Leveraging the Restricted Isometry Property: Improved Low-Rank Subspace Decomposition for Hybrid Millimeter-Wave Systems

Wei Zhang, *Student Member, IEEE*, Taejoon Kim, *Member, IEEE*, David J. Love, *Fellow, IEEE*, and Erik Perrins, *Senior Member, IEEE*

Abstract—Communication at millimeter wave frequencies will be one of the essential new technologies in 5G. Acquiring an accurate channel estimate is the key to facilitate advanced millimeter wave hybrid multiple-input multiple-output (MIMO) precoding techniques. Millimeter wave MIMO channel estimation, however, suffers from a considerably increased channel use overhead. This happens due to the limited number of radio frequency (RF) chains that prevent the digital baseband from directly accessing the signal at each antenna. To address this issue, recent research has focused on adaptive closed-loop and two-way channel estimation techniques. In this paper, unlike the prior approaches, we study a non-adaptive, hence rather simple, open-loop millimeter wave MIMO channel estimation technique. We present a random phase rotation design of channel subspace sampling signals and show that they obey the restricted isometry property (RIP) with high probability. We then formulate the channel estimation as a low-rank subspace decomposition problem and, based on the RIP, show that the proposed framework reveals resilience to a low signal-to-noise ratio. It is revealed that the required number of channel uses ensuring a bounded estimation error is linearly proportional to the degrees of freedom of the channel, whereas it converges to a constant value if the number of RF chains can grow proportionally to the channel dimension while keeping the channel rank fixed. In particular, we show that the tighter the RIP characterization the lower the channel estimation error is. We also devise an iterative technique that effectively finds a suboptimal, but stationary, solution to the formulated problem. The proposed technique is shown to have improved channel estimation accuracy with a substantially low channel use overhead as compared to that of previous closed-loop and two-way adaptation techniques.

Index Terms—Millimeter-wave, hybrid multiple-input multiple-output (MIMO), restricted isometry property (RIP), channel estimation, low-rank subspace decomposition, mean squared error (MSE), open-loop.

I. INTRODUCTION

WITH the growth in the demand for wireless broadband, the operating frequency of modern wireless systems is steadily shifting upward to the millimeter wave band to provide a much wider bandwidth [2]. In the millimeter wave bands, the radio channel experiences severe pathloss that is compensated for by using large-scale antenna arrays [3]–[5]. To make the system available at low cost, large-sized analog arrays that are fed by a limited number of radio frequency (RF) chains are popularly discussed. This system is often referred to as a millimeter wave hybrid multiple-input multiple-output (MIMO) system. Despite the limited RF chains, a remarkable throughput boost is possible if the system employs advanced hybrid MIMO precoding techniques [5]–[9]. To enable such techniques, accurate channel state information (CSI) is the key.

Useful CSI has to be attained via channel estimation. However, millimeter wave hybrid MIMO channel estimation is limited by an insufficient number of RF chains. This limitation increases the channel use overhead required to extract a useful channel estimate of the system. To address this difficulty, recent research has focused on adaptive channel estimation frameworks, including adaptive closed-loop [5], [8], [10]–[12] and two-way [9], [13] channel estimation methods. Closed-loop channel estimation includes techniques such as the multi-level adaptive subspace search techniques [5], [8], [10]–[12], where the receiver at each search level conveys feedback to the transmitter to guide the next level subspace sampling signal design [14]–[16]. The two-way technique [9], [13] exploits the uplink-downlink channel reciprocity to iteratively refine the channel estimate via uplink-downlink channel echoing [17]–[20]. Improved estimation accuracy compared to other non-adaptive techniques, e.g., open-loop techniques, has been evidenced in [5], [8]–[13].

Nevertheless, these processes still require a relatively large number of channel uses because of the repeated channel sampling [5], [8]–[13]. Achieving lower channel use overhead is particularly important to the millimeter wave communications that are limited by a shortage of channel coherence

resources. Open-loop MIMO channel estimation has been thoroughly studied in the literature [21], [22]. It is non-adaptive and simple. However, it is impractical for estimating large-dimensional channels, such as sub-6 GHz massive MIMO channels [14]–[16], due to its inaccuracy. Meanwhile, there have been convincing experimental studies showing that the practical millimeter wave MIMO channel exhibits substantial rank sparsity [23], [24]. A general belief is that the sparsity of a signal, either in its support or subspace, provides tremendous clues when estimating it from distorted observations. Useful directions for open-loop millimeter wave MIMO channel estimation are recently suggested by the low-rank matrix reconstruction techniques [25]–[33].

Methods for low-rank matrix reconstruction have recently spurred considerable interest due to their exploitation of rank sparsity [26]–[33]. The reliability of low-rank matrix reconstruction is best described by the restricted isometry property (RIP) [26], [31]. The RIP is a probabilistic model for the norm convergence of a subspace-sampled matrix to its original matrix. It has also been used to analyze the estimation error performances of various algorithms [29], [31], [34]. The RIP concept is useful and can be exploited in devising new design criteria for the open-loop channel estimation. However, low-rank matrix reconstruction techniques [26]–[33], in their current form, face difficulties in estimating the millimeter wave MIMO channels. Reliable performance of low-rank matrix reconstruction techniques [26]–[33] is only ensured in the no noise or asymptotically high signal-to-noise ratio (SNR) regime. Reliable operation at low SNR is critical for millimeter wave systems that are limited by heavy mixed signal processing with an excessive power consumption [4]. Moreover, the probabilistic analysis of when the RIP holds has been studied for the subspace samples with independently and identically distributed (i.i.d.) Gaussian or Bernoulli entries [26], [31]. Yet, the subspace samples that the millimeter wave hybrid arrays generate are not well modeled by these distributions. Thus far, its RIP characterization has not been reported in the literature.

In this paper, we study an open-loop channel estimation technique by leveraging low-rank matrix reconstruction for millimeter wave hybrid MIMO systems. This is in contrast with prior approaches like [5] and [8]–[13] that focus on closed-loop and two-way techniques. We begin by introducing a random subspace sampling signal design by using analog arrays, which does not rely on any prior knowledge of channel subspace information. We adopt the precoder and combiner gain maximization criterion [35] to formulate the channel estimation as a low-rank subspace decomposition problem. We show through an analysis that the random phase rotation ensembles of the subspace sampling signals satisfy the RIP with high probability. On the basis of the established RIP, we carry out channel estimation error analysis and show that the proposed low-rank subspace decomposition reveals resilience to low SNRs.

In particular, we reveal that a tighter RIP bound characterization results in a lower estimation error for the devised low-rank subspace decomposition problem. It is also shown that the required number of channel uses ensuring a bounded

estimation error performance scales in proportional to the degrees of freedom of the channel when the number of RF chains is fixed, whereas it can converge to a constant value if the number of RF chains grows proportionally to the channel dimension while keeping the channel rank fixed. It is worthwhile to note that the proposed technique does not assume any explicit channel model and statistics, making it useful in a wide range of sparse scenarios. The subspace decomposition is a non-convex operation. To effectively solve the problem, we devise an alternating optimization technique that finds a suboptimal, but stationary, solution to the problem. The proposed technique results in improved channel estimation accuracy compared with prior approaches while consuming a considerably reduced channel use overhead.

This result is interesting since it is achieved without advanced adaptation techniques such as closed-loop or two-way adaptations [5], [8]–[12]. We are not aware of similar work establishing the RIP of a specific wireless signaling and estimation problem such as the low-rank millimeter wave MIMO channel estimation problem by drawing a connection between the RIP and estimation performance. Though we focus exclusively on millimeter wave MIMO channel estimation in this work, we hope that our approach will motivate the use of the RIP in other related communication problems.

The rest of the paper is organized as follows. Section II provides the system model and a brief review of the low-rank matrix reconstruction. Section III presents the low-rank subspace decomposition problem and the subspace sampling signal design. In Section IV, the RIP of the subspace sampling signals is analyzed. Section V presents the performance analysis and the algorithm to effectively solve the low-rank subspace decomposition problem. Finally, the numerical simulations and concluding remarks are presented in Sections VI and VII, respectively.

Notations: A bold lower case letter $\mathbf{a}$ is a vector, a bold capital letter $\mathbf{A}$ is a matrix. $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$, tr$(\mathbf{A})$, $|\mathbf{A}|$, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_*$, and $\|\mathbf{a}\|_2$ are, respectively, the transpose, conjuagate transpose, inverse, trace, determinant, Frobenius norm, nuclear norm (i.e., the sum of the singular values of $\mathbf{A}$) of $\mathbf{A}$, and l_2-norm of $\mathbf{a}$. $[\mathbf{A}]_{:,i}$, $[\mathbf{A}]_{i,:}$, $[\mathbf{A}]_{i,j}$, and $[\mathbf{a}]_i$ are, respectively, the ith column, ith row, ith row and jth column entry of $\mathbf{A}$, and ith entry of vector $\mathbf{a}$. vec$(\mathbf{A})$ stacks the columns of $\mathbf{A}$ and form a long column vector. diag$(\mathbf{A})$ extracts the diagonal entries of $\mathbf{A}$ to form a column vector. $\mathbf{I}_M \in \mathbb{R}^{M \times M}$ is the identity matrix. $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$.

