

Received 22 June 2023, accepted 13 July 2023, date of publication 25 July 2023, date of current version 2 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3298702

## APPLIED RESEARCH

# Sparse Kalman Filter-Based Channel Estimation for RIS-Aided Millimeter Wave Multiple-Input Multiple-Output Systems

JING ZHANG<sup>1</sup>, YING SU<sup>1</sup>, AND DONGJUN QIAN<sup>2</sup>

<sup>1</sup>College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 201418, China

<sup>2</sup>Shanghai Ideal Information Company Ltd., Shanghai 201315, China

Corresponding author: Ying Su (980154894@qq.com)

This work was supported by the Project of Shanghai Normal University under Grant 309-C-9000-21-309199.

**ABSTRACT** The acquisition of accurate channel state information is critical for enhancing the transmission quality and energy efficiency of reconfigurable intelligent surface-aided millimeter wave systems with multiple-input multiple-output antenna arrays at the transceiver ends. Due to the cascaded doubly sparse channel property in angular and delay domains, it is difficult to acquire nonzero angular gains sampled by overcomplete dictionaries with low pilot overhead. In this work, a channel-estimation method based on the sparse Kalman filter (SKF) framework is proposed, in which a state-space model is established for a sparse complex angular gain vector in both static and dynamic scenarios, and a zero-padding block-diagonal pilot pattern is designed. A linear  $l_1$ -norm subdifferential pseudo-measurement equation is deduced. The on-grid angular gain vector, along with the corresponding support, is recovered by the sequential processing of multiple pilot tone observations. To enhance the  $l_1$ -norm convergence and thus reduce the pilot overhead, an accelerated algorithm is developed, which incorporates an exponential factor to accelerate the gradient descent procedure. The simulation results demonstrate that the method can capture the angular support using a few pilot blocks. The normalized mean square error performance for the angular gain vector was better than that for the orthogonal matching pursuit. The accelerated SKF algorithm achieves faster convergence and higher accuracy compared with the ordinary SKF algorithm, and it is easy to implement online.


**INDEX TERMS** Channel estimation, compressed sensing, Kalman filters, millimeter-wave communication, reconfigurable intelligent surfaces.

## I. INTRODUCTION

Low-cost and low-power reconfigurable intelligent surfaces (RISs) are promising candidates for future sensing, computing, and communication applications. They can be employed to adjust wireless environments between the transmitter and receiver, such as user equipment (UE) and base stations (BS). An RIS typically comprises numerous reflective elements that introduce additional phase shifts into the incident signals and passively reflect the signals without processing or energy consumption. The phase shifts of RIS elements can

be adjusted to maximize spectral and energy efficiencies [1], [2].

Efficient channel estimation (CE) with low pilot overhead for the RIS auxiliary link is essential. The methods for achieving efficient CE include tensor decomposition and sparse recovery [3]. In the former category, e.g., in a multiple user multiple-input single-output (MISO) millimeter wave (mmWave) uplink system, the parallel factorization model is utilized to unfold the cascaded channel. By designing a specific structured spatiotemporal pilot pattern, the CE can be performed using a bilinear alternating least square (LS) algorithm and the approximate message passing (AMP) method [4], [5], [6]. The Cramer–Rao bound was derived to evaluate the normalized mean square error (NMSE) [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Halil Ersin Soken .

The time-scale channel separation and estimation were covered in [8]. Typically, an alternating two-stage method is adopted for cascaded CE. A decomposition and completion method for sparse channel matrices is proposed by exploring slowly changing channel components and hidden sparsity [9]. Considering the similar sparse channel characteristics in the angular domain, the bilinear general AMP method was investigated for passive RIS-aided systems [10]. The CE problem is interpreted as a sparse matrix decomposition problem solved using an AMP algorithm based on Bayesian inference [11]. In [12], RIS auxiliary and direct links were simultaneously considered. The CE was performed for each subsurface channel, and the entire channel was estimated by adjusting the RIS phase shift.

The latter category is based on compressed sensing. The inherent sparse characteristics of RIS-aided mmWave channels are fully modeled. When considering the sparsity in the angular domain, the cascaded gain matrix can be vectored using the Khatri–Rao and Kronecker product formulas. Orthogonal matching pursuit (OMP) and AMP have been used for recovery [13]. For a doubly sparse channel matrix with common nonzero rows and partially common nonzero columns, the dual OMP algorithm was utilized to jointly estimate the support and gains [14]. To achieve improved recovery, an iterative method was conducted to solve a non-convex optimization problem [15]. The problem was modeled as an  $l_1$ -norm regularization problem with a fixed rank constraint [16]. By establishing a block-sparse structure for all users, alternating minimization and manifold optimization were implemented to determine the optimum solution.

Detailed channel-state information in the angular domain is desired for cascaded CE. By considering the broadband model, including the angle of departure (AoD), angle of arrival (AoA), complex gain, and delay, a five-step Newtonized OMP was introduced. Angle parameters were obtained through an oversampled grid search. The cascaded gain was updated using LS [17]. Notably, the two-stage OMP algorithm improves performance using the mutual correlation function between spatial steering vectors [18]. The high-dimensional broadband channels of the downlink BS-RIS and RIS-UE were jointly estimated using a distributed OMP [19]. These channels can also be acquired using an iterative reweighting algorithm [20]. In [21], the atomic norm minimization method is used to estimate AoAs/AoDs. The alternating direction multiplier method could restore the cascaded angular gains [22]. A two-stage CE was designed based on partial knowledge of the RIS-BS channel, and it estimated the direct links of UE-BS and UE-RIS in Stage 1. Subsequently, RIS phase shifts were designed to form an initial estimate of the global channel. In Stage 2, additional pilots were utilized to refine the estimation [23]. In addition, data-transmission protocols for practical applications were designed using limited feedback channel information [24].

To solve the cascaded CE problem, it is necessary to design a pilot pattern to reduce the pilot overhead and improve accu-

racy. A hybrid RIS with a few active elements was obtained for a CE solution [25], [26]. Nonuniformly spaced comb-type pilots for compressed CE in MISO orthogonal frequency division multiplexing systems are presented in [27]. It is possible to accurately estimate many channel coefficients using only a few three-phase pilots [28]. The reflection coefficient of the RIS element can be adjusted in advance for CE [29].

Other CE studies on RIS-aided systems have also been conducted. In intelligent-learning CE, an enhanced limit learning machine network is constructed to improve performance [30], in addition to deep reinforcement learning networks [31] and convolutional neural networks combined with long short-term memory [32]. In terahertz RIS-aided, large-scale multiple-input multiple-output (MIMO) systems, the CE procedure includes coarse-grained and fine-grained stages for downlink and uplink, respectively [33]. In mobility scenarios, the channel is modeled as a state-space model and estimated using a Kalman filter (KF) [34], [35], an extended KF, and a particle filter [36]. The best reflection coefficient is obtained using the minimum mean-squared error criterion. Additionally, the KF is utilized to estimate off-grid AoAs in MIMO radar systems [37].

This paper focuses on the CE for a RIS-aided mmWave channel that is sparse in both the angular and delay domains. This type of doubly sparse channel model has been explored for single-hop mmWave and RIS-aided systems [38], [39], [40]. A CE method based on the sparse Kalman filter (SKF) framework was developed. By analyzing the sparse channel structure in the angular and delay domains and designing a zero-padding block diagonal pilot pattern, SKF algorithms were developed to recover the sparse angular gain vector. The algorithms incorporate the constraint of the  $l_1$ -norm using a pseudo-measurement technique. Simulations show that the accelerated SKF algorithm outperforms Oracle LS, OMP, and an ordinary sparse filtering algorithm in terms of the NMSE, support capture, convergence speed, and pilot overhead.

The main contributions of this study are as follows.

(1) A CE method within the SKF framework is proposed for RIS-aided MIMO systems. The method transforms the common sparse channel recovery problem into sequential retrieval through multiple pilot tone observations. As the method is insensitive to the coherence of angular dictionaries, it is robust in practice. Moreover, it can recover a cascaded channel when each row or column of the gain matrix has either zero or only one nonzero entry. This permits the construction of angular dictionaries without restriction.

(2) SKF algorithms with  $l_1$ -norm constraints were developed. The  $l_1$ -norm subdifferential pseudo-measurement equation was deduced for sparse complex angular gain vectors. An accelerated SKF algorithm is proposed to accomplish CE using a few pilot blocks, and its superior performance is validated by simulation studies. Moreover, these algorithms are easy to implement online.

(3) The proposed method is suitable for both static and dynamic scenarios that involve noisy pilot tone observa-

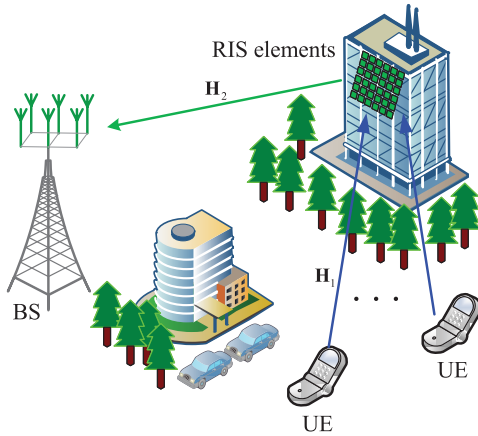


FIGURE 1. RIS-aided uplink transmission scenarios.

tions. In particular, its ability to achieve a dynamic sparse gain vector along with an accurate support renders the CE method suitable for various scenarios. Although a zero-padding block-diagonal pilot pattern is required to exploit the delay sparsity, accurate and detailed angular information can be obtained.

