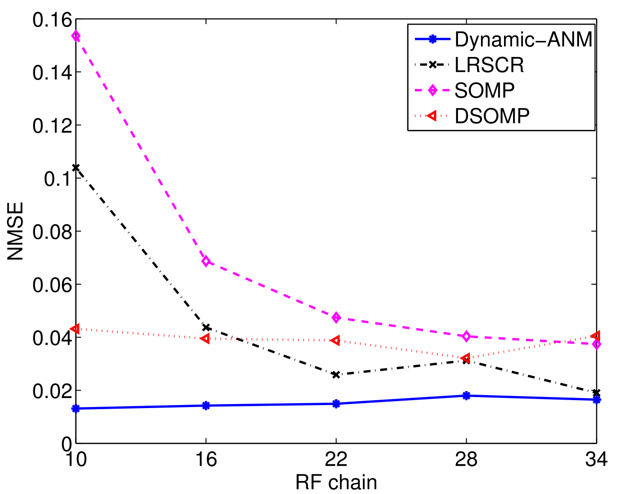


Fig. 6. NMSE of CSI estimation versus M for LRSCR, Dynamic-ANM, SOMP and DSOMP, when $N = 64, T = 20, I = 35, K = 4$, SNR = 10 dB.

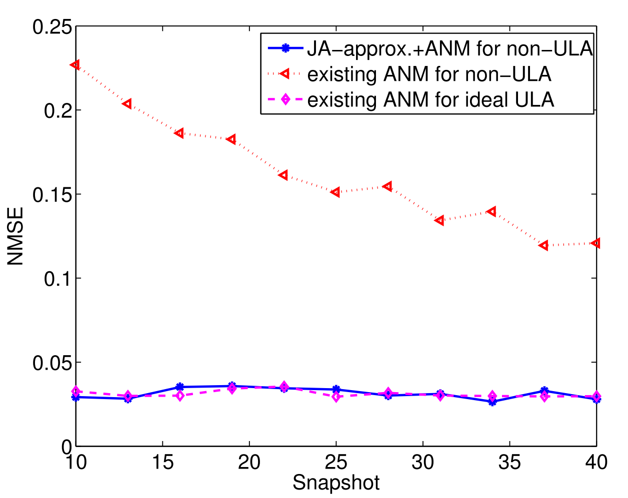


Fig. 7. NMSE of CSI estimation versus T for Dynamic-ANM, existing ANM for non-ULA, and existing ANM for the same aperture size ideal ULA, when $N = 16, M = 8, I = 60, K = 4$, SNR = 10 dB.

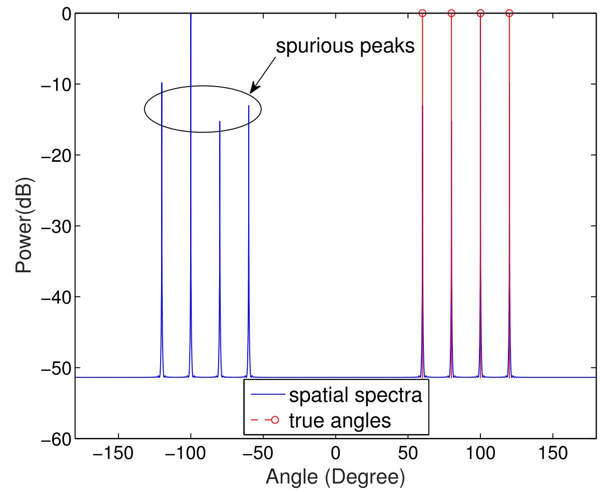


Fig. 8. Spatial spectra of LRSCR for non-ULA.

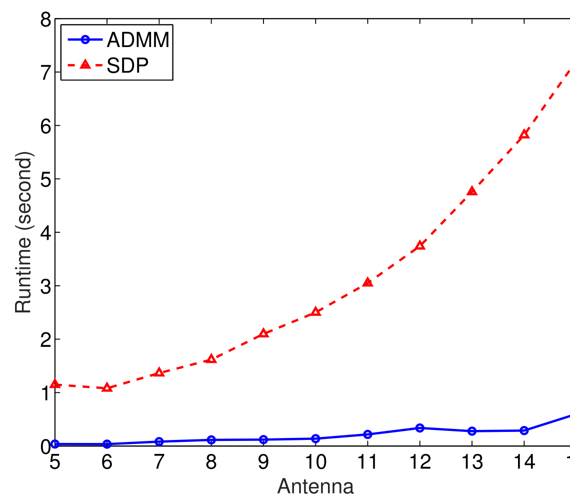


Fig. 9. Runtime versus N for ADMM and SDP implementations.

as the benchmark for the best performance that the proposed techniques can achieve.