II. System Model

We provide the signal model and briefly review the low-rank matrix reconstruction techniques under consideration.

A. Signal Model

Consider a point-to-point millimeter wave hybrid MIMO system. An independent block fading channel with a coherence block length T_c (channel uses) is assumed. The open-loop channel estimation is based on sounding a set of predefined channel subspace sampling signals from the transmitter to the receiver by using K ($K \ll T_c$) channel uses. We denote by
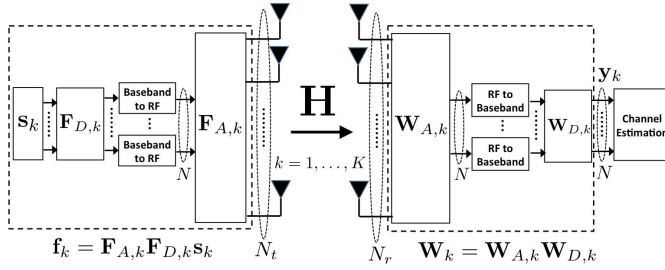
Fig. 1.   Open-loop millimeter wave hybrid MIMO channel sounding.

N_r, N_t, and N, respectively, the numbers of receive antennas, transmit antennas, and RF chains ($N \ll \min(N_r, N_t)$). Fig. 1 illustrates the open-loop channel sounding at the kth channel use, $k = 1, \ldots, K$. The received signal $\mathbf{y}_k$ is described as

$$\mathbf{y}_k = \mathbf{W}_k^H \mathbf{H} \mathbf{f}_k + \mathbf{W}_k^H \mathbf{n}_k, \tag{1}$$

where $\mathbf{W}_k \in \mathbb{C}^{N_r \times N}$, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, $\mathbf{f}_k \in \mathbb{C}^{N_t \times 1}$, and $\mathbf{n}_k \in \mathbb{C}^{N_r \times 1}$ are, respectively, the receive subspace sampling signal, channel, transmit subspace sampling signal, and the noise.

Seen from Fig. 1, the subspace sampling signals $\mathbf{W}_k$ and $\mathbf{f}_k$ in (1) are further divided into the analog and digital components, i.e.,

$$\mathbf{W}_k = \mathbf{W}_{A,k} \mathbf{W}_{D,k} \text{ and } \mathbf{f}_k = \mathbf{F}_{A,k} \mathbf{F}_{D,k} \mathbf{s}_k, \tag{2}$$

where $\mathbf{W}_{A,k} \in \mathbb{C}^{N_r \times N}$ ($\mathbf{F}_{A,k} \in \mathbb{C}^{N_t \times N}$) and $\mathbf{W}_{D,k} \in \mathbb{C}^{N \times N}$ ($\mathbf{F}_{D,k} \in \mathbb{C}^{N \times N}$) are the analog and digital receive (transmit) sampling signals, respectively. $\mathbf{s}_k \in \mathbb{C}^{N \times 1}$ is the transmitted symbol. The entries of $\mathbf{F}_{A,k}$ and $\mathbf{W}_{A,k}$ obey the constant modulus property such that $|[\mathbf{F}_{A,k}]_{m,n}| = 1/\sqrt{N_t}$ and $|[\mathbf{W}_{A,k}]_{m,n}| = 1/\sqrt{N_r}$, $\forall m, n$, representing the analog phased-array elements. We assume the entries of $\mathbf{n}_k$ are independent and identically distributed (i.i.d.) Gaussian with zero mean and σ^2 variance. The low-rank channel $\mathbf{H}$ has $\mathrm{rank}(\mathbf{H}) = L$ with $L \leq N \ll \min(N_r, N_t)$. With the power constraint $E[\|\mathbf{f}_k\|_2^2] = 1$, the SNR per channel use is given by $\frac{1}{\sigma^2}$.

For the generation of the subspace sampling signal pair $(\mathbf{W}_k, \mathbf{f}_k)$, we take a simple approach. The transmitter only activates one RF chain per channel use for channel sounding. This is to say the transmit subspace sampling signal is in the form $\mathbf{f}_k = \mathbf{f}_{A,k} f_{D,k} s_k$ where $\mathbf{f}_{A,k} \in \mathbb{C}^{N_t \times 1}$, $f_{D,k} \in \mathbb{C}$, and $s_k \in \mathbb{C}$ for simplicity. Moreover, we admit the following random generation,

$$[\mathbf{W}_{A,k}]_{m,n} = \frac{1}{\sqrt{N_r}} e^{j\eta_{k,m,n}}, \quad [\mathbf{f}_{A,k}]_n = \frac{1}{\sqrt{N_t}} e^{j\phi_{k,n}},$$
$$\mathbf{W}_{D,k} = \mathbf{I}_N, \ f_{D,k} = 1, \text{ and } s_k = 1, \quad \forall m, n, \tag{3}$$

where random variables $\{\eta_{k,m,n}\}$ and $\{\phi_{k,n}\}$ are mutually independent and i.i.d. according to the uniform distribution over $[0, 2\pi)$, i.e., $\eta_{k,m,n}, \phi_{k,n} \in \mathcal{U}[0, 2\pi)$, $\forall k, m, n$. The $\mathbf{f}_k = \mathbf{f}_{A,k} f_{D,k} s_k$ in (3) satisfies $E[\|\mathbf{f}_k\|_2^2] = 1$. Although simple, this random design, as we will show in the following sections, permits theoretical insights to conduct channel estimation performance analysis based on the RIP.

The digital baseband signal in (1) cannot directly access $\mathbf{H}$, but it instead observes a noisy version of the analog distorted effective channel $\mathbf{W}_k^H \mathbf{H} \mathbf{f}_k \in \mathbb{C}^{N \times 1}$ per channel use. Hence,

the channel sounding has to rely on the subspace sampling signal set $\mathcal{S} = \{(\mathbf{W}_k, \mathbf{f}_k)\}_{k=1}^K$. In general, observing the entire $\mathbf{H}$ using $\mathcal{S}$ requires $K = O(N_r N_t)$ channel uses. This overhead could overwhelm the channel coherence time T_c in millimeter wave bands, which is over ten-fold shorter than that in sub-6 GHz bands. On the other hand, rank sparsity of the channel matrix is an inherent characteristic of radio channels in the millimeter wave bands [23], [24].

B. Low-Rank Matrix Reconstruction

Low-rank matrix reconstruction has received much recent attention in the field of large-scale sparse signal processing [25]–[33]. In its problem formulation, distorted observations are modeled by using an affine map $\mathcal{A}$. For the sounding model in (1), we can define the affine map $\mathcal{A}_{\mathcal{S}} : \mathbb{C}^{N_r \times N_t} \mapsto \mathbb{C}^{KN \times 1}$ of $\mathbf{H}$ associated with the subspace sampling signal set $\mathcal{S} = \{(\mathbf{W}_k, \mathbf{f}_k)\}_{k=1}^K$ as

$$\mathcal{A}_{\mathcal{S}}(\mathbf{H}) = \left[(\mathbf{W}_1^H \mathbf{H} \mathbf{f}_1)^T, \ldots, (\mathbf{W}_K^H \mathbf{H} \mathbf{f}_K)^T \right]^T \in \mathbb{C}^{KN \times 1}. \tag{4}$$

Stacking $\{\mathbf{y}_k\}_{k=1}^K$ in (1) into a long column vector gives

$$\mathbf{y} = \mathcal{A}_{\mathcal{S}}(\mathbf{H}) + \tilde{\mathbf{n}}, \tag{5}$$

where $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T]^T \in \mathbb{C}^{KN \times 1}$ and the effective noise

$$\tilde{\mathbf{n}} = [(\mathbf{W}_1^H \mathbf{n}_1)^T, \ldots, (\mathbf{W}_K^H \mathbf{n}_K)^T]^T \in \mathbb{C}^{KN \times 1}. \tag{6}$$

We will use the notation $M \triangleq KN$ to indicate the total number of observations. In low-rank matrix reconstruction, popularly adopted approaches include the nuclear norm minimization (NNM) [26] and matrix factorization (MF) [32]–[34]. These approaches can directly be extended to estimate $\mathbf{H}$ in (5).

1) Nuclear Norm Minimization (NNM): The NNM problem in [26] is formulated as

$$\widehat{\mathbf{H}} = \underset{\mathbf{H}}{\operatorname{argmin}} \|\mathbf{H}\|_* \quad \text{subject to} \quad \mathbf{y} = \mathcal{A}_{\mathcal{S}}(\mathbf{H}). \tag{7}$$

This problem is convex and can be solved reliably. However, as seen from (7), it does not guarantee the low-rank solution. One critical drawback is that it does not scale to large-dimensional matrices due to prohibitively high computational complexity [1], [26].