The remainder of this paper is organized as follows. In Section II, the doubly sparse channel models of UE-RIS and RIS-BS are introduced, and the state-space model of the cascaded channel is explained. In Section III, the pilot pattern and the received signal model are described. In Section IV, the proposed SKF algorithm and its computational complexity are presented. In Section V, the simulation results are presented. Finally, conclusions are drawn in Section VI.

**Notations.** Bold capital letters and symbols denote matrices. Bold lowercase letters denote column vectors. The italic letter  $(j)$  denotes  $\sqrt{-1}$ .  $(\cdot)^H$ ,  $(\cdot)^T$ ,  $(\cdot)^*$ , and  $(\cdot)^{-1}$  denote the conjugate transpose, transpose, conjugate, and inverse, respectively.  $\|\cdot\|_2$ ,  $\|\cdot\|_1$ , and  $|\cdot|$  denote the  $l_2$ -norm,  $l_1$ -norm, and absolute value, respectively. The diagonal matrix is denoted as  $diag(\mathbf{a})$ , where a vector is placed on the diagonal. The vectored matrix is denoted as  $vec(\mathbf{A})$  by the column stacking of the matrix. The symbols,  $\otimes$  and  $\odot$ , denote the Kronecker and Khatri–Rao products, respectively.  $\mathbf{I}_N$  denotes an  $N$ -dimensional identity matrix. The  $N \times M$  zero matrix is  $\mathbf{0}_{NM}$ . The zero row vector is  $\mathbf{0}_N$ .  $\lfloor \cdot \rfloor$  denotes the round-down operation.  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  represents the complex Gaussian distribution with mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ .  $\mathbb{E}\{\cdot\}$  represents the mathematical expectation.

## II. CASCADED CHANNEL MODEL

The uplink of the RIS-aided mmWave transmission system is shown in Fig. 1, where the number of antenna arrays at the BS is denoted as  $N_R$ . Each UE has a single antenna. Multiple UEs form a virtual  $N_T$  multi-antenna that shares wireless resources, particularly time slots. When the direct link of the line of sight is blocked, signals from the UEs to the BS are reflected by the RIS. The RIS is composed of  $N_I$  elements to adjust the phase shifts of the incident signals. It is assumed that each element can be adjusted to the

on-off state and provide a discrete phase shift. The reflective coefficient of each element is constant. The RIS-BS channel ( $\mathbf{H}_2 \in \mathbb{C}^{N_R \times N_I}$ ) exhibits a long timescale property because the BS and RIS are both at fixed deployment locations. The entries remain unchanged within multiple symbol frames. For the mobility of the UE, the UE-RIS channel ( $\mathbf{H}_1 \in \mathbb{C}^{N_I \times N_T}$ ) varies on a considerably smaller timescale than that for the quasi-static channel ( $\mathbf{H}_2$ ).

Without loss of generality, suppose that the antennas of the UE, BS, and RIS elements form half-wavelength uniform planar arrays. Channel  $\mathbf{H}_1$  at the  $d$ -th delay tap in the  $s$ -th transmission block is given by

$$\mathbf{H}_1(s) = \sqrt{\frac{N_T N_I}{L_1}} \sum_{i=1}^{L_1} \rho_{1i,s} \delta(dT_s - \tau_{1i}) \times \mathbf{a}_r(\varphi_{1i}^r, \theta_{1i}^r) \mathbf{a}_t^H(\varphi_{1i}^t, \theta_{1i}^t), \quad (1)$$

where  $L_1$  is the number of distinguishable paths.  $\rho_{1i,s}$  is the complex gain of the  $i$ -th spatial multipath component in the  $s$ -th transmission block.  $\tau_{1i}$  denotes the propagation delay, which obeys a uniform distribution.  $\delta(\cdot)$  denotes the band-limited pulse-shaping filter response.  $\varphi_{1i}^r(\theta_{1i}^r)$  and  $\varphi_{1i}^t(\theta_{1i}^t)$  are the azimuth (elevation) angles of arrival and departure, respectively.  $\mathbf{a}_r(\cdot)$  and  $\mathbf{a}_t(\cdot)$  represent the receive and transmit steering vectors, respectively. For a planar array with size  $N = N_y N_z$ , the steering vector is given by

$$\mathbf{a}(\varphi, \theta) = \frac{1}{\sqrt{N_y N_z}} (1, \dots, e^{jkd(m \sin \varphi \sin \theta + n \cos \theta)}, \dots, e^{jkd((N_y - 1) \sin \varphi \sin \theta + (N_z - 1) \cos \theta)})^T, \quad (2)$$

where  $0 \leq m < N_y$  and  $0 \leq n < N_z$  are the  $y$  and  $z$  indices of the antenna element, respectively.  $k = 2\pi/\lambda$ ,  $\lambda$  is the signal wavelength.  $d$  is the spacing of the antennas,  $d = 0.5\lambda$ .

The incident signals from the different transmitting antennas are reflected by the RIS elements. Subsequently, additional phase shifts are added to the incident signals. Let the phase shift vector be  $\mathbf{g} = (g_1 e^{j\phi_1}, \dots, g_{N_I} e^{j\phi_{N_I}})^T$ , where  $g_n, n = 1, \dots, N_I$  denotes the reflective coefficient of the  $n$ -th RIS element.  $\phi_n \in (0, 2\pi]$ .

Channel  $\mathbf{H}_2$  is modeled similarly to (1) and (2). Owing to the scattering nature of mmWave, the number of distinguishable paths of UE-RIS and RIS-BS is significantly smaller than the dimension of the channel matrix. When the UE is sufficiently close, the channel responses of UE-RIS and RIS-BS are considered to share common angular-domain and delay-domain sparse properties.

### A. SPARSE PROPERTY IN THE ANGULAR DOMAIN

The geometric channel model (1) is expressed in compact form, as follows:

$$\mathbf{H}_1(s) = \mathbf{A}_{1R} diag(\mathbf{g}_1(s)) \mathbf{A}_{1T}^H, \quad (3)$$

where  $\mathbf{A}_{1R} = [\mathbf{a}_r(\varphi_{11}^r, \theta_{11}^r), \dots, \mathbf{a}_r(\varphi_{1L_1}^r, \theta_{1L_1}^r)] \in \mathbb{C}^{N_I \times L_1}$  and  $\mathbf{A}_{1T} = [\mathbf{a}_t(\varphi_{11}^t, \theta_{11}^t), \dots, \mathbf{a}_t(\varphi_{1L_1}^t, \theta_{1L_1}^t)] \in \mathbb{C}^{N_T \times L_1}$

denote the steering matrices that remain unchanged during the CE stage.  $\mathbf{g}_1(s)$  denotes the beam response vector in the  $s$ -th transmission block.

$$\mathbf{g}_1(s) = \sqrt{\frac{N_T N_I}{L_1}} [\rho_{11,s} \delta(dT_s - \tau_{11}), \dots, \rho_{1L,s} \delta(dT_s - \tau_{1L})]^T, \quad (4)$$

where  $\mathbf{g}_1(s)$  contains all the distinguishable nonzero responses associated with the  $L_1$  paths.

Due to the limited angular quantization and resolution,  $\mathbf{A}_{1R}$  and  $\mathbf{A}_{1T}$  are sampled using overcomplete dictionary matrices constructed as

$$\Theta_{1T} = [\mathbf{a}_{r,0}, \mathbf{a}_{r,1}, \dots, \mathbf{a}_{r,\beta_1-1}], \quad (5a)$$

$$\Theta_{1R} = [\mathbf{a}_{r,0}, \mathbf{a}_{r,1}, \dots, \mathbf{a}_{r,\beta_1-1}], \quad (5b)$$

where  $\beta_1$  denotes the planar angular grid size of  $\mathbf{H}_1(s)$ .  $\beta_1 = \beta_{y1} \beta_{z1}$ ,  $\beta_{y1} = 2\pi/F_{a1}$ ,  $\beta_{z1} = \pi/F_{a1}$ , where  $F_{a1}$  is the resolution for both the azimuth and the elevation angles. In accordance with the angular grids, the directional response matrix ( $\Theta_{1g}(s)$ ) is expressed as

$$\Theta_{1g}(s) = \begin{bmatrix} g_{1,11}(s) & \cdots & g_{1,1\beta_1}(s) \\ \vdots & \ddots & \vdots \\ g_{1,\beta_1}(s) & \cdots & g_{1,\beta_1\beta_1}(s) \end{bmatrix}. \quad (6)$$

The nonzero angular gain ( $g_{1,ij}(s)$ ,  $i, j \in \{1, \dots, \beta_1\}$ ) is modeled as the state-space equation:

$$g_{1,ij}(s) = \psi g_{1,ij}(s-1) + \sqrt{1 - \psi^2} w_{1,ij}(s-1), \quad (7)$$

where  $\psi$  denotes the temporal correlation coefficient of the gain. It is evaluated from Jake's model as  $\psi = J_0(2\pi f_D T_s)$ , where  $J_0(\cdot)$  is the zeroth order Bessel function of the first kind, and  $f_D$  denotes the maximum Doppler frequency.  $w_{1,ij}(s-1)$  denotes the innovation noise, and  $w_{1,ij}(s-1) \sim \mathcal{CN}(0, q)$ . The initial nonzero components of  $\Theta_{1g}(0)$  are irrelevant to the corresponding noise matrix ( $\mathbf{W}_1(s)$ ).