Further, we study the side effect of the Jacobi-Anger approximation in terms of the spurious peaks generated in the case of non-ULAs. Fig. 8 shows the spatial spectra result of LRSCR, where the peaks on the right indicate the true angles. Meanwhile, the spurious peaks appear at the symmetric angles, which can be simply removed given the prior knowledge of the angle range as $[0°, 180°)$.

C. Computational Complexity

In addition, we test the computational cost of the proposed fast algorithm via ADMM, compared with the SDP-based solver [49]. By counting the runtime[6] versus the number of antennas, Fig. 9 clearly indicates that as the number of antennas increases the slope of the runtime curve of the ADMM-based solution is much smaller than that of the SDP counterpart. Obviously,

[6]All simulations are conducted in Matlab 2013b on a computer with a 4-core Intel i7-6500U 2.50 GHz CPU and 8 GB memory.

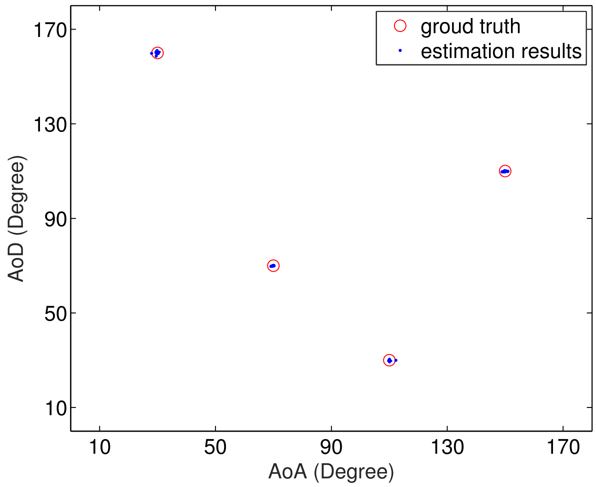


Fig. 10. 2D path angle estimation via D-ANM in a multiple-antenna MS case, when $N_{\text{MS}} = N_{\text{BS}} = 32, M_{\text{MS}} = M_{\text{BS}} = 16, I_{\text{MS}} = I_{\text{BS}} = 35, K = 4, \text{SNR} = 10$ dB.

the proposed fast implementation via ADMM has low computational complexity and is hence well suited for large arrays in practice.

D. 2D Path Angle Estimation

Last but not least, we extend our proposed work to the multiple-antenna MS case where both the BS and the MS are equipped with arbitrary multiple antennas, which results in a 2D path angle estimation scenario. We carry out 50 trials and show the estimation results in Fig. 10 where our proposed D-ANM method based on the Jacobi-Anger approximation can precisely retrieve both the AoDs and the AoAs of the sparse scattering paths, which indicates the high performance of our proposed super-resolution 2D channel estimation technique.

VI. Conclusions

Recognizing the imperfect geometry of arbitrary arrays and the hardware constraint of hybrid mmWave massive MIMO systems, this paper has proposed a new super-resolution channel estimation framework that achieves the benefits of the array manifold separation techniques and the structure-based optimization approaches. Through the Jacobi-Anger approximation, the Vandermonde structure is effected in the virtual antenna domain of arbitrary arrays, which enables super-resolution channel estimation based on gridless CS techniques to obtain a high performance at low training costs. In particular, we develop two channel covariance estimation approaches via LRSCR and Dynamic-ANM. Further, considering that angles change relatively slower than path gains, we design a two-step CSI estimation scheme which separates long-term angle estimation from frequent path gain estimation. Theoretical results are provided to investigate the fundamental limits of the proposed super-solution technique in terms of the minimum number of RF chains required for channel estimation and the bounds on the mode order for the Jacobi-Anger approximation. To reduce the computational complexity of the structure-based optimization via SDP, an iterative algorithm is developed through ADMM for fast implementation. To combat the side effect due to the Jacobi-Anger approximation occurring in non-ULAs, we provide a mechanism to efficiently remove the spurious peaks. Finally, we extend our work to the 2D-angle scenarios, where both the BS and the MS are equipped with multiple antennas.

References

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, “MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?” *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[3] F. Rusek *et al.*, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter-wave cellular wireless networks: Potentials and challenges,” *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[5] Y. Wang and Z. Tian, “Big data in 5G,” in *Encyclopedia of Wireless Networks*. New York, NY, USA: Springer, 2018. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-32903-1_58-1

[6] A. F. Molisch *et al.*, “Hybrid beamforming for massive MIMO: A survey,” *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[7] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, “MIMO precoding and combining solutions for millimeter-wave systems,” *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[8] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels,” *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[9] P. Schniter and A. Sayeed, “Channel estimation and precoder design for millimeter-wave communications: The sparse way,” in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 273–277.