2) Matrix Factorization (MF): The development of MF [32]–[34] has revealed a new path towards handling large-scale low-rank matrices. It achieves accurate reconstruction for various classes of matrices at high SNR [25]. Given (5), the MF problem can be formulated as

$$(\mathbf{L}^\star, \mathbf{R}^\star) = \underset{\mathbf{L} \in \mathbb{C}^{N_r \times L}, \mathbf{R} \in \mathbb{C}^{N_t \times L}}{\operatorname{argmin}} \left\| \mathbf{y} - \mathcal{A}_{\mathcal{S}}(\mathbf{L}\mathbf{R}^H) \right\|_2^2. \tag{8}$$

Solving (8) yields the estimate $\widehat{\mathbf{H}} = \mathbf{L}^\star(\mathbf{R}^\star)^H$. The problem in (8) is non-convex. A widely adopted suboptimal treatment is to use the alternating minimization with power factorization [33]. The complexity of the latter approach is much lower than that of NNM. Interestingly, MF outperforms NNM when the SNR is high and the matrix is substantially low-rank [1], [33], [34]. A critical drawback of MF is that it has very poor performance at low SNR [1]. This

is because the objective function in (8) is highly susceptible to noise. For example, detrimental over-matching can frequently happen when the noise overpowers the signal, i.e., $\|\tilde{\mathbf{n}}\|_2 \gg \|\mathcal{A}_{\mathcal{S}}(\mathbf{H})\|_2$ [1].

3) Restricted Isometry Property (RIP): A reliability guarantee of the low-rank matrix reconstruction is best described by the RIP [26], [31]. For any $\mathbf{H}$ whose rank is at most L and an arbitrary affine map $\mathcal{A}$, the RIP can be expressed as the bound,

$$\left| \|\mathcal{A}(\mathbf{H})\|_2^2 - \|\mathbf{H}\|_F^2 \right| \leq \delta_L \|\mathbf{H}\|_F^2, \tag{9}$$

where $0 \leq \delta_L < 1$ is the L-RIP constant [26]. The bound in (9) is justified when $\mathcal{A}(\mathbf{H})$ is as significant as if $\mathbf{H}$ were directly observed. The reliable reconstruction of NNM [26], [29] and MF [34] is guaranteed with high probability when the $\mathcal{A}(\mathbf{H})$ obeys the RIP. The RIP also provides an operational characterization that any $\mathcal{A}$ obeying (9) ensures a bounded estimation error for $K = O(L \cdot \max(N_r, N_t))$ [31].

C. General Statement of Technique

The desired open-loop millimeter wave MIMO channel estimation must leverage the sparse property of the channel while consuming far less channel use overhead than $K = O(N_r N_t)$. The RIP offers a useful statistical characterization that ensures a bounded estimation error performance while consuming a reduced amount of channel uses. However, the RIP has only been modeled for the $\mathcal{S}$ with i.i.d. Gaussian or Bernoulli entries [26]. The $\mathcal{S}$ that hybrid arrays generate based on the random design in (3), however, is in different random ensembles. Moreover, the conventional low-rank matrix reconstructions are critically limited at low SNR. For instance, the rank constraint in (8), i.e., $\text{rank}(\widehat{\mathbf{H}}) \leq L$, is insufficient to capture the sparsity under the noise. The original MF [32], [33] was originally devised for the noise-free scenario where the rank constraint was sufficient. Note that with the noise, the matrix becomes full rank. The desired open-loop millimeter wave MIMO channel estimation will require further stringent approaches to incorporate the sparsity under the noise.

III. LOW-RANK SUBSPACE DECOMPOSITION

We begin by addressing the questions related to how to enhance the low SNR performance. In MIMO transmission, high resilience to noise is conventionally achieved by matching the precoder and combiner to the channel subspace [35]. This points that when recovering a sparse matrix under the noise, constraints on the channel subspace would be more essential than the rank constraint in (8).

A. Problem Formulation: Subspace Constaraints

Consider the unitary precoding and combining gain maximization problem,

$$\left(\widehat{\mathbf{W}}, \widehat{\mathbf{F}}\right) = \underset{\mathbf{W}, \mathbf{F}}{\arg\max} \|\mathbf{W}^H \mathbf{H} \mathbf{F}\|_F^2$$
$$\text{subject to } \mathbf{W}^H \mathbf{W} = \mathbf{I}_d, \ \mathbf{F}^H \mathbf{F} = \mathbf{I}_d, \tag{10}$$

where $\mathbf{W} \in \mathbb{C}^{N_r \times d}$ and $\mathbf{F} \in \mathbb{C}^{N_t \times d}$ are semi-unitary matrices, and $d \leq L$ denotes the dimension of $\mathbf{W}$ and $\mathbf{F}$. Equation (10) can be written in an alternative form [1]

$$\left(\widehat{\mathbf{U}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\mathbf{V}}\right) = \underset{\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}}{\arg\min} \|\mathbf{H} - \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H\|_F^2$$
$$\text{subject to } \mathbf{U}^H \mathbf{U} = \mathbf{I}_d, \ \mathbf{V}^H \mathbf{V} = \mathbf{I}_d, \tag{11}$$

where $\mathbf{U} \in \mathbb{C}^{N_r \times d}$ and $\mathbf{V} \in \mathbb{C}^{N_t \times d}$. In (11), $\boldsymbol{\Sigma} \in \mathbb{C}^{d \times d}$ is a diagonal matrix.

Remark 1: The relation between (10) and (11) is that the optimal solution $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ in (11) will also be optimal for (10). However, $\widehat{\mathbf{W}}$ and $\widehat{\mathbf{F}}$ in (10) are not necessarily optimal for (11). An important implication of (10) and (11) is that maximizing the precoding gain is equivalent to approximating $\mathbf{H}$ to a subspace decomposition $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$.

Sampling $\mathbf{H}$ using $\mathcal{S} = \{(\mathbf{W}_k, \mathbf{f}_k)\}_{k=1}^K$, incorporating the affine map $\mathcal{A}_{\mathcal{S}}$ in (4), and including the effective noise in (6) to (11) yields the low-rank subspace decomposition problem,

$$\left(\widehat{\mathbf{U}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\mathbf{V}}\right) = \underset{\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}}{\arg\min} \|\mathbf{y} - \mathcal{A}_{\mathcal{S}}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H)\|_2^2$$
$$\text{subject to } \mathbf{U}^H \mathbf{U} = \mathbf{I}_d, \ \mathbf{V}^H \mathbf{V} = \mathbf{I}_d. \tag{12}$$

The significance of (12) compared to (8) is that now (12) explicitly contains the subspace constraints and $\boldsymbol{\Sigma}$ can be exploited to improve robustness at low SNR.

When the noise overpowers the signal, the objective in (12) is likely misled by $\|\tilde{\mathbf{n}} - \mathcal{A}_{\mathcal{S}}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H)\|_2^2$. This results in the critical over-matching problem [1]. As a remedy, we propose to control $\boldsymbol{\Sigma}$ to prevent over-matching by adding a condition, $\|\boldsymbol{\Sigma}\|_F^2 \leq \beta$, to (12) with β being a constant, yielding

$$\left(\widehat{\mathbf{U}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\mathbf{V}}\right) = \underset{\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}}{\arg\min} \|\mathbf{y} - \mathcal{A}_{\mathcal{S}}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H)\|_2^2$$
$$\text{subject to } \mathbf{U}^H \mathbf{U} = \mathbf{I}_d, \ \mathbf{V}^H \mathbf{V} = \mathbf{I}_d, \ \|\boldsymbol{\Sigma}\|_F^2 \leq \beta. \tag{13}$$

The channel estimate is given by $\widehat{\mathbf{H}} = \widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}^H$. For the selection of β, we first notice that the true channel $\mathbf{H}$ is feasible to the problem in (13). This being said, choosing the smallest β satisfying the condition $\|\mathbf{H}\|_F^2 \leq \beta$ is the desired choice. However, we will see in the simulation study in Section VI, that a rough guess of β is sufficient to provide robust estimation performance. Notice that the subspace decomposition in (13) is non-convex due to the coupled variables and the quadratic equality constraints. An algorithm for effective minimization of (13) will be discussed in Section V.

B. Random Subspace Sampling Signal

Based on the design in (3), inserting (3) in (1) gives

$$\mathbf{y}_k = \mathbf{W}_{A,k}^H \mathbf{H} \mathbf{f}_{A,k} + \mathbf{W}_{A,k}^H \mathbf{n}_k. \tag{14}$$

The ith entry of $\mathbf{W}_{A,k}^H \mathbf{H} \mathbf{f}_{A,k} \in \mathbb{C}^{N \times 1}$ can be rewritten as

$$[\mathbf{W}_{A,k}^H \mathbf{H} \mathbf{f}_{A,k}]_i = \text{tr}(\mathbf{f}_{A,k}[\mathbf{W}_{A,k}]_{:,i}^H \mathbf{H}) = \text{tr}(\mathbf{X}_{k,i}^H \mathbf{H}), \tag{15}$$

where $\mathbf{X}_{k,i} = [\mathbf{W}_{A,k}]_{:,i}\mathbf{f}_{A,k}^H \in \mathbb{C}^{N_r \times N_t}$, $1 \leq i \leq N$. Provided $M = KN$ observations, the affine map $\mathcal{A}_{\mathcal{S}}(\mathbf{H}) : \mathbb{C}^{N_r \times N_t} \to \mathbb{C}^{M \times 1}$ in (4) can be rewritten as

$$\mathcal{A}_{\mathcal{S}}(\mathbf{H}) = [\mathrm{tr}(\mathbf{X}_{1,1}^H\mathbf{H}), \ldots, \mathrm{tr}(\mathbf{X}_{1,N}^H\mathbf{H}), \mathrm{tr}(\mathbf{X}_{2,1}^H\mathbf{H}), \ldots$$
$$\ldots, \mathrm{tr}(\mathbf{X}_{K,1}^H\mathbf{H}), \ldots, \mathrm{tr}(\mathbf{X}_{K,N}^H\mathbf{H})]^T. \quad (16)$$

The following lemma characterizes the distribution of each entry of $\mathbf{X}_{k,i} \in \mathbb{C}^{N_r \times N_t}$.