Based on the dictionaries with  $\beta_1$  grids, the UE-RIS channel can be rewritten as

$$\mathbf{H}_1(s) = \Theta_{1R} \Theta_{1g}(s) \Theta_{1T}^H. \quad (8)$$

Dictionaries with predefined grids allow the detection of the nonzero entries of  $\Theta_{1g}(s)$ , which indicate the angular gains along with the implicit AoAs/AoDs of the dominant paths [41], [42], [43], [44]. If the real beam does not fall into one of the grids, off-grid leakage will occur, resulting in extra nonzero entries in  $\Theta_{1g}(s)$ . When the grid density is sufficiently high, the mismatch error is conventionally neglected [45]. It is assumed that only the  $L_1 (\ll \beta_1^2)$  entries of  $\Theta_{1g}(s)$  are nonzero, which indicates the sparsity property in the angular domain.

Similarly, channel  $\mathbf{H}_2$  is treated in the same way as  $\mathbf{H}_1(s)$ . The angular dictionary matrices are formed using  $\beta_2$  grids, and the time-invariant  $\mathbf{H}_2$  is expressed as

$$\mathbf{H}_2 = \Theta_{2R} \Theta_{2g} \Theta_{2T}^H. \quad (9)$$

It is assumed that only the  $L_2 (\ll \beta_2^2)$  entries of  $\Theta_{2g}$  are nonzero.

### B. SPARSE PROPERTY in the DELAY DOMAIN

In a multipath wireless environment, the delay spread increases the duration of the received signal. This time dispersion results in frequency-selective fading and inter-symbol interference (ISI), which is conventionally eliminated by the guard interval in block transmission methods. As depicted in [38], most mmWave channel taps are extremely weak. Very few dominant paths are observed in the frame period ( $T_s$ ), which is less than the channel coherence time.

For channel  $\mathbf{H}_1$ , by dividing the delay on the  $\gamma_1$  time-slot grids with resolution  $F_d = T_s/\gamma_1$ , the  $l$ -th path delay can fall into one grid ( $\gamma_{1,l} = \lfloor \tau_{1l}/F_d \rfloor$ ,  $l \in \{1, \dots, L_1\}$ ). Similarly, the  $L_2$  path delays fall into the corresponding grids divided by the  $\gamma_2$  grids. The cascaded  $L = L_1 L_2$  path delays are associated with the  $L$  grids in the total  $\gamma_1 \gamma_2$  grids. Thus, the cascaded channel exhibits a sparse property in the delay domain. As delays are slow-varying parameters, the delay support in consecutive  $s = 1, \dots, S$  frames is considered to be constant.

The channel sparsity property in the delay domain is hidden in transmission and used to design a pilot pattern in a block transmittal.

### C. CASCADED CHANNEL MODEL

The cascaded channel is defined as

$$\mathbf{H}(s) = \mathbf{H}_2 \mathbf{D}_s \mathbf{H}_1(s), \quad (10)$$

where  $\mathbf{D}_s \triangleq \text{diag}(\mathbf{g}(s))$ . The cascaded channel includes the phase configuration of the RIS elements. It is chosen as a low-resolution discrete set.

Through the dictionary matrices, we have

$$\mathbf{H}(s) = \Theta_{2R} \Theta_{2g} \Theta_{2T}^H \mathbf{D}_s \Theta_{1R} \Theta_{1g}(s) \Theta_{1T}^H. \quad (11)$$

Sparse  $\Theta_{1g}(s)$  and  $\Theta_{2g}$  are to be acquired in the CE. The advantage of the angular gain estimation lies in its detailed beam direction and gain information. After obtaining directional gains, beamforming can be implemented to improve the signal-to-noise ratio (SNR). Once the gains are accurately known, the cascaded MIMO channel can be obtained.

Using  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$  and  $\text{vec}(\mathbf{A} \text{diag}(\mathbf{b}) \mathbf{C}) = (\mathbf{C}^T \odot \mathbf{A}) \mathbf{b}$ , the channel is expressed in vector format [22]:

$$\text{vec}(\mathbf{H}(s)) = ((\Theta_{2T}^* \odot \Theta_{1R}) \mathbf{g}^*(s))^T \otimes (\Theta_{1T}^* \otimes \Theta_{2R}) \text{vec}(\Theta_g(s)), \quad (12)$$

where  $\text{vec}(\Theta_g(s)) = \text{vec}(\Theta_{1g}^T(s) \otimes \Theta_{2g})$ .

It is determined that  $\Theta_g(s)$  is considerably sparser than  $\Theta_{1g}(s)$  and  $\Theta_{2g}$ . A sparse example of  $\Theta_g(s)$  is shown in Fig. 2 when  $N_T = N_R = N_I = \beta_1 = \beta_2 = 4$ , with  $L_1 = 2$  and  $L_2 = 3$ . An arbitrary row or column has either zeros or one nonzero entry. In addition, notably, in one-hop mmWave transmissions, the dictionary size is selected to be

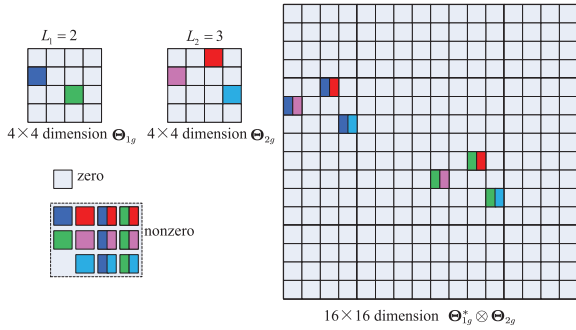


FIGURE 2. Sparse structure in the angular domain.

2 ~ 4 times the array size [38]. However, in this two-hop system, the dictionary matrices need to be constructed with a comparatively coarse grid to overcome high-dimensional processing. The dictionaries can be first constructed using one times the array size and subsequently refined around the coarse AoAs/AoDs. Using this strategy, the off-grid error can be ignored.

We define the vector to be recovered as  $\mathbf{h}(s) \triangleq \text{vec}(\Theta_g(s))$ . The state-space model in (7) is established as follows:

$$\mathbf{h}(s) = \psi \mathbf{h}(s-1) + \sqrt{1-\psi^2} \mathbf{w}(s-1), \quad (13)$$

where  $\mathbf{w}(s-1) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ . Notably, zero entries are always zero. The sparsity structure is irrelevant to  $s$ . It is a static sparse model when  $\psi = 1$ , whereas it is a dynamic sparse model when  $\psi < 1$ .

### III. RECEIVED SIGNAL MODEL

For this doubly sparse cascaded CE, a block-diagonal pilot pattern is designed. As the multitap response incurs ISI, zero padding in the pilot frames is implemented to separate the convolution of responses without a discrete Fourier transformation [38]. This pattern is illustrated in Fig. 3. One frame is divided into two subframes for the pilot and traffic data transmission. The pilot subframe is divided into  $N_T L N_g$  subslots. Each  $N_g$ -length subslot includes one pilot signal and  $N_g - 1$  guard interval,  $N_g > L$ .

For the  $t$ -th UE, the pilot is transmitted as

$$\mathbf{x}_{s,t} = \left( \overbrace{x_{s,1}, 0 \cdots 0}^{N_g-1}, \overbrace{x_{s,2}, 0 \cdots 0}^{N_g-1}, \dots, \overbrace{x_{s,L}, 0 \cdots 0}^{N_g-1} \right). \quad (14)$$

Inter-antenna interference can also be eliminated using a block-diagonal pilot pattern. Let  $\mathbf{X}_{s,N_T}$  be the block pilot symbols in a single-frame period.

$$\mathbf{X}_{s,N_T} = \begin{pmatrix} \mathbf{x}_{s,1} & \mathbf{0}_{N_g} & \cdots & \mathbf{0}_{N_g} \\ \mathbf{0}_{N_g} & \mathbf{x}_{s,2} & \cdots & \mathbf{0}_{N_g} \\ \mathbf{0}_{N_g} & \mathbf{0}_{N_g} & \ddots & \mathbf{0}_{N_g} \\ \mathbf{0}_{N_g} & \mathbf{0}_{N_g} & \cdots & \mathbf{x}_{s,N_T} \end{pmatrix}. \quad (15)$$

The pilot pattern results in multiple one-tap transmissions, as follows:

$$\mathbf{Y}_s(m, l) = \mathbf{H}(s) \mathbf{X}(s, l) + \mathbf{U}_s(m), \quad (16)$$

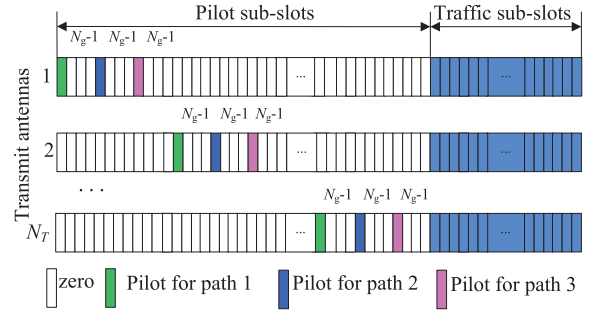


FIGURE 3. Block-diagonal pilot pattern.

where  $m$  denotes the probing instant.  $\mathbf{X}(s, l)$  is

$$\mathbf{X}(s, l) = \begin{pmatrix} x_{1,l} & 0 & \cdots & 0 \\ 0 & x_{2,l} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & x_{N_T,l} \end{pmatrix}. \quad (17)$$

Hence, the pilot is designed as  $\mathbf{X}(s, l) = \frac{P}{L} \mathbf{I}_{N_T}$ , where  $P$  is the total transmit power,  $P = \sum_{l=1}^L \sum_{t=1}^{N_T} x_{t,l}$ .