[10] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[11] J. Lee, G.-T. Gil, and Y. H. Lee, “Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications,” *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

[12] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, “Channel estimation and hybrid combining for mmwave: Phase shifters or switches?” in *Proc. Inf. Theory Appl. Workshop*, Feb. 2015, pp. 90–97.

[13] Y. Wang, Z. Tian, S. Feng, and P. Zhang, “A fast channel estimation approach for millimeter-wave massive MIMO systems,” in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2016, pp. 1413–1417.

[14] Z. Gao, C. Hu, L. Dai, and Z. Wang, “Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels,” *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

[15] H. Zhang, S. Venkateswaran, and U. Madhow, “Channel modeling and MIMO capacity for outdoor millimeter wave links,” in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2010, pp. 1–6.

[16] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, “Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications,” *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[17] M. Shafi *et al.*, “Microwave vs. millimeter-wave propagation channels: Key differences and impact on 5G cellular systems,” *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 14–20, Dec. 2018.

[18] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2014.

[19] A. M. Sayeed, “Deconstructing multiantenna fading channels,” *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[20] Y. Wang, Z. Tian, S. Feng, and P. Zhang, “Efficient channel statistics estimation for millimeter-wave MIMO systems,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2016, pp. 3411–3415.

[21] S. Park and R. W. Heath, "Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing-based approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8047–8062, Dec. 2018.

[22] S. Park, J. Park, A. Yazdan, and R. W. Heath, "Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3818–3832, Jul. 2017.

[23] Z. Li, S. Han, and A. F. Molisch, "Optimizing channel-statistics-based analog beamforming for millimeter-wave multi-user massive MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4288–4303, Jul. 2017.

[24] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[25] H. Xie, F. Gao, S. Zhang, and S. Jin, "UL/DL channel estimation for TDD/FDD massive MIMO systems using DFT and angle reciprocity," in *Proc. IEEE 83rd Veh. Technol. Conf.*, May 2016, pp. 1–5.

[26] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[27] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[28] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[29] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[30] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, Dec. 2012.

[31] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, Jun. 2014.

[32] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[33] Y. Wang, P. Xu, and Z. Tian, "Efficient channel estimation for massive MIMO systems via truncated two-dimensional atomic norm minimization," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.

[34] Y. Tsai, L. Zheng, and X. Wang, "Millimeter-wave beamformed full-dimensional MIMO channel estimation based on atomic norm minimization," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6150–6163, Dec. 2018.

[35] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 303–318, Jan. 2017.

[36] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

[37] F. Belloni, A. Richter, and V. Koivunen, "DoA estimation via manifold separation for arbitrary array structures," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4800–4810, Oct. 2007.

[38] A. B. Gershman, M. Rübsamen, and M. Pesavento, "One- and two-dimensional direction-of-arrival estimation: An overview of search-free techniques," *Signal Process.*, vol. 90, no. 5, pp. 1338–1349, May 2010.

[39] A. Govinda Raj and J. H. McClellan, "Single snapshot super-resolution DOA estimation for arbitrary array geometries," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 119–123, Jan. 2019.

[40] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[41] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. Mineola, NY, USA: Dover, 1972.

[42] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[43] S. Pejoski and V. Kafedziski, "Estimation of sparse time dispersive channels in pilot aided OFDM using atomic norm," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 397–400, Aug. 2015.

[44] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1257–1269, Mar. 2016.

[45] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5145–5157, Oct. 2016.

[46] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.

[47] X. Wu, W. Zhu, and J. Yan, "A Toeplitz covariance matrix reconstruction approach for direction-of-arrival estimation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8223–8237, Sep. 2017.

[48] C. P. Mathews and M. D. Zoltowski, "Eigenstructure techniques for 2-D angle estimation with uniform circular arrays," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2395–2407, Sep. 1994.

[49] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Mar. 2014. [Online]. Available: http://cvxr.com/cvx

[50] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[51] O. Toeplitz, "Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen," *Mathematische Annalen*, vol. 70, no. 3, pp. 351–376, Sep. 1911.

[52] S. Li, D. Yang, G. Tang, and M. B. Wakin, "Atomic norm minimization for modal analysis from random and compressed samples," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1817–1831, Apr. 2018.

[53] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1030–1042, Feb. 2015.