Lemma 1: The entries of $\mathbf{X}_{k,i} = [\mathbf{W}_{A,k}]_{:,i}\mathbf{f}_{A,k}^H \in \mathbb{C}^{N_r \times N_t}$, $1 \leq i \leq N$ in (16) are i.i.d. and each entry follows

$$[\mathbf{X}_{k,i}]_{m,n} \stackrel{d}{=} \frac{1}{\sqrt{N_r N_t}}e^{j\theta}, \quad (17)$$

where $\stackrel{d}{=}$ denotes equality in distribution and $\theta \sim \mathcal{U}[0, 2\pi)$.

Proof: From (16) and (3), one can write

$$[\mathbf{X}_{k,i}]_{m,n} = [\mathbf{W}_{A,k}]_{m,i}[\mathbf{f}_{A,k}]_n^H$$
$$= \frac{1}{\sqrt{N_r N_t}}e^{j(\eta_{k,m,i} - \phi_{k,n})}, \quad (18)$$

where $\eta_{k,m,i}$ and $\phi_{k,n}$ are independent and $\eta_{k,m,i}, \phi_{k,n} \in \mathcal{U}[0, 2\pi)$. Let $\theta_t = \eta_{k,m,i} - \phi_{k,n}$. Then θ_t follows the probability distribution function (PDF)

$$f(\theta_t) = \begin{cases} \frac{1}{4\pi^2}(2\pi - \theta_t), & \text{if } \theta_t \in [0, 2\pi), \\ \frac{1}{4\pi^2}(2\pi + \theta_t), & \text{if } \theta_t \in [-2\pi, 0), \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

Since $e^{j\theta_t} \stackrel{d}{=} e^{j\,\mathrm{mod}(\theta_t, 2\pi)}$, where $\mathrm{mod}(\theta_t, 2\pi)$ is the modulo 2π of θ_t, defining $\theta \triangleq \mathrm{mod}(\theta_t, 2\pi)$ yields

$$f(\theta) = \begin{cases} \frac{1}{2\pi}, & \text{if } \theta \in [0, 2\pi) \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

Since $\eta_{k,m,i}$ and $\phi_{k,n}$ are i.i.d., it is straightforward to conclude that $[\mathbf{X}_{k,i}]_{m,n}$ is i.i.d. $\forall m, n$. ∎

The distribution in Lemma 1 will be used in the next section to establish the RIP.

IV. RESTRICTED ISOMETRY PROPERTY

In this section, we show that the affine map $\mathcal{A}_{\mathcal{S}}$ in (16) satisfies the RIP in (9) with high probability.

A. Preliminaries

The following theorem states a sufficient condition for the RIP of an arbitrary affine map $\mathcal{A}$.

Theorem 1 [31]: Let an affine map $\mathcal{A} : \mathbb{C}^{N_r \times N_t} \mapsto \mathbb{C}^{M \times 1}$ be a random ensemble obeying the following condition: for any $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, whose rank is at most L, and fixed $0 < \alpha < 1$,

$$\mathrm{Pr}(|\|\mathcal{A}(\mathbf{H})\|_2^2 - \|\mathbf{H}\|_F^2| > \alpha\|\mathbf{H}\|_F^2) \leq Ce^{-cM}, \quad (21)$$

for $C, c > 0$ which depends on α. Then, if $M \geq CL(N_t + N_r + 1)$, $\mathcal{A}$ satisfies the RIP in (9) with probability exceeding $1 - Ce^{-qM}$ for a fixed constant $q > 0$, where $q = c - \frac{\ln(36\sqrt{2}/\delta_L)}{C}$ and δ_L is the L-RIP constant.

In (21), the probability is taken over the affine map $\mathcal{A}$. Here, $\mathcal{A}$ is not restricted to the sounder set $\mathcal{S}$ in (16) but it is modeled for any arbitrary $\mathcal{S}$. Reindexing $\mathbf{X}_{k,i}$ in (16) as $\mathbf{X}_m$ with m being a function of (k, i) such that $m = (k-1)N + i$, $k = 1, \ldots, K$ and $i = 1, \ldots, N$, we have

$$[\mathcal{A}(\mathbf{H})]_m = \mathrm{tr}(\mathbf{X}_m^H\mathbf{H}), \quad m = 1, \ldots, M. \quad (22)$$

It is tractable to transform the sufficient condition in (21) in a vector form

$$\mathrm{Pr}(|\|\mathbf{A}\mathrm{vec}(\mathbf{H})\|_2^2 - \|\mathrm{vec}(\mathbf{H})\|_2^2| > \alpha\|\mathrm{vec}(\mathbf{H})\|_2^2) \leq Ce^{-cM}, \quad (23)$$

where $\mathbf{A} \in \mathbb{C}^{M \times N_r N_t}$ and

$$[\mathbf{A}]_{m,:} = \mathrm{vec}(\mathbf{X}_m)^H, \quad \forall m.$$

The condition in (23) specifies that the random ensembles of $\|\mathbf{A}\mathrm{vec}(\mathbf{H})\|_2^2$ should be concentrated around $\|\mathrm{vec}(\mathbf{H})\|_2^2$ in order to meet the RIP.

One such example of $\mathbf{A}$ is the matrix with i.i.d. Gaussian entries [36], i.e., $[\mathbf{A}]_{m,n} \sim \mathcal{N}(0, \frac{1}{M})$. Another example is when $\mathbf{A}$ has entries sampled from an i.i.d. symmetric Bernoulli distribution [37], i.e.,

$$[\mathbf{A}]_{m,n} \in \{1/\sqrt{M}, -1/\sqrt{M}\} \quad \text{with probability } 1/2. \quad (24)$$

Specifically, in [37], it has been shown that the random construction in (24) satisfies (23) with $C = 2$ and $c = \alpha^2/4 - \alpha^3/6$.

Accoridng to Lemma 1, the matrix $\mathbf{A}$ associated with $\mathcal{S} = \{\mathbf{W}_k, \mathbf{f}_k\}_{k=1}^K$, which we will denote as $\mathbf{A}_{\mathcal{S}}$, has i.i.d. entries following

$$[\mathbf{A}_{\mathcal{S}}]_{m,n} \stackrel{d}{=} \frac{1}{\sqrt{N_r N_t}}e^{j\theta} \quad (25)$$

with $\theta \in \mathcal{U}[0, 2\pi)$. Each entry of (25) can be viewed as a random phase rotation of $[\mathbf{A}]_{m,n}$ in (24). In what follows, we will establish the RIP of $\mathbf{A}_{\mathcal{S}}$ in (25) by showing (23) (i.e., (21) in Theorem 1).

B. RIP Proof

The following lemma establishes a connection between the vector drawn from the distribution in (25) and the vector consisting of Bernoulli random variables in (24).

Lemma 2: Let $\mathbf{A}_{\mathcal{S}} \in \mathbb{C}^{M \times N_r N_t}$ and $\mathbf{B} \in \mathbb{R}^{M \times N_r N_t}$ be random matrices with i.i.d. entries $[\mathbf{A}_{\mathcal{S}}]_{m,n} = 1/\sqrt{N_r N_t}e^{j\theta_{m,n}}$, $\theta_{m,n} \in \mathcal{U}[0, 2\pi)$, and $[\mathbf{B}]_{m,n} \in \{-1/\sqrt{N_r N_t}, 1/\sqrt{N_r N_t}\}$ with equal probability, respectively. Then, for any fixed unit norm vector $\mathbf{x} \in \mathbb{C}^{N_r N_t \times 1}$, $\|\mathbf{x}\|_2^2 = 1$,

$$E\left[|[\mathbf{A}_{\mathcal{S}}]_{m,:}\mathbf{x}|^{2u}\right] \leq E\left[|[\mathbf{B}]_{m,:}\bar{\mathbf{x}}|^{2u}\right], \quad (26)$$

where u is a non-negative integer, $\bar{\mathbf{x}}$ is the element-wise absolute of $\mathbf{x}$, and the expectations are taken over $[\mathbf{A}_{\mathcal{S}}]_{m,:}$ and $[\mathbf{B}]_{m,:}$, respectively.