The multitap channel responses were probed in every sub-slot and combined. The received signal model is expressed as follows:

$$\mathbf{Y}_s(m) = \mathbf{H}_2 \mathbf{D}_s \mathbf{H}_1(s) \mathbf{X}(s) + \mathbf{U}_s(m), \quad (18)$$

where  $\mathbf{Y}_s(m) \in \mathbb{C}^{N_R \times N_T}$ ,  $\mathbf{X}(s) \in \mathbb{C}^{N_T \times N_T}$ , and  $\mathbf{H}(s) = \mathbf{H}_2 \mathbf{D}_s \mathbf{H}_1(s)$  contain all the  $L$  path responses. Consequently, the sensing matrix is formed as  $\mathbf{A}(s) \triangleq ((\Theta_{2T}^* \odot \Theta_{1R}) \mathbf{g}^*(s))^T \otimes ((\mathbf{X}^T(s) \Theta_{1T}^*) \otimes \Theta_{2R})$ . The pilot tone observation was  $\mathbf{y}(s) \triangleq \text{vec}(\mathbf{Y}_s(m))$ . The observation noise was  $\mathbf{u}(s) \triangleq \text{vec}(\mathbf{U}_s(m))$ . Thus, we have

$$\mathbf{y}(s) = \mathbf{A}(s) \mathbf{h}(s) + \mathbf{u}(s), \quad (19)$$

where  $\mathbf{u}(s) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ .

### IV. SPARSE RECOVERY OF THE CASCADED CHANNEL

According to the sparse-property analysis described in Section II, the cascaded channel is structurally sparse. It has a fixed support in the angular and delay domains. The sparse  $\mathbf{h}(s)$  is expected to be recovered using a single measurement vector (SMV) and multiple measurement vectors (MMVs). Owing to the cascaded setting, the pilot tone observation ( $\mathbf{y}(s)$ ) is an  $N_R N_T \times 1$  vector, whereas  $\mathbf{h}(s)$  is a  $\beta_1^2 \beta_2^2 \times 1$  vector. Sensing matrix  $\mathbf{A}(s)$  is mutually coherent. It is difficult to use a single  $\mathbf{y}(s)$  to recover  $\mathbf{h}(s)$ . It is necessary to seek help from MMV  $\mathbf{y}(1), \dots, \mathbf{y}(s)$ .

#### A. RELATED SPARSE RECOVERY METHODS

For an underdetermined sensing equation, the sparsest solution is to minimize the  $l_0$ -norm of the unknown sparse vector. The  $l_0$ -norm is a nonconvex function, and its minimization is an NP-hard problem. The  $l_1$ -norm is the optimal convex approximation of the  $l_0$ -norm. The sparsity obtained by minimizing the  $l_1$ -norm is smaller than that obtained by minimizing the  $l_2$ -norm. Based on the norm-optimization metric,

there are many algorithms for recovering a sparse vector. For the SMV, typical algorithms include Oracle LS, the greedy class of matching pursuit (e.g., OMP), and the class of basis pursuit algorithms including iterative  $l_1$ -norm minimization (e.g., the least absolute shrinkage and selection operator). Oracle LS is the LS solution when the sensing matrix is underdetermined. OMP conventionally requires prior knowledge of sparsity. In addition, AMP, along with its improved algorithms, is a promising method for solving the recovery problem. However, its pilot overhead is typically high, owing to its iterative structure and probability inference. Sparse Bayesian learning and its sequential approaches also suffer from high pilot overheads.

The retrieval of  $\mathbf{h}(s)$  is challenging. Among the current estimation methods, KF belongs to the category of maximum posterior estimation. It can track the time-varying state and provide a recursive optimal estimate in a linear state-space model. KF incorporated with norm constraints has been applied to sparse signal reconstruction using the temporal correlation of sparse signals [46], [47], [48]. The SKF framework exploits the updating and norm-correction structures, providing a solution for sequentially estimating the static or dynamic  $\mathbf{h}(s)$  from the MMV. Compared with a method that is dependent on SMV, which obtains the solution in MMV by averaging multiple recovered sparse vectors regardless of their temporal correlation, SKF is the choice for the recovery of structurally sparse  $\mathbf{h}(s)$ . The SKF framework and corresponding algorithms are preferred because they present a filtering structure for static and dynamic  $\mathbf{h}(s)$  in a sequential processing manner, and they are easy to implement.

### B. ORDINARY SPARSE KALMAN FILTER ALGORITHM

The optimization problem with the  $l_1$ -norm constraint for the recovery of sparse  $\mathbf{h}(s)$  in the SKF framework is expressed as

$$\min \sum_{s=1}^S \left\{ \left\| \hat{\mathbf{h}}(s) - \mathbf{h}(s) \right\|_2^2 \right\} \quad \text{s.t.} \quad \left\| \hat{\mathbf{h}}(s) \right\|_1 \leq \epsilon. \quad (20)$$

The inequality constraint is incorporated into the filtering process using the pseudo-measurement technique:

$$0 = \|\mathbf{h}(s)\|_1 - \epsilon, \quad (21)$$

where  $\epsilon$  serves as the measurement noise. The variance ( $r_\epsilon$  of  $\epsilon$ ) is regarded as a tuning factor that regulates the tightness of the constraint ( $\|\hat{\mathbf{h}}(s)\|_1 \leq \epsilon$ ). Interestingly, taking  $\epsilon$  to be high intensity, e.g.,  $r_\epsilon = \gamma^2$  with  $\gamma \geq 200$ , yields satisfactory accuracy [49].

For linear filtering, Eq. (21) must be linearized by the subdifferential of the  $l_1$ -norm, which is deduced as follows:

$$\frac{\partial \|\mathbf{h}(s)\|_1}{\partial h_n(s)} = \frac{\partial \sum \sqrt{h_n(s)h_n^*(s)}}{\partial h_n(s)} = \frac{h_n(s)^{-\frac{1}{2}}h_n^*(s)^{\frac{1}{2}}}{2}, \quad (22)$$

where  $h_n(s)$  is the  $n$ -th element of  $\mathbf{h}(s)$  in the  $s$ -th recursion. Thus, the Jacobian matrix is written as

$$\mathbf{b}(s) = \left( \frac{\partial \|\mathbf{h}(s)\|_1}{\partial h_1(s)}, \dots, \frac{\partial \|\mathbf{h}(s)\|_1}{\partial h_{\beta_1^2 \beta_2^2}(s)} \right)_{h_n(s)=\hat{h}_n(s)}. \quad (23)$$

### Algorithm 1 O-SKF for the Angular Gain Recovery

**Input:**  $\psi, \mathbf{y}(s), \mathbf{A}(s), \mathbf{Q}, \mathbf{R}, r_\epsilon, \zeta$

**Output:**  $\mathbf{h}(s)$

- 1: Initialize  $\mathbf{h}(0)$  as a complex Gaussian random vector,  $\mathbf{P}(0)$  as  $100\mathbf{I}_{\beta_1^2 \beta_2^2}$
- 2: **while**  $s \leq S$  **do**
- 3:   Calculate (27)–(31)
- 4:    $\tau = 0$ ;
- 5:   **while**  $\left| \left\| \hat{\mathbf{h}}_s(\tau) \right\|_1 - \left\| \hat{\mathbf{h}}_s(\tau - 1) \right\|_1 \right| < \zeta$  **do**
- 6:      $\tau = \tau + 1$ ;
- 7:      $\eta_s(\tau) = 1$ ;
- 8:     Calculate (32)–(36)
- 9:   **end while**
- 10: **end while**

The linearized pseudo-measurement is constructed as

$$0 = \mathbf{b}(s)\mathbf{h}(s) - \epsilon. \quad (24)$$

Pseudo-measurement (24) and pilot tone observation (19) cannot be utilized simultaneously to construct a complete filtering algorithm. The complete algorithm suffers from a long convergence time and a large reconstruction error. Consequently, two-stage correction is utilized. The first stage is a common five-step KF procedure using (13) and (19), whereas the second stage is implemented to iteratively correct the  $l_1$ -norm using (24). The ordinary sparse filtering algorithm is described in Algorithm 1.

### C. ACCELERATED SPARSE KALMAN FILTER ALGORITHM

To reduce the  $l_1$ -norm of the angular gain vector rapidly with a few pilot tone observations, an adaptive factor is introduced into the pseudo-measurement equation. This renders the  $l_1$ -norm correction to reach a stable value faster. Considering the dynamic model of (13), the sparse reconstruction problem is given by

$$\min \sum_{s=1}^S \left\{ \left\| \hat{\mathbf{h}}(s) - \mathbf{h}(s) \right\|_2^2 \right\} \quad \text{s.t.} \quad \left\| \hat{\mathbf{h}}(s) \right\|_1 \leq \eta(s)\epsilon. \quad (25)$$

The pseudo-measurement equation is established as

$$0 = \|\mathbf{h}(s)\|_1 - \eta(s)\epsilon, \quad (26)$$

where  $\eta(s)$  is another factor introduced to tune the variance ( $r_\epsilon$ ) to  $\eta^2(s)r_\epsilon$ . This is intended to force the  $l_1$ -norm to reduce at a high speed in early iterations. To guarantee that the  $l_1$ -norm is convergent,  $\eta(s)$  should be designed to increase gradually with  $s$ . The factor should reach one when  $s$  tends toward infinity [46]. The typical function for  $\eta(s)$  is chosen as  $\eta(s) = 1 - \eta(0)\exp(-\mu s/T)$ , where  $\eta(0)$ ,  $T$ , and  $\mu$  are the profile parameters used to adjust the  $l_1$ -norm correction procedure.

The filtering process follows the same stages as those of O-SKF. It comprises the state correction by pilot tone

observation and the  $l_1$ -norm correction by the current pseudo-measurement. The procedure is as follows.