[54] Z. Yang, L. Xie, and P. Stoica, "Vandermonde decomposition of multilevel Toeplitz matrices with application to multidimensional super-resolution," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3685–3701, Jun. 2016.

[55] Z. Zhang, Y. Wang, and Z. Tian, "Efficient two-dimensional line spectrum estimation based on decoupled atomic norm minimization," *Signal Process.*, vol. 163, pp. 95–106, 2019.

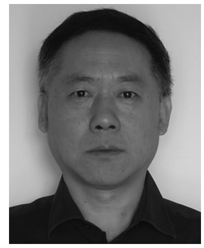
[56] Z. Tian, Z. Zhang, and Y. Wang, "Low-complexity optimization for two-dimensional direction-of-arrival estimation via decoupled atomic norm minimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2017, pp. 3071–3075.

Yue Wang (M'11) received the Ph.D. degree in communication and information system from the Beijing University of Posts and Telecommunications, Beijing, China, in 2011. He is currently a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, USA. Prior to that, he was a Senior Engineer with Huawei Technologies Co., Ltd., Beijing, China. From 2009 to 2011, he was a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI, USA. His general interests are in the areas of wireless communications and signal processing. His specific research focuses on compressive sensing, massive MIMO, millimeter-wave communications, DoA estimation, cognitive radios, and optimization.

Yu Zhang (S'19) received the B.E. degree in 2013 from the College of Electronics and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, where he is currently working toward the Ph.D. degree in communication and information system. He is currently a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, USA. His research interests include compressive sensing and array signal processing in radar and communications.

Zhi Tian (M'00–SM'06–F'13) has been a Professor with the Electrical and Computer Engineering Department, George Mason University, Fairfax, VA, USA since 2015. Prior to that, she was on the Faculty of Michigan Technological University from 2000 to 2014. She was a Program Director with the U.S. National Science Foundation from 2012 to 2014. Her research interest lies in the areas of statistical signal processing, wireless communications, and estimation and detection theory. Her current research focuses on compressed sensing for random processes, statistical inference of network data, distributed network optimization and learning, and millimeter-wave communications. She was an IEEE Distinguished Lecturer for both the IEEE Communications Society and the IEEE Vehicular Technology Society. She was an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON SIGNAL PROCESSING. She was a General Co-Chair of the 2016 IEEE GlobalSIP Conference, and is a Chair of the IEEE Signal Processing Society Big Data Special Interest Group. She is a Member-at-Large of the IEEE Signal Processing Society Board of Governors for the term of 2019–2021.

Geert Leus (F'12) received the M.Sc. and Ph.D. degrees in electrical engineering from Katholieke Universiteit Leuven, Leuven, Belgium, in June 1996 and May 2000, respectively. He is currently an "Antoni van Leeuwenhoek" Full Professor with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands. His research interests are in the broad areas of signal processing, with a specific focus on wireless communications, array processing, sensor networks, and graph signal processing. He was the recipient of the 2002 IEEE Signal Processing Society Young Author Best Paper Award and the 2005 IEEE Signal Processing Society Best Paper Award. He is a Fellow of EURASIP. He was a Member-at-Large of the Board of Governors of the IEEE Signal Processing Society, the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, a Member of the IEEE Sensor Array and Multichannel Technical Committee, and the Editor-in-Chief of the *EURASIP Journal on Advances in Signal Processing*. He was also on the editorial boards of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE SIGNAL PROCESSING LETTERS, and the *EURASIP Journal on Advances in Signal Processing*. He is currently the Chair of the EURASIP Special Area Team on Signal Processing for Multisensor Systems, a Member of the IEEE Signal Processing Theory and Methods Technical Committee, a Member of the IEEE Big Data Special Interest Group, an Associate Editor of *Foundations and Trends in Signal Processing*, and the Editor-in-Chief of *EURASIP Signal Processing*.



Gong Zhang (M'07) received the Ph.D. degree in electronic engineering from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2002. From 1990 to 1998, he was a Member of Technical Staff with No. 724 Institute China Shipbuilding Industry Corporation, Nanjing, China. Since 1998, he has been with the College of Electronic and Information Engineering, NUAA, where he is currently a Professor. He is a Member of the Committee of Electromagnetic Information, Chinese Society of Astronautics, a Member of Editorial Board for *Journal of Radar*, China, and a Senior Member of the Chinese Institute of Electronics. His research interests include signal processing for communications, radar, and sensor processing, statistical and array signal processing, applications of linear algebra and optimization methods in signal processing and communications, estimation and detection theory, sampling theory, classification and recognition, and cooperative and cognitive systems.