Proof: See Appendix A. ∎

Now, we show that $\mathbf{A}_{\mathcal{S}}$ in (25) meets the bound (21). The following theorem is inspired by a similar theorem in [37, Theorem 5.1] and refines the results for $\mathbf{A}_{\mathcal{S}}$ in (25).

Theorem 2: Let $\mathbf{A}_{\mathcal{S}} \in \mathbb{C}^{M \times N_r N_t}$ be a random matrix with i.i.d. entries $[\mathbf{A}_{\mathcal{S}}]_{m,n} = \frac{1}{\sqrt{N_r N_t}}e^{j\theta}$, $\forall m, n$, where $\theta \in \mathcal{U}[0, 2\pi)$.

Then, for any fixed $\mathbf{x} \in \mathbb{C}^{N_r N_t \times 1}$ with $\|\mathbf{x}\|_2^2 = 1$ and $0 < \alpha < 1$, the following holds

$$\Pr\left(\left|\left\|\sqrt{N_t N_r/M}\,\mathbf{A}_\mathcal{S}\mathbf{x}\right\|_2^2 - 1\right| \geq \alpha\right) \leq 2e^{-\frac{M}{2}\left(\frac{\alpha^2}{2} - \frac{\alpha^3}{3}\right)}. \quad (27)$$

Proof: See Appendix B. ∎

Based on Theorem 1 and Theorem 2, we provide our main theorem of this section as follows.

Theorem 3: Suppose the affine map $\mathcal{A}_\mathcal{S}$ in (16) and $\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H}) = \sqrt{\frac{N_r N_t}{M}}\mathcal{A}_\mathcal{S}(\mathbf{H})$. For arbitrary $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, whose rank is at most L, and for any $0 < \alpha < 1$, the following holds

$$\Pr\left(\left|\|\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})\|_2^2 - \|\mathbf{H}\|_F^2\right| \geq \alpha\|\mathbf{H}\|_F^2\right) \leq 2e^{-\frac{M}{2}\left(\frac{\alpha^2}{2} - \frac{\alpha^3}{3}\right)}. \quad (28)$$

Moreover, if $M \geq 2L(N_t + N_r + 1)$, then $\hat{\mathcal{A}}_\mathcal{S}$ satisfies the RIP in (9) with probability exceeding $1 - 2e^{-qM}$ and isometry constant δ_L, where $q = \frac{\alpha^2}{4} - \frac{\alpha^3}{6} - \frac{\ln(36\sqrt{2}/\delta_L)}{2}$.

Proof: Rewrite $\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H}) = \sqrt{\frac{N_r N_t}{M}}\mathbf{A}_\mathcal{S}\,\mathrm{vec}(\mathbf{H})$ where, by Lemma 1, each entry of $\mathbf{A}_\mathcal{S} \in \mathbb{C}^{M \times N_r N_t}$ is i.i.d. and follows the same distribution as $\frac{1}{\sqrt{N_r N_t}}\exp(j\theta)$ for $\theta \in \mathcal{U}[0, 2\pi)$. Then, by Theorem 2, substituting $\mathbf{x} = \mathrm{vec}(\mathbf{H})/\|\mathbf{H}\|_F$ into (27), the following holds

$$\Pr\left(\left|\left\|\frac{\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})}{\|\mathbf{H}\|_F}\right\|_2^2 - 1\right| \geq \alpha\right) = \Pr\left(\left|\|\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})\|_2^2 - \|\mathbf{H}\|_F^2\right| \geq \alpha\|\mathbf{H}\|_F^2\right)$$
$$\leq 2e^{-\frac{M}{2}\left(\frac{\alpha^2}{2} - \frac{\alpha^3}{3}\right)}, \quad (29)$$

i.e., the sufficient condition in Theorem 1. Then, by Theorem 1, if (29) holds and

$$M \geq 2L(N_t + N_r + 1), \quad (30)$$

$\hat{\mathcal{A}}_\mathcal{S}$ satisfies the RIP for L-RIP constant δ_L with probability exceeding $1 - 2\,e^{-qM}$ and fixed constant $q = \frac{\alpha^2}{4} - \frac{\alpha^3}{6} - \frac{\ln(36\sqrt{2}/\delta_L)}{2}$. This concludes the proof. ∎

Remark 2: Theorem 3 signifies that the subspace samples in (16) become as significant as directly observing $\mathbf{H}$ when the number of channel uses K grows as large as $2\frac{L}{N}(N_t + N_r + 1)$. For fixed rank L, as N_r, N_t, and N tend to infinity while keeping the ratio $\frac{N_r + N_t}{N} = \gamma$ fixed, the channel $\mathbf{H}$ becomes rank-sparse and $K \to 2\,L\gamma$. On the other hand, for fixed L and N, as N_r, N_t grow large, the overhead K is on the order of $O(\max(N_r, N_t))$. The RIP will further be consolidated in Section V by analyzing the estimation error performance of the low-rank subspace decomposition problem in (13).

C. Asymptotic Analysis

To gain intuition on the bound in (28), we model the distribution of $\|\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})\|_2^2$ in the asymptotically large M, N_r, and N_t regime. The central limit theorem (CLT) [38] provides a tractable means for this purpose. From (16), we have

$$[\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})]_m = \sqrt{\frac{N_r N_t}{M}}\,\mathrm{tr}(\mathbf{X}_m^H\mathbf{H}), \quad m = 1, \ldots, M.$$

As $N_r, N_t \to \infty$, by the CLT, $\frac{\sqrt{M}}{\|\mathbf{H}\|_F}[\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})]_m$ converges to the Gaussian normal distribution, i.e., $\frac{\sqrt{M}}{\|\mathbf{H}\|_F}[\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})]_m \sim \mathcal{CN}(0, 1)$, resulting in

$$[\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})]_m \sim \mathcal{CN}(0, \|\mathbf{H}\|_F^2/M), \quad \forall m. \quad (31)$$

This implies $\frac{M}{\|\mathbf{H}\|_F^2}\|\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})\|_2^2 \sim \chi^2(M)$, where $\chi^2(M)$ represents the central chi-squared distribution with M degrees of fredom. As $M \to \infty$, the central chi-squared distribution behaves as $\chi^2(M) \sim \mathcal{N}(M, 2M)$ [39]. Hence, the asymptotic distribution of $\|\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})\|_2^2$ follows

$$\|\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})\|_2^2 \sim \mathcal{N}(\|\mathbf{H}\|_F^2, 2\|\mathbf{H}\|_F^4/M). \quad (32)$$

Based on (32), the l.h.s. of (28) is upper bounded by

$$\Pr\left(\left|\left\|\frac{\hat{\mathcal{A}}_\mathcal{S}(\mathbf{H})}{\|\mathbf{H}\|_F}\right\|_2^2 - 1\right| \geq \alpha\right) = \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{-\alpha\sqrt{M/2}} e^{-\frac{t^2}{2}}dt$$
$$\leq 2e^{-\frac{M}{2}\frac{\alpha^2}{2}}, \quad (33)$$

where the last step follows from the Chernoff bound of the normal Gaussian random variable [38].

Remark 3: It is interesting to examine the two-sided tail bounds in (29) and (33). The comparison can be made by looking at the two exponents. It can be seen that $\frac{\alpha^2}{2} - \frac{\alpha^3}{3} \leq \frac{\alpha^2}{2}$ and $\frac{\alpha^2}{2} - \frac{\alpha^3}{3} \to \frac{\alpha^2}{2}$ as $\alpha \to 0$, meaning for asymptotically large M, N_r, and N_t, the bound in (29) converges to the CLT bound in (33) as α tends to zero. This is to say, in the asymptotic sense, the RIP characterization is equivalent to the CLT whose convergence is in probability. A fundamental difference is that the RIP in Theorem 3 gives a tighter characterization than the convergence in probability since it does not require that M, N_r, and N_t tend to infinity, while the bound in (33) only holds for asymptotically large M, N_r, and N_t.

V. ESTIMATION ERROR ANALYSIS AND AN ALGORITHM FOR LOW-RANK SUBSPACE DECOMPOSITION

On the basis of the RIP, the estimation error performance of the low-rank subspace decomposition problem in (13) is analyzed. We also provide an algorithm for effective minimization of the low-rank subspace decomposition problem.

A. Estimation Error Bound

Denote the compact sigular value decomposition (SVD) of the channel as $\mathbf{H} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{\Phi}^H$, where $\mathbf{\Psi} \in \mathbb{C}^{N_r \times L}$ and $\mathbf{\Phi} \in \mathbb{C}^{N_t \times L}$ are semi-unitary matrices, i.e., $\mathbf{\Psi}^H\mathbf{\Psi} = \mathbf{I}_L$ and $\mathbf{\Phi}^H\mathbf{\Phi} = \mathbf{I}_L$, and $\mathbf{\Lambda} \in \mathbb{R}^{L \times L}$ is the diagonal matrix with singular values $\lambda_1 \geq \cdots \geq \lambda_L$ at position (l, l), $l = 1, \ldots, L$. Define the desired channel estimate of the subspace decomposition problem in (13) as $\mathbf{H}_d$, where $\mathbf{H}_d = \sum_{l=1}^d \lambda_l\psi_l\phi_l^H$ with ψ_l and ϕ_l being the lth columns of $\mathbf{\Psi}$ and $\mathbf{\Phi}$ and $d \leq L$. We also define $\mathbf{H}_{L\backslash d} = \mathbf{H} - \mathbf{H}_d = \sum_{l=d+1}^L \lambda_l\psi_l\phi_l^H$ as the residual. For the squared error $\|\mathbf{H}_d - \hat{\mathbf{H}}\|_F^2$ of the low-rank subspace decomposition in (13), we have the following theorem.