In the first stage, the one-step state prediction is given by

$$\mathbf{h}(s|s-1) = \psi \mathbf{h}(s-1). \quad (27)$$

The one-step prediction of estimate error covariance is expressed as

$$\mathbf{P}(s|s-1) = \psi^2 \mathbf{P}(s-1) + (1 - \psi^2) \mathbf{Q}. \quad (28)$$

The gain matrix is given by

$$\mathbf{K}(s) = \mathbf{P}(s|s-1) \mathbf{A}^H(s) (\mathbf{A}(s) \mathbf{P}(s|s-1) \mathbf{A}^H(s) + \mathbf{R})^{-1}. \quad (29)$$

Thus, the current estimate is updated as

$$\hat{\mathbf{h}}(s) = \mathbf{h}(s|s-1) + \mathbf{K}(s)(\mathbf{y}(s) - \mathbf{A}(s)\mathbf{h}(s|s-1)). \quad (30)$$

The estimate error covariance is updated as

$$\mathbf{P}(s) = \mathbf{P}(s|s-1) - \mathbf{K}(s) \mathbf{A}(s) \mathbf{P}(s|s-1). \quad (31)$$

In the second stage, the  $l_1$ -norm correction of  $\hat{\mathbf{h}}(s)$  is performed in an iterative manner. It starts with the current estimate ( $\hat{\mathbf{h}}(s)$ ) in (30). The  $\hat{\mathbf{h}}(s)$  is regarded as  $\mathbf{h}_s(0)$ , and the estimated error covariance ( $\mathbf{P}(s)$ ) in (31) is regarded as  $\mathbf{P}_s(0)$ . The one-step  $l_1$ -norm prediction is as follows:

$$\mathbf{h}_s(\tau|\tau-1) = \mathbf{h}_s(\tau-1). \quad (32)$$

The estimate error covariance prediction is given by

$$\mathbf{P}_s(\tau|\tau-1) = \mathbf{P}_s(\tau-1). \quad (33)$$

Thus, the  $l_1$ -norm correction gain is obtained as

$$\begin{aligned} \mathbf{k}_s(\tau) &= \mathbf{P}_s(\tau-1) \mathbf{b}_s^H(\tau) (\mathbf{b}_s(\tau) \mathbf{P}_s(\tau-1) \mathbf{b}_s^H(\tau) + \eta^2(\tau) r_\epsilon)^{-1}. \end{aligned} \quad (34)$$

The estimate is updated as

$$\hat{\mathbf{h}}_s(\tau) = \mathbf{h}_s(\tau-1) - \mathbf{k}_s(\tau) \mathbf{b}_s(\tau) \mathbf{h}_s(\tau). \quad (35)$$

The estimate error covariance is updated as

$$\mathbf{P}_s(\tau) = \mathbf{P}_s(\tau-1) - \mathbf{k}_s(\tau) \mathbf{b}_s(\tau) \mathbf{P}_s(\tau-1). \quad (36)$$

In the second stage of the  $l_1$ -norm correction, the process is performed with each SMV. Let  $\eta(\tau) = 1 - \exp(-\mu\tau/T)$ . The halting condition is determined by the absolute difference between two adjacent  $l_1$ -norm values. In other words,  $\left| \|\hat{\mathbf{h}}_s(\tau)\|_1 - \|\hat{\mathbf{h}}_s(\tau-1)\|_1 \right| < \zeta$ , where  $\zeta$  is a small positive value. After processing with the current SMV,  $\mathbf{h}_s(\tau)$  and  $\mathbf{P}_s(\tau)$  are considered the current estimates of  $\mathbf{h}(s)$  and  $\mathbf{P}(s)$ , respectively. Subsequently, filtering continues when the next SMV is available. The proposed algorithm is described in Algorithm 2.

#### Algorithm 2 A-SKF for Angular Gain Recovery

**Input:**  $\psi, \mathbf{y}(s), \mathbf{A}(s), \mathbf{Q}, \mathbf{R}, r_\epsilon, \eta(0), \mu, T, \zeta$

**Output:**  $\mathbf{h}(s)$

- 1: Initialize  $\mathbf{h}(0)$  as a complex Gaussian random vector,  $\mathbf{P}(0)$  as  $100 \mathbf{I}_{\beta_1^2 \beta_2^2}$
- 2: **while**  $s \leq \mathcal{S}$  **do**
- 3:   Calculate (27)-(31)
- 4:    $\tau = 0$ ;
- 5:   **while**  $\left| \|\hat{\mathbf{h}}_s(\tau)\|_1 - \|\hat{\mathbf{h}}_s(\tau-1)\|_1 \right| < \zeta$  **do**
- 6:      $\tau = \tau + 1$ ;
- 7:      $\eta_s(\tau) = 1 - \eta(0) \exp(-\mu\tau/T)$
- 8:     Calculate (32)-(36)
- 9:   **end while**
- 10: **end while**

#### D. COMPUTATIONAL COMPLEXITY

The computational complexity is evaluated by the floating-point operation (FLOP) for addition, multiplication, and square-root calculations. The inverse of a square matrix is calculated using the Cholesky decomposition. The computational complexity is closely related to the dimensions of the state and measurement, which are  $\beta_1^2 \beta_2^2$  and  $N_R N_T$ , respectively. According to the details in [50] for a real domain model with a high state dimension ( $a$ ) and a low measurement dimension ( $b$ ), in the one-step implementation of (27)–(31), the addition FLOP is  $a^3 + 0.5b^3 + 3a^2b + 2ab^2 - 0.5b^2 - ab$ , the multiplication FLOP is  $a^3 + 0.5b^3 + 3a^2b + 2ab^2 - 3a^2 + 3.5b^2 + 2ab$ , and the square-root FLOP is  $b$ . Each FLOP is doubled in the complex-domain channel calculations. The FLOPs of (32)–(36) and (23) are ignored because the pseudo-measurement yields a numerical value. The inversion calculation in (34) is reciprocal. Hence, the computational FLOP of the filtering algorithm is mainly dominated by  $\mathcal{O}(2S\beta_1^6\beta_2^6)$ . The complexity of the OMP algorithm is  $\mathcal{O}(N_T N_R N_I S)$  [22]. Although the complexity of SKF is higher than that of an OMP algorithm, the filtering framework improves the NMSE, as shown in the simulations.

#### V. SIMULATION AND ANALYSIS

In the simulation, the algorithm performance was first evaluated in terms of the NMSE, estimation convergence procedure, support capture, and pilot overhead when  $\psi = 1$ . After that, the performances of the two SKF-based algorithms in dynamic scenarios with  $\psi < 1$  were compared with respect to the NMSE, tracking ability, and the  $l_1$ -norm convergence speed. For the evaluation, mmWave channel models (1) and (2) were adopted for both the UE-RIS and RIS-BS channels. The parameters of AoAs/AoDs, along with the corresponding angular gains in two scenarios, are listed in Table 1. The angular dictionaries are constructed using the same parameters as those in Table 1. Each antenna array is  $2 \times 2$  the planar one, and  $N_R = N_T = N_I = 4$ . The RIS phase shift vector is periodically set

TABLE 1. The cascaded channel parameters for the two scenarios.

$S_1$	$\varphi_{1i}^r$	$\theta_{1i}^r$	$\varphi_{2i}^t$	$\theta_{2i}^t$	gain
$H_1$	2.9864	-0.4996	0.9045	-0.8301	-0.0528 + 4.3276j
	1.5948	0.1456	1.293	-1.2185	-1.7249 + 0.9577j
	2.4526	0.4885	0.6166	0.005	0
	-0.0079	-1.2644	2.3942	-0.0711	0
$H_2$	0.6116	-1.5021	0.3065	-1.3788	-1.9978 + 0.4179j
	-2.6687	0.3105	-1.9356	-0.6088	0.4492 - 1.8453j
	-1.3321	0.5235	-2.6555	-0.2679	-0.0393 - 1.8219j
	2.7529	-1.5417	-1.0252	0.547	0
$S_2$	$\varphi_{1i}^r$	$\theta_{1i}^r$	$\varphi_{2i}^t$	$\theta_{2i}^t$	gain
$H_1$	-2.6157	1.1496	-1.2967	1.5031	-1.0047 - 0.9579j
	-0.603	-0.2335	-3.1013	0.7124	0
	-0.366	1.4003	-2.5026	0.2968	0
	2.735	0.1947	-0.674	0.7231	0
$H_2$	-0.5031	-1.0512	2.4332	0.627	-5.8373 - 1.8592j
	0.8371	-1.3346	-0.1097	-1.3795	0
	2.6054	-1.3771	1.081	1.3171	0
	2.7913	1.3105	2.2873	1.2354	0

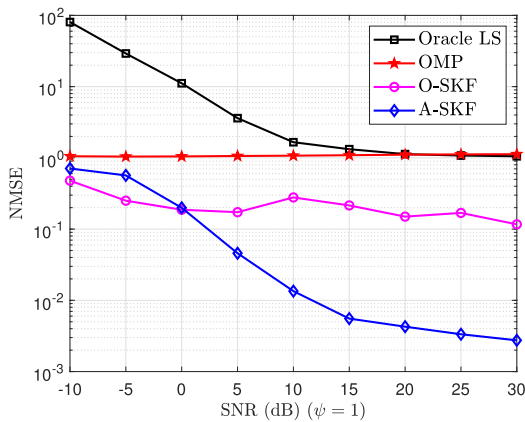


FIGURE 4. NMSE versus SNR ( $S_1$ ).

to  $(1, 1, 1, 1)$ ,  $(1, j, -1, -j)$ ,  $(1, -1, 1, -1)$ ,  $(1, -j, -1, j)$  in MMV. The A-SKF parameters are  $\mathbf{P}(0) = 100\mathbf{I}_{\beta_1^2\beta_2^2}$ ,  $\mathbf{Q} = \sqrt{1 - \psi^2}\mathbf{I}_{\beta_1^2\beta_2^2}$ ,  $\mathbf{R} = \mathbf{I}_{N_T N_R}$ , and  $r_\epsilon = 40000$ . The profile parameters for  $\eta(\tau)$  are  $\eta(0) = 0.9$ ,  $T = 10$ , and  $\mu = 0.02$ . The initial O-SKF parameters were the same as those for A-SKF. The estimates of Oracle LS and OMP were averaged using the MMV. The estimate in the last training frame ( $S$ ) is utilized for O-SKF and A-SKF. NMSE is defined as

$$NMSE(\hat{\mathbf{h}}(S)) = \frac{\mathbb{E}\{\|\mathbf{h}(S) - \hat{\mathbf{h}}(S)\|_2^2\}}{\mathbb{E}\{\|\mathbf{h}(S)\|_2^2\}}. \quad (37)$$

A. PERFORMANCE ANALYSIS in STATIC SCENARIOS

1)  $S_1$ :  $L_1 = 2$ ,  $L_2 = 3$ .