Theorem 4: Consider the problem in (13) where $\mathbf{H} = \mathbf{H}_d + \mathbf{H}_{L\backslash d}$ and $\mathcal{A}_\mathcal{S}$ is generated according to (16). Given that $\sqrt{\frac{N_r N_t}{M}}\mathcal{A}_\mathcal{S}(\cdot)$ meets the RIP with $2d$-RIP constant δ_{2d} for $M \geq 4d(N_t + N_r + 1)$, we have

$$\left\|\mathbf{H}_d - \hat{\mathbf{H}}\right\|_F^2 \leq \min\left\{\frac{4N_t N_r(\|\mathcal{A}_\mathcal{S}(\mathbf{H}_{L\backslash d}) + \tilde{\mathbf{n}}\|_2^2)}{(1 - \delta_{2d})M}, 2\beta\right\}, \quad (34)$$

where the bound (34) holds with probability exceeding $1 - 2e^{-qM}$ for $q = \frac{\alpha^2}{4} - \frac{\alpha^3}{6} - \frac{\ln(36\sqrt{2}/\delta_{2d})}{2}$ and $0 < \alpha < 1$.

Proof: See Appendix C. ∎

Remark 4: In (34), as M grows, $\|\mathcal{A}_S(\mathbf{H}_{L\backslash d}) + \tilde{\mathbf{n}}\|_2^2 / M \to E\left[\left|[\mathcal{A}_S(\mathbf{H}_{L\backslash d}) + \tilde{\mathbf{n}}]_i\right|^2\right]$, $i = 1, 2, \cdots, M$, almost surely. Meanwhile, the constant δ_{2d} in (34) can decrease as M increases. This can be explained as follows. According to Theorem 4, the bound (34) is guaranteed with probability exceeding $1 - 2\ e^{-qM} \triangleq \mu(\delta_{2d}, M)$ where $q = \frac{\alpha^2}{4} - \frac{\alpha^3}{6} - \frac{\ln(36\sqrt{2}/\delta_{2d})}{2}$. For $M_1 \geq M_2 \geq 4d(N_t + N_r + 1)$, in order to have the same level of probability guarantee, i.e., $\mu(\delta_{2d,1}, M_1) = \mu(\delta_{2d,2}, M_2)$, it requires $\delta_{2d,1} \leq \delta_{2d,2}$ because the exponent q is proportional to δ_{2d}. This means that as M grows the upper bound in (34), which is proportional to δ_{2d}, decreases.

Remark 5: The fact that the bound (34) is proportional to δ_{2d} suggests that optimizing the subspace sampling signal set S by minimizing the RIP bound in (9) would result in better execution of the channel estimation because the tighter the RIP bound, the lower δ_{2d} will be. In this work, we do not aim to optimize S, but designing S to minimize the RIP constant in (9) (i.e., reducing the error bound in (34)) is an interesting future research problem.

Remark 6: When $L = d$, $\mathbf{H}_{L\backslash d} = 0$ in (34), and this results in the least error bound

$$\left\|\mathbf{H}_d - \widehat{\mathbf{H}}\right\|_F^2 \leq \min\left\{\frac{4N_tN_r\|\tilde{\mathbf{n}}\|_2^2}{(1 - \delta_{2d})M}, 2\beta\right\}. \tag{35}$$

Now, taking the expectation yields the mean squared error (MSE) bound

$$E\left[\left\|\mathbf{H}_d - \widehat{\mathbf{H}}\right\|_F^2\right] \leq \min\left\{\frac{4N_tN_r\sigma^2}{(1 - \delta_{2d})}, 2\beta\right\}, \tag{36}$$

where we use the fact that $E[\|\tilde{\mathbf{n}}\|_2^2] = M\sigma^2$. It is noteworthy that, in the low SNR regime (i.e., when σ^2 is large), the MSE bounded by 2β, preventing the estimation error from growing along with the noise power σ^2. This can sharpen the MSE performance at low SNR. Overall, Theorem 4 formulates a required number of channel uses, guaranteeing a bounded estimation error performance with high probability.

B. Alternating Optimization

It is difficult to directly find the solution to (13) due to the semi-unitary constraints, which are non-convex. We turn to a suboptimal approach that effectively solves the problem. To this end, we first consider the convex relaxation of (13) as

$$\left(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{\Sigma}}\right) = \underset{\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}}{\operatorname{argmin}} \left\|\mathbf{y} - \mathcal{A}_S(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H)\right\|_2^2$$

$$\text{subject to } \operatorname{tr}(\mathbf{U}^H\mathbf{U}) \leq d, \operatorname{tr}(\mathbf{V}^H\mathbf{V}) \leq d, \|\mathbf{\Sigma}\|_F^2 \leq \beta, \tag{37}$$

where $\mathbf{\Sigma} \in \mathbb{C}^{d \times d}$ is the same diagonal matrix as in (13). The equality constraints are relaxed and the feasible set of (37) becomes convex. However, the problem is still non-convex since the optimization parameters are coupled with each other. The coupling among $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$ in the objective function

---

Algorithm 1 Alternating Minimization for Low-Rank Subspace Decomposition

1: Input: Affine map $\mathcal{A}_S$, observations $\mathbf{y}$, and maximum number of iteration ℓ_{max}.
2: Initialization: Set $\ell = 0$, initialize $\widehat{\mathbf{H}}_{(0)} = \mathcal{A}_S^*(\mathbf{y})$, let $\mathbf{U}_{(0)}$, $\mathbf{V}_{(0)}$, and $\mathbf{\Sigma}_{(0)}$ be the top d left singular vectors, right singular vectors, and singular value matirx of $\widehat{\mathbf{H}}_{(0)}$, respectively.
3: **repeat**
4:    Update the column subspace $\mathbf{U}_{(\ell)}$ to $\mathbf{U}_{(\ell+1)}$ by solving

$$(\text{S1}) \begin{cases} \mathbf{U}_{(\ell+1)} = \underset{\mathbf{U}}{\operatorname{argmin}} \left\|\mathbf{y} - \mathcal{A}_S(\mathbf{U}\mathbf{\Sigma}_{(\ell)}\mathbf{V}_{(\ell)}^H)\right\|_2^2 \\ \qquad\qquad \text{subject to } \operatorname{tr}(\mathbf{U}^H\mathbf{U}) \leq d, \end{cases}$$

5:    Update the row subspace $\mathbf{V}_{(\ell)}$ to $\mathbf{V}_{(\ell+1)}$ by solving

$$(\text{S2}) \begin{cases} \mathbf{V}_{(\ell+1)} = \underset{\mathbf{V}}{\operatorname{argmin}} \left\|\mathbf{y} - \mathcal{A}_S(\mathbf{U}_{(\ell+1)}\mathbf{\Sigma}_{(\ell)}\mathbf{V}^H)\right\|_2^2 \\ \qquad\qquad \text{subject to } \operatorname{tr}(\mathbf{V}^H\mathbf{V}) \leq d, \end{cases}$$

6:    Update the subspace power allocation matrix $\mathbf{\Sigma}_{(\ell)}$ to $\mathbf{\Sigma}_{(\ell+1)}$ by solving

$$(\text{S3}) \begin{cases} \mathbf{\Sigma}_{(\ell+1)} = \underset{\mathbf{\Sigma}}{\operatorname{argmin}} \left\|\mathbf{y} - \mathcal{A}_S(\mathbf{U}_{(\ell+1)}\mathbf{\Sigma}\mathbf{V}_{(\ell+1)}^H)\right\|_2^2, \\ \text{subject to } \|\mathbf{\Sigma}\|_F^2 \leq \beta, \end{cases}$$

7:    $\widehat{\mathbf{H}}_{(\ell+1)} = \mathbf{U}_{(\ell+1)}\mathbf{\Sigma}_{(\ell+1)}\mathbf{V}_{(\ell+1)}^H$ and $\ell = \ell + 1$,
8: **until** ℓ exceeds ℓ_{max} or the iterations stagnate.
9: Output: $\widehat{\mathbf{H}} = \widehat{\mathbf{H}}_{(\ell)}$.