The NMSE curves for  $\hat{\mathbf{h}}(S)$  versus the SNR for Oracle LS, OMP, O-SKF, and A-SKF are shown in Fig. 4 using 100 MMV. Evidently, the NMSE of Oracle LS is extremely high; thus, it fails to perform CE. The NMSE curve of the OMP was flat, owing to incorrect support capture. The A-SKF achieved the best NMSE at SNR values greater than 5 dB. The NMSE gap between the O-SKF and A-SKF is large because the O-SKF is unable to reach its stable state at 100 MMV.

The estimation convergence procedure curves of the SKF-based algorithms at SNR = 10 dB are shown in Fig. 5 for

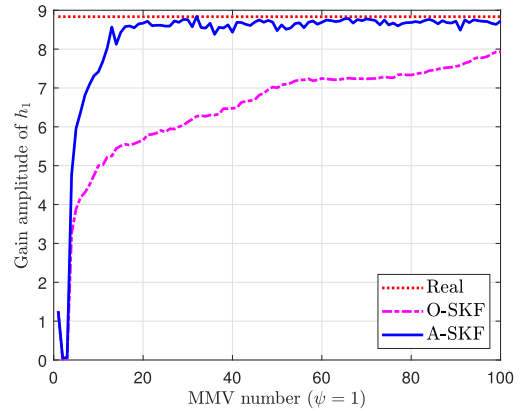


FIGURE 5. Estimation procedure versus the MMV number ( $S_1$ ).

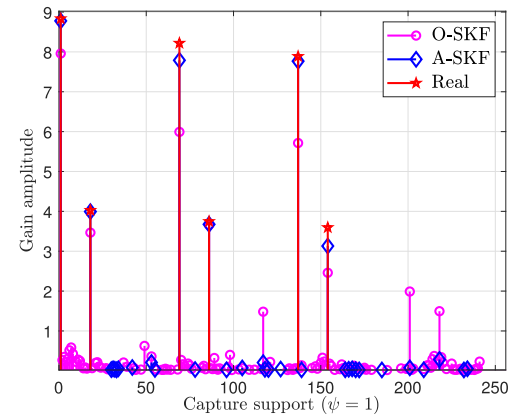


FIGURE 6. Capture of angular support ( $S_1$ ).

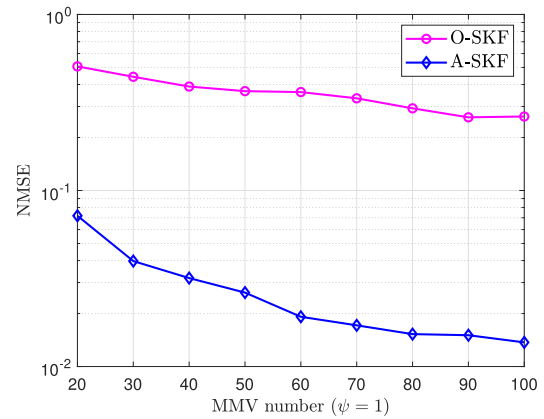


FIGURE 7. NMSE versus the MMV number ( $S_1$ ).

a complex angular gain denoted as  $h_1$ . As can be observed, the transition process of A-SKF is faster than that of O-SKF. A-SKF only needs 20 MMV to reach a stable state, and O-SKF requires more MMV owing to its slow convergence. This phenomenon indicates that the pilot overhead can be as small as 20 pilot blocks using A-SKF. Therefore, it is beneficial to perform CE using A-SKF.

The support capture result of A-SKF for  $\mathbf{h}(S)$  at SNR = 10 dB, compared with that of O-SKF, and the real support are shown in Fig. 6. The A-SKF performance is quite good when the SNR is as low as 10 dB. Moreover, the corresponding angular gains were estimated accurately. The capture error of



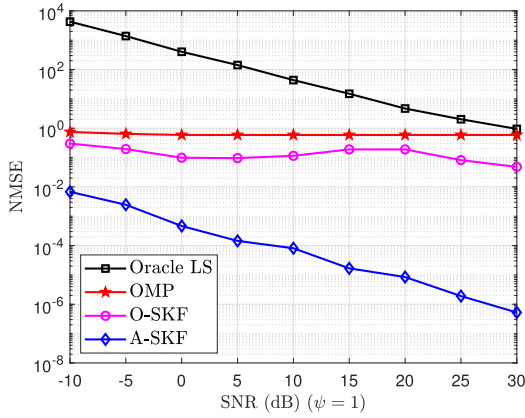


FIGURE 8. NMSE versus SNR ( $S_2$ ).

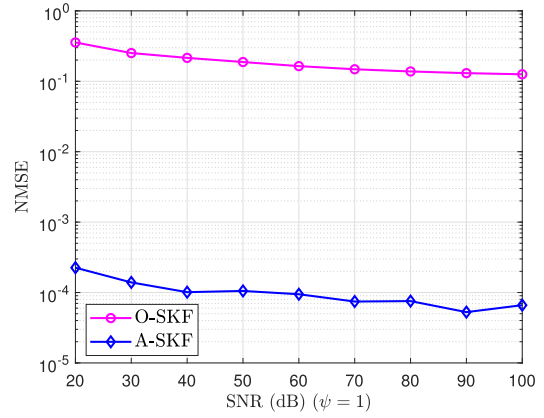


FIGURE 11. NMSE versus the MMV number ( $S_2$ ).

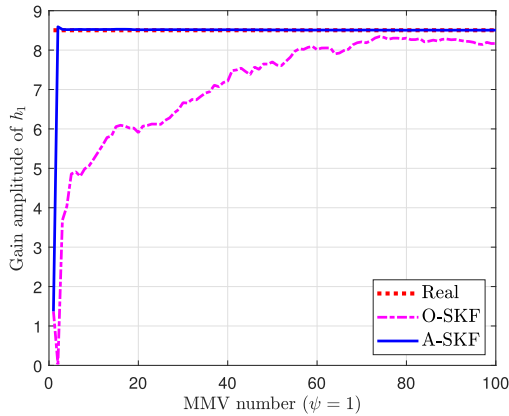


FIGURE 9. Estimation procedure versus the MMV number ( $S_2$ ).

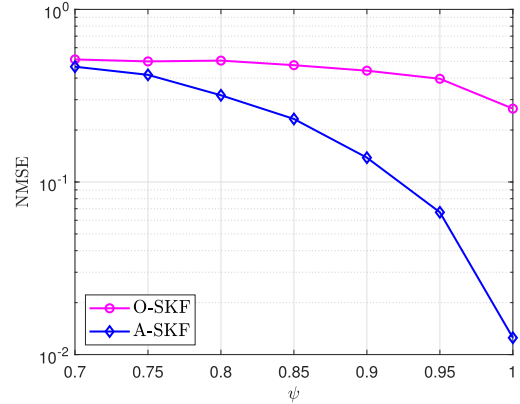


FIGURE 12. NMSE versus  $\psi$  ( $S_1$ ).

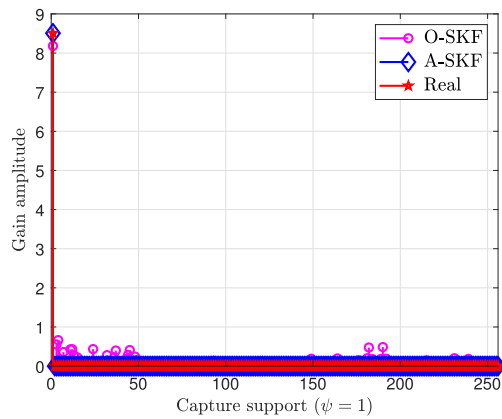


FIGURE 10. Capture of angular support ( $S_2$ ).

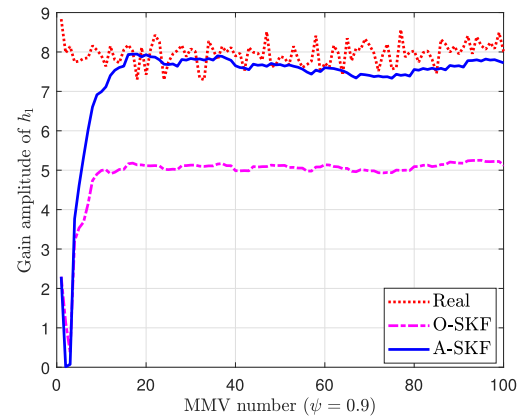


FIGURE 13. Dynamic tracking of the angular gain ( $S_1$ ).