---

makes block coordinate decent [40] an ideal approach for (37) to iteratively optimize $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$. In particular, optimizing one parameter by fixing the other two parameters in (37) is convex. Denoting ℓ as the interation index, we can iteratively optimize $\mathbf{U}_\ell$, $\mathbf{V}_\ell$, and $\mathbf{\Sigma}_\ell$ for $\ell = 1, 2, \ldots$ by solving the following subproblems:

(S1) Fix the row subspace $\mathbf{V}_\ell$ and power allocation $\mathbf{\Sigma}_\ell$, optimize the column subspace $\mathbf{U}_\ell$;
(S2) Fix the column subspace $\mathbf{U}_\ell$ and power allocation $\mathbf{\Sigma}_\ell$, optimize the row subspace $\mathbf{V}_\ell$;
(S3) Fix the row subspace $\mathbf{V}_\ell$ and column subspace $\mathbf{U}_\ell$, optimize the power allocation $\mathbf{\Sigma}_\ell$.

A formal description of the alternating optimization is provided in Algorithm 1. In Step 2 of Algorithm 1, $\mathcal{A}_S^* : \mathbb{C}^{M \times 1} \mapsto \mathbb{C}^{N_r \times N_t}$ denotes the adjoint operator of the affine map $\mathcal{A}_S : \mathbb{C}^{N_r \times N_t} \mapsto \mathbb{C}^{M \times 1}$. Specifically, based on (22), it follows $\mathcal{A}_S^*(\mathbf{y}) = \sum_{m=1}^M [\mathbf{y}]_m \mathbf{X}_m$. Hence, the initialization $\widehat{\mathbf{H}}_{(0)} = \mathcal{A}_S^*(\mathbf{y}) = \sum_{m=1}^M [\mathbf{y}]_m \mathbf{X}_m$ in Step 2 can be viewed as to linearly estimate the channel $\mathbf{H}$ by taking the sounder set $S = \{(\mathbf{W}_k, \mathbf{f}_k)\}_{k=1}^K$ as its basis and the noisy subspace samples $[\mathbf{y}]_m$ as their combining weights.

Since (S1), (S2), and (S3) are all convex, we can obtain the optimal solution of each subproblem in closed-form. We limit our discussion to the column subspace optimization (S1) and the power matrix optimization (S3), keeping in mind that the same approach for solving (S1) applies to the row subspace

optimization (S2). For simplicity, we omit the iteration index ℓ attached to the variables. The following lemma provides the solution to (S1).

Lemma 3: Consider the following quadratic program,

$$\widehat{\mathbf{U}} = \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \mathbf{y} - \mathcal{A}_{\mathcal{S}}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H) \right\|_2^2$$
$$\text{subject to } \operatorname{tr}(\mathbf{U}^H\mathbf{U}) \leq d, \qquad (38)$$

where $\mathbf{y} \in \mathbb{C}^{M \times 1}$, $\mathbf{U} \in \mathbb{C}^{N_r \times d}$, $\mathbf{V} \in \mathbb{C}^{N_t \times d}$, $\boldsymbol{\Sigma} \in \mathbb{C}^{d \times d}$, and $\mathcal{S} = \{(\mathbf{W}_k, \mathbf{f}_k)\}_{k=1}^K$. Let $\mathbf{G} \in \mathbb{C}^{M \times (dN_r)}$ be

$$\mathbf{G} = [((\boldsymbol{\Sigma}\mathbf{V}^H\mathbf{f}_1)^T \otimes \mathbf{W}_1^H)^T, \ldots, ((\boldsymbol{\Sigma}\mathbf{V}^H\mathbf{f}_K)^T \otimes \mathbf{W}_K^H)^T]^T. \qquad (39)$$

Then, the optimal solution $\widehat{\mathbf{U}}$ is given by either $\operatorname{vec}(\widehat{\mathbf{U}}) = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{y}$ such that $\| \operatorname{vec}(\widehat{\mathbf{U}})\|_2^2 \leq d$ or $\operatorname{vec}(\widehat{\mathbf{U}}) = (\mathbf{G}^H\mathbf{G} + \mu\mathbf{I}_{dN_r})^{-1}\mathbf{G}^H\mathbf{y}$ such that $g(\mu) \triangleq \| \operatorname{vec}(\widehat{\mathbf{U}})\|_2^2 = d$, where $\mu > 0$ is the unique solution to the fixed point equation $g(\mu) = d$ and $g(\mu)$ is monotonically decreasing in $\mu > 0$.

Proof: Suppose the affine map $\mathcal{A}_{\mathcal{S}}$ in (4). By the Kronecker product equality $\operatorname{vec}(\mathbf{W}_k^H\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H\mathbf{f}_k) = ((\boldsymbol{\Sigma}\mathbf{V}^H\mathbf{f}_k)^T \otimes \mathbf{W}_k^H)\operatorname{vec}(\mathbf{U})$, $k = 1, \ldots, K$ and collecting them as a long column vector yields an equivalent problem to (38) as

$$\widehat{\mathbf{U}} = \underset{\mathbf{U}}{\operatorname{argmin}} \| \mathbf{y} - \mathbf{G}\operatorname{vec}(\mathbf{U})\|_2^2 \text{ subject to } \|\operatorname{vec}(\mathbf{U})\|_2^2 \leq d,$$

where $\mathbf{G}$ is given in (39). This problem is a standard differentiable convex optimization problem, which can be solved by applying Karush-Kuhn-Tucker (KKT) conditions [41]. Since it is a standard procedure, we omit the details here. ∎

We now turn our attention to the subspace power allocation problem (S3). To obtain a closed-form solution to (S3), we have the following lemma.

Lemma 4: Consider the following quadratic program,

$$\widehat{\boldsymbol{\Sigma}} = \underset{\boldsymbol{\Sigma}}{\operatorname{argmin}} \left\| \mathbf{y} - \mathcal{A}_{\mathcal{S}}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H) \right\|_2^2,$$
$$\text{subject to } \|\boldsymbol{\Sigma}\|_F^2 \leq \beta, \qquad (40)$$

where $\mathbf{y} \in \mathbb{C}^{M \times 1}$, $\mathbf{U} \in \mathbb{C}^{N_r \times d}$, $\mathbf{V} \in \mathbb{C}^{N_t \times d}$, and $\boldsymbol{\Sigma} \in \mathbb{C}^{d \times d}$ is diagonal. Define the matrix $\mathbf{P} \in \mathbb{C}^{M \times d}$ with the ith column

$$[\mathbf{P}]_{:,i} = \mathcal{A}_{\mathcal{S}}([\mathbf{U}]_{:,i}[\mathbf{V}]_{:,i}^H). \qquad (41)$$

The solution $\widehat{\boldsymbol{\Sigma}}$ is given by either $\operatorname{diag}(\widehat{\boldsymbol{\Sigma}}) = (\mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H\mathbf{y}$ such that $\| \operatorname{diag}(\widehat{\boldsymbol{\Sigma}})\|_2^2 \leq \beta$ or $\operatorname{diag}(\widehat{\boldsymbol{\Sigma}}) = (\mathbf{P}^H\mathbf{P} + \rho\mathbf{I}_d)^{-1}\mathbf{P}^H\mathbf{y}$ such that $g(\rho) \triangleq \| \operatorname{diag}(\widehat{\boldsymbol{\Sigma}})\|_2^2 = \beta$, where $\rho > 0$ is the unique solution of the equation $g(\rho) = \beta$.

Proof: Rewrite $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H = \sum_{i=1}^d [\mathbf{U}]_{:,i}[\boldsymbol{\Sigma}]_{i,i}[\mathbf{V}]_{:,i}^H$. Then,

$$\mathcal{A}_{\mathcal{S}}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H) = \sum_{i=1}^d [\boldsymbol{\Sigma}]_{i,i}\mathcal{A}_{\mathcal{S}}([\mathbf{U}]_{:,i}[\mathbf{V}]_{:,i}^H). \qquad (42)$$

Defining $[\mathbf{P}]_{:,i} = \mathcal{A}_{\mathcal{S}}([\mathbf{U}]_{:,i}[\mathbf{V}]_{:,i}^H)$ in (42), the problem in (40) is given by

$$\operatorname{diag}(\widehat{\boldsymbol{\Sigma}}) = \underset{\boldsymbol{\Sigma}}{\operatorname{argmin}} \left\| \mathbf{y} - \mathbf{P}\operatorname{diag}(\boldsymbol{\Sigma}) \right\|_2^2,$$
$$\text{subject to } \|\boldsymbol{\Sigma}\|_F^2 \leq \beta. \qquad (43)$$

The optimal solution to (43) is found by applying the KKT conditions [41]. ∎

Notice that since the subsequent minimizations in (S1), (S2), and (S3) are optimally solved, a sequence of $\{\| \mathbf{y} - \mathcal{A}_{\mathcal{S}}(\mathbf{U}_\ell\boldsymbol{\Sigma}_\ell\mathbf{V}_\ell^H)\|_2^2\}_{\ell \geq 1}$ is non-increasing and converges to a stationary point over the iterations in Algorithm 1.