A-SKF was significantly lower than that of O-SKF. Although errors occur because of the  $l_1$ -norm constraint approach, the support capture of A-SKF is better than those of Oracle LS and OMP, which are unable to capture the support in this case. This indicates that accurate AoAs/AoDs, along with the angular gains, can be estimated by SKF-type algorithms.

In Fig. 7, the curves of NMSE versus the MMV numbers of O-SKF and A-SKF are shown. As can be observed, the NMSE decreases with the MMV number. A-SKF requires fewer pilot blocks than O-SKF to achieve a tolerant NMSE.

In addition to the results, the profile parameters of the adaptive factor ( $\eta(\tau)$ ) are critical for the convergence of  $l_1$ -norm.

For the chosen exponentially increasing function, a large  $\eta(0)$  implies fast correction of the  $l_1$ -norm at early iterations with an SMV. A small  $\eta(0)$  suffers from slow convergence, and a suitable  $\eta(0)$  can be selected by balancing the sparsity and convergent speed. Moreover, parameter  $T$  should be less than the MMV number, and  $\mu$  can be selected according to the profile.

2)  $S_2$ :  $L_1 = 1, L_2 = 1$ .

In the scenario of one path in  $S_2$ , the results are shown in Fig. 8, Fig. 9, Fig. 10, and Fig. 11, where it can be observed that A-SKF still presents the best NMSE. A-SKF also achieves faster convergence than O-SKF. Notably, A-

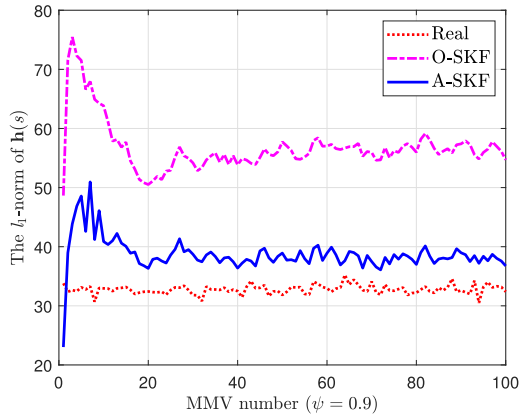


FIGURE 14. The  $l_1$ -norm of the angular gain vector ( $S_1$ ).

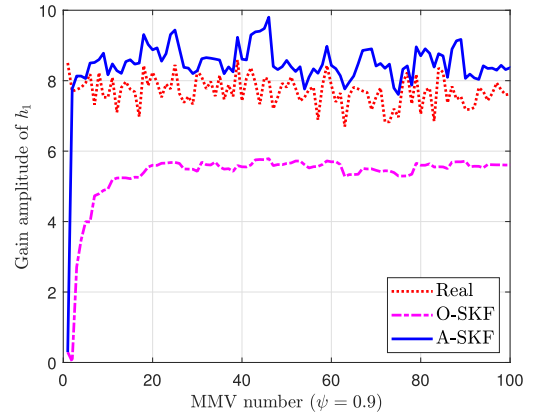


FIGURE 16. Dynamic tracking of the angular gain ( $S_2$ ).

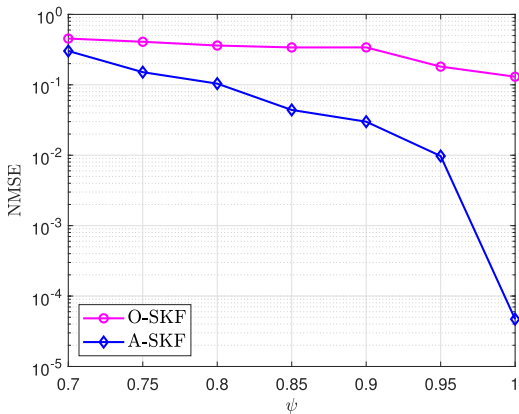


FIGURE 15. NMSE versus  $\psi$  ( $S_2$ ).

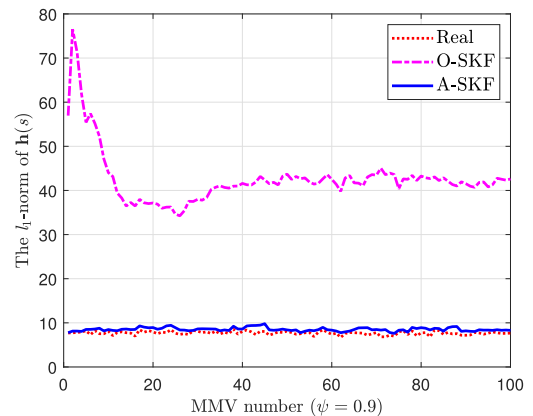


FIGURE 17. The  $l_1$ -norm of the angular gain vector ( $S_2$ ).

SKF can capture the support and accurately recover the angular gain. Similarly, O-SKF exhibits a small support error. Conversely, Oracle LS and OMP still failed to perform CE.

Simulations were also conducted in  $S_1$  and  $S_2$ , with a fixed RIS phase-shift vector  $((1, -j, -1, j))$ . A-SKF failed to capture the support in  $S_1$ , whereas it was successful in  $S_2$ . However, O-SKF was unable to capture the support in either scenario. The corresponding figures are not shown here for brevity. This indicates that a variant RIS phase-shift setting is necessary for the cascaded CE.

Based on the above results in static scenarios, A-SKF is suitable for CE as it can capture the support with low pilot overhead at low SNRs. Further, it can provide better beam information than other algorithms that ignore the temporal correlation of nonzero angular gains. The constructed pseudo-measurement by introducing an adaptive factor  $(\eta(s))$  can be implemented to enhance performance.

### B. PERFORMANCE ANALYSIS in DYNAMIC SCENARIOS

The angular gains listed in Table 1 were considered as the initial real parameters in dynamic scenarios. The gains vary recursively with  $s$  when  $\psi < 1$ .

1)  $S_1$ :  $L_1 = 2, L_2 = 3$ .

With the parameters in  $S_1$ , the NMSE curves versus  $\psi$  at SNR = 10 dB are shown in Fig. 12, where the MMV number

was 100. As can be observed, a small  $\psi$  corresponds to a large NMSE. Notably, A-SKF and O-SKF can capture the angular support, although more errors are obtained. In Fig. 13, the tracking curve of A-SKF for  $h_1(s)$  is shown in comparison with that of O-SKF, as well as the real gain. It was observed that A-SKF can track the channel to some extent. Fig. 14 shows the  $l_1$ -norm convergence curves of the two filtering algorithms for the angular gain vector. The transition process is almost the same as that in the static scenarios. This indicates that A-SKF does not require more pilot overhead in dynamic scenarios.

2)  $S_2$ :  $L_1 = 1, L_2 = 1$ .

Similar results in  $S_2$  are shown in Fig. 15, Fig. 16, and Fig. 17. As shown, A-SKF exhibits better performance in terms of NMSE and the necessary MMV number in comparison with O-SKF, and it successfully captured the support. Additionally, the convergence speed of A-SKF is faster than that of O-SKF. The pilot overhead can be reduced to less than 10 blocks for a certain CE.

From the simulation results, it is concluded that cascaded CE can be accomplished by SKF-type algorithms to find implicit AoAs/AoDs and angular gains. A-SKF can acquire accurate angular support at low SNRs when the available pilot overhead is low. Since SKF-type algorithms do not require any prior sparsity and support information, they can be implemented online in both static and dynamic scenarios.

## VI. CONCLUSION

A channel-estimation method based on the SKF framework is proposed for RIS-aided mmWave MIMO systems. The sparse angular gain vector sampled by overcomplete dictionaries is recovered through state-space modeling and block-diagonal pilot-pattern design. The  $l_1$ -norm pseudo-measurement equation is linearized by the deduced subdifferential of the sparse complex gain vector. In the proposed algorithm, an adaptive factor is introduced to control the noise covariance of the  $l_1$ -norm in the pseudo-measurement equation. Simulations demonstrated that the method has the advantages of a superior NMSE, accurate support capture, rapid convergence speed, and low pilot overhead. Its computational complexity is moderate, owing to the cascaded settings. The method is robust to angular dictionaries and noisy environments. It is suitable for implementation under both static and dynamic scenarios.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Zhangjie Peng and anonymous reviewers for their suggestions on the article.

## REFERENCES

- [1] J. Zhao, "A survey of intelligent reflecting surfaces (IRSs): Towards 6G wireless communication networks," 2019, *arXiv:1907.04789*.
- [2] M. Dajer, Z. Ma, L. Piazzzi, N. Prasad, X. Qi, B. Sheen, J. Yang, and G. Yue, "Reconfigurable intelligent surface: Design the channel—A new opportunity for future wireless networks," *Digit. Commun. Netw.*, vol. 8, no. 2, pp. 87–106, Nov. 2021.
- [3] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications—Part I: Fundamentals, solutions, and future opportunities," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1398–1402, May 2021.
- [4] G. T. de Araújo and A. L. F. de Almeida, "PARAFAC-based channel estimation for intelligent reflective surface assisted MIMO system," in *Proc. IEEE 11th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2020, pp. 1–5, doi: [10.1109/SAM48682.2020.9104260](https://doi.org/10.1109/SAM48682.2020.9104260).
- [5] L. Wei, C. Huang, G. C. Alexandropoulos, and C. Yuen, "Parallel factor decomposition channel estimation in RIS-assisted multi-user MISO communication," in *Proc. IEEE 11th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2020, pp. 1–5, doi: [10.1109/SAM48682.2020.9104305](https://doi.org/10.1109/SAM48682.2020.9104305).
- [6] L. Mo, F. Saggese, X. Lu, Z. Wang, and P. Popovski, "Direct tensor-based estimation of broadband mmWave channels with RIS," *IEEE Commun. Lett.*, vol. 27, no. 7, pp. 1849–1853, Jul. 2023, doi: [10.1109/LCOMM.2023.3270513](https://doi.org/10.1109/LCOMM.2023.3270513).
- [7] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, Jun. 2021.
- [8] S. Yang, W. Lyu, Y. Xiu, Z. Zhang, and C. Yuen, "Active 3D double-RIS-aided multi-user communications: Two-timescale-based separate channel estimation via Bayesian learning," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3605–3620, Jun. 2023, doi: [10.1109/TCOMM.2023.3265115](https://doi.org/10.1109/TCOMM.2023.3265115).
- [9] H. Liu, X. Yuan, and Y. A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2621–2636, Nov. 2020.
- [10] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.
- [11] E. Shtaiwi, H. Zhang, S. Vishwanath, M. Youssef, A. Abdelhadi, and Z. Han, "Channel estimation approach for RIS assisted MIMO systems," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 2, pp. 452–465, Jun. 2021.
- [12] S. Xia, Y. Shi, Y. Zhou, and X. Yuan, "Reconfigurable intelligent surface for massive connectivity: Joint activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 69, pp. 5693–5707, 2021.
- [13] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, 2020.
- [14] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications—Part II: An improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, May 2021.
- [15] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, early access, Feb. 24, 2023, doi: [10.1109/TWC.2023.3246264](https://doi.org/10.1109/TWC.2023.3246264).
- [16] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for IRS-assisted millimeter-wave MIMO systems: Sparsity-inspired approaches," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4078–4092, Jun. 2022.
- [17] Y. Liu, S. Zhang, F. Gao, J. Tang, and O. A. Dobre, "Cascaded channel estimation for RIS assisted mmWave MIMO transmissions," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 2065–2069, Sep. 2021.
- [18] S. Ma, W. Shen, J. An, and L. Hanzo, "Wideband channel estimation for IRS-aided systems in the face of beam squint," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6240–6253, Oct. 2021.
- [19] Z. Wan, Z. Gao, and M.-S. Alouini, "Broadband channel estimation for intelligent reflecting surface aided mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6, doi: [10.1109/ICC40277.2020.9149146](https://doi.org/10.1109/ICC40277.2020.9149146).
- [20] J. He, M. Leinonen, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6, doi: [10.1109/GLOBECOM42002.2020.9348112](https://doi.org/10.1109/GLOBECOM42002.2020.9348112).
- [21] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems via atomic norm minimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5786–5797, Sep. 2021.
- [22] H. Liu, J. Zhang, Q. Wu, H. Xiao, and B. Ai, "ADMM based channel estimation for RISs aided millimeter wave communications," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 2894–2898, Sep. 2021.
- [23] C. L. Miller, P. A. Dmochowski, and P. J. Smith, "Efficient channel estimation for RIS," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6, doi: [10.1109/ICC42927.2021.9500469](https://doi.org/10.1109/ICC42927.2021.9500469).
- [24] Y. Xu, H. Chu, and P. Xu, "Joint channel estimation and passive beamforming for reconfigurable intelligent surface aided multi-user massive MIMO system," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, May 2021, pp. 1–3, doi: [10.1109/BLACKSEACOM52164.2021.9527799](https://doi.org/10.1109/BLACKSEACOM52164.2021.9527799).
- [25] R. Schroeder, J. He, and M. Juntti, "Passive RIS vs. Hybrid RIS: A comparative study on channel estimation," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Apr. 2021, pp. 1–7, doi: [10.1109/VTC2021-SPRING51267.2021.9448802](https://doi.org/10.1109/VTC2021-SPRING51267.2021.9448802).
- [26] G. Lee, H. Lee, J. Oh, J. Chung, and J. Choi, "Channel estimation for reconfigurable intelligent surface with a few active elements," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8170–8174, Jun. 2023, doi: [10.1109/TVT.2023.3241928](https://doi.org/10.1109/TVT.2023.3241928).
- [27] R. Jiang, Z. Fei, S. Huang, X. Wang, Q. Wu, and S. Ren, "Bivariate pilot optimization for compressed channel estimation in RIS-assisted multiuser MISO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9115–9130, Jul. 2023, doi: [10.1109/TVT.2023.3250252](https://doi.org/10.1109/TVT.2023.3250252).
- [28] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, Oct. 2020.
- [29] J. An, C. Xu, L. Gan, and L. Hanzo, "Low-complexity channel estimation and passive beamforming for RIS-assisted MIMO systems relying on discrete phase shifts," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1245–1260, Feb. 2022.
- [30] C. Qing, L. Wang, L. Dong, and J. Wang, "Enhanced ELM based channel estimation for RIS-assisted OFDM systems with insufficient CP and imperfect hardware," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 153–157, Jan. 2022.
- [31] K. Kim, Y. K. Tun, M. S. Munir, W. Saad, and C. S. Hong, "Deep reinforcement learning for channel estimation in RIS-aided wireless networks," *IEEE Commun. Lett.*, early access, May 29, 2023, doi: [10.1109/LCOMM.2023.3280821](https://doi.org/10.1109/LCOMM.2023.3280821).
- [32] C. Nguyen, T. M. Hoang, and A. A. Cheema, "Channel estimation using CNN-LSTM in RIS-NOMA assisted 6G network," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 1, pp. 43–60, 2023.

- [33] Z. Wan, Z. Gao, F. Gao, M. D. Renzo, and M.-S. Alouini, "Terahertz massive MIMO with holographic reconfigurable intelligent surfaces," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4732–4750, Jul. 2021.
- [34] Z. Mao, M. Peng, and X. Liu, "Channel estimation for reconfigurable intelligent surface assisted wireless communication systems in mobility scenarios," *China Commun.*, vol. 18, no. 3, pp. 29–38, Mar. 2021.
- [35] Y. Liu, M. Chen, C. Pan, Y. Pan, Y. Wang, Y. Huang, T. Cao, and J. Wang, "Channel tracking for RIS-aided mmWave communications under high mobility scenarios," *IEEE Commun. Lett.*, vol. 27, no. 5, pp. 1397–1401, May 2023.
- [36] Z. Li, W. Yuan, B. Li, J. Wu, C. You, and F. Meng, "Reconfigurable intelligent surface aided OTFS: Transmission scheme and channel estimation," *IEEE Internet Things J.*, early access, Apr. 25, 2023, doi: [10.1109/JIOT.2023.3270335](https://doi.org/10.1109/JIOT.2023.3270335).
- [37] E. Baidoo, J. Hu, and L. Zhan, "Kalman filtering method for sparse off-grid angle estimation for bistatic multiple-input multiple-output radar," *IET Radar, Sonar Navigat.*, vol. 14, no. 2, pp. 313–319, Feb. 2020.
- [38] S. Gao, X. Cheng, and L. Yang, "Estimating doubly-selective channels for hybrid mmWave massive MIMO systems: A doubly-sparse approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5703–5715, Sep. 2020.
- [39] S. Srivastava and A. K. Jagannatham, "Sparse Bayesian learning-based Kalman filtering (SBL-KF) for group-sparse channel estimation in doubly selective mmWave hybrid MIMO systems," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5, doi: [10.1109/SPAWC.2019.8815509](https://doi.org/10.1109/SPAWC.2019.8815509).
- [40] Y. Chen, Y. Wang, and Z. Wang, "Reconfigurable intelligent surface aided high-mobility millimeter wave communications with dynamic dual-structured sparsity," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4580–4599, Jul. 2022.
- [41] 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.
- [42] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
- [43] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [44] 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.
- [45] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.
- [46] D. A. Hage, M. Heredia Conde, and O. Loffeld, "Sparse signal recovery via Kalman-filter-based  $\ell_1$  minimization," *Signal Process.*, vol. 171, Jun. 2020, Art. no. 107487.
- [47] D. Wei, D. Li, and J. Huang, "Improved force identification with augmented Kalman filter based on the sparse constraint," *Mech. Syst. Signal Process.*, vol. 167, Mar. 2022, Art. no. 108561.
- [48] W. Feng, Q. Li, and Q. Lu, "Force localization and reconstruction based on a novel sparse Kalman filter," *Mech. Syst. Signal Process.*, vol. 144, Oct. 2020, Art. no. 106890.
- [49] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2405–2409, Apr. 2010.
- [50] G. Xu, Y. Huang, Z. Gao, and Y. Zhang, "A computationally efficient variational adaptive Kalman filter for transfer alignment," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13682–13693, Nov. 2020.



**JING ZHANG** received the B.S. degree from the University of Shanghai for Science and Technology, China, in 1992, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, in 2004. She was a Visiting Scholar with the University of Wyoming, USA, from February 2007 to June 2007. She was a Postdoctoral Researcher with the Department of Electronic Engineering, Shanghai Jiao Tong University, from 2007 to 2009. She is currently an Associate Professor with the College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai. Her research interests include massive MIMO, relay transmission, channel estimation, and signal detection.



**YING SU** received the Ph.D. degree from Northwestern Polytechnical University, Xi'an, China, in 2004. She is currently a Researcher with the College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai. Her research interests include wireless communication, modeling, simulation, and evaluation of complex systems.



**DONGJUN QIAN** received the M.S. degree from Shanghai Normal University, China, in 2013. He is the Technical Director of the Big Data Department, China Telecommunication Shanghai Ideal Information Industry (Group) Company Ltd. He mainly engages in big data-related works, including the development and construction of big data centers, analysis and mining of massive data, and natural language processing.

...