VI. SIMULATION RESULTS

A. Simulation Setup

In this section, we evaluate the performance of low-rank subspace decomposition (SD) in Algorithm 1. We adopt the physical representation of sparse millimeter wave MIMO channels [4], [8], where L scatters are assumed to constitute the propagation paths between the transmitter and receiver,

$$\mathbf{H} = \sqrt{\frac{N_r N_t}{L}} \sum_{l=1}^L h_l \mathbf{a}_r(\xi_{r,l})\mathbf{a}_t^H(\xi_{t,l}). \qquad (44)$$

In (44), $h_l \in \mathbb{C}$ is the complex gain of the lth path, i.i.d. according to $h_l \sim \mathcal{CN}(0, 1)$, and $\xi_{t,l}$ and $\xi_{r,l}$ are the angles of departures (AoDs) and arrivals (AoAs) at the transmitter and receiver, respectively. Both $\xi_{t,l}$ and $\xi_{r,l}$ are i.i.d. and uniform over $[-\pi/2, \pi/2]$. The $\mathbf{a}_t(\xi_{t,l}) \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{a}_r(\xi_{r,l}) \in \mathbb{C}^{N_r \times 1}$ are the array response vectors at the transmitter and receiver, respectively. We assume, for simplicity, $\mathbf{a}_t(\xi_{t,l})$ and $\mathbf{a}_r(\xi_{r,l})$ are uniform linear arrays (ULAs) with the antenna spacing set to half of the wavelength. The channel model in (44) satisfies $E[\|\mathbf{H}\|_F^2] = N_r N_t$. Throughout the simulation, we assume $N_r = 36$ and $N_t = 144$, and the parameter β in (13) for the SD method is set to $\beta = E[\|\mathbf{H}\|_F^2] = N_r N_t$.

We consider four other channel estimation techniques as our benchmarks, two open-loop channel estimation techniques, i.e., the MF in [34] and CS-based method in [42], and two adaptive channel estimation techniques, i.e., the two-way Arnoldi approach in [9] and closed-loop adaptive channel estimation (ACE) in [8]. The work in [42] provides an open-loop compressed sensing (CS) framework. Unlike the proposed SD, the CS-based method in [42] assumes the sparse representation of millimeter wave channels in angular domain. Hence, in [42], the orthogonal matching pursuit (OMP) is employed to recover the angle supports of the channel matrix. To enhance the sensing capability, it designs the measurement matrix to meet the incoherency property [43].

For the comparison, we admit two performance metrics, i.e., normalized MSE (NMSE) and spectrum efficiency. Specifically, the NMSE is defined as

$$\text{NMSE} = E\left[\left\|\mathbf{H}_d - \widehat{\mathbf{H}}\right\|_F^2 / \|\mathbf{H}_d\|_F^2\right]. \qquad (45)$$

The spectrum efficiency is defined as

$$R = \log_2\left|\mathbf{I}_d + \frac{1}{\sigma^2}\mathbf{R}_n^{-1}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right|, \qquad (46)$$

where $\tilde{\mathbf{H}} = \mathbf{W}_D^H\mathbf{W}_A^H\mathbf{H}\mathbf{F}_A\mathbf{F}_D$ and $\mathbf{R}_n = \mathbf{W}_D^H\mathbf{W}_A^H\mathbf{W}_A\mathbf{W}_D$ with $\mathbf{F}_A \in \mathbb{C}^{N_t \times N}$, $\mathbf{F}_D \in \mathbb{C}^{N \times d}$, $\mathbf{W}_A \in \mathbb{C}^{N_r \times N}$, and $\mathbf{W}_D \in \mathbb{C}^{N \times d}$, respectively, being the analog precoder, digital precoder, analog combiner, and digital combiner. Here, we consider d-stream multiplexing. For a fair comparison, we admit the near-optimal precoding technique proposed in [7]. This is to say, given the channel estimate $\widehat{\mathbf{H}}$ attained by the proposed
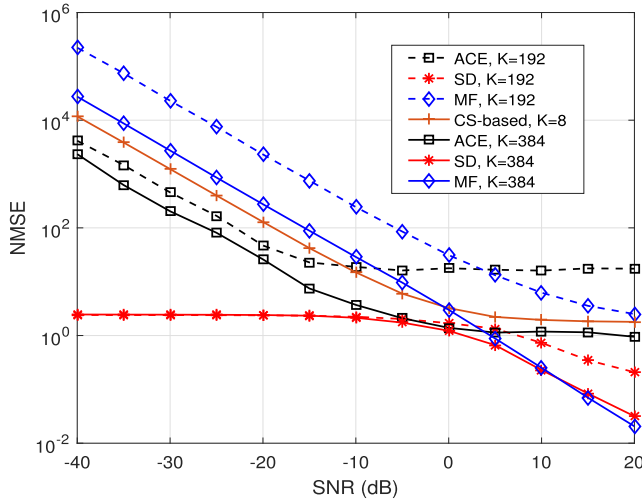
Fig. 2.   NMSE vs. SNR (dB) ($N_t = 144, N_r = 36, L = 4, d = 4, N = 4$).

Fig. 3.   Spectrum efficiency vs. SNR (dB) ($N_t = 144, N_r = 36, L = 4, d = 4, N = 4$).

SD, MF [34], CS-based method [42], ACE [8], and Arnoldi [9] approaches, the precoders and combiners are designed based on the technique in [7] and then R in (46) for each approach is evaluated. For the NMSE evaluation, we involve the proposed SD, MF, CS-based method, and ACE,[1] while we consider all the benchmarks for the spectrum efficiency.

B. Numerical Results

In Fig. 2, the NMSE of the proposed SD is compared with the MF, CS-based method, and ACE. We assume the number of paths $L = 4$, the rank of the desired channel estimate $d = 4$, and the number of RF chains $N = 4$. According to Theorem 3, channel use overhead K satisfies $K \geq 2\frac{L}{N}(N_r + N_t + 1)$ to meet the RIP. We set the channel uses for SD, MF and ACE at $K = 192$ and $K = 384$, which are approximately half and equal to the bound defined in (30), respectively. Because of the assumed angle grid channel model in [42], the number of channel uses for the CS-based method is reduced to $K = 2N$. Seen from Fig. 2, the proposed SD results in the most accurate channel estimation performance. Moreover, the NMSE of the proposed SD shows sharp improvement at low SNR and it further decreases as the SNR grows, which is consistent with the analysis in (36). Specifically, when the SNR is high (i.e., 0 dB ~ 20 dB) and the number of channel uses is sufficiently large (i.e., $K = 384$), MF and SD can achieve accurate estimation. The latter scenario represents an ideal case of high SNR with a sufficiently large number of observations. On the other hand, when $K = 192$, the MF experiences large performance degradation, while the proposed SD is still compatible with the performance of $K = 384$, demonstrating its robustness. For ACE and CS-based method, even if the SNR and/or the number of channel uses are high, they can not provide an accurate channel estimate. This happens because the channel estimation of these methods is based on the angle grids [8] [42]. When the true AoDs and AoAs do not fall exactly on the grid points, these two methods are limited by residual NMSE even at very high SNR, thus the NMSE

flooring in Fig. 2. When the SNR is less than 0 dB in Fig. 2, the performance of SD does not improve as the number of channel uses K increases. This phenomenon is consistent with the mean square error bound analysis in (36). When the SNR is low, the NMSE in (36) is noise-limited and dominated by 2β, which is irrespective of K. As a result, the proposed SD significantly outperforms MF, especially at low SNR.

In Fig. 3, the spectrum efficiency of the proposed SD, MF, ACE, CS-based method, and Arnoldi approaches are illustrated under the same simulation settings as in Fig. 2. The Perfect curve in Fig. 3 is based on perfect CSI and fully digital precoding to evaluate the spectrum efficiency. Given the system dimension, the minimum channel use overhead required for Arnoldi [9] is given by $K = 360$. When the K for ACE increases to $K = 384$, its performance becomes compatible with Arnoldi. Seen from Fig. 3, the proposed SD achieves the best spectrum efficiency and both MF and SD closely approach perfect CSI performance as SNR grows. The proposed technique outperforms Arnoldi by only consuming about 50% of the channel use overhead of Arnoldi. Seen from Fig. 2 and 3, the performance gap of the proposed SD between $K = 192$ and $K = 384$ is slight and half of the bound in (30), i.e., $K = 192$ can still offer near optimal performance. Note that in Fig. 3, though the CS-based method requires only $K = 8$ channel uses, the spectrum efficiency achieved by it is very low. This is because the CS-based method, assuming the discrete AoDs and AoAs as their channel priors and using the OMP, can not be extended to the practical channel model assumed in (44). Thus, in what follows, we shall only evaluate ACE, Arnoldi, MF, and the proposed SD, and exclude the CS-based method in the rest of the simulations. In Fig. 2, when the SNR is in between -10 dB and 5 dB for $K = 384$, the NMSE performance gap between ACE and SD is minor. However, this is not the case when comparing the spectrum efficiency in Fig. 3. In Fig. 3, the proposed SD achieves considerably improved spectrum efficiency than that of the ACE. This is the direct consequence of the improved subspace estimation of SD.

---

[1]The Arnoldi approach in [9], is incapable of computing the NMSE in (45), since it only extracts the estimate of $\mathbf{H}^H\mathbf{H}$ or $\mathbf{H}\mathbf{H}^H$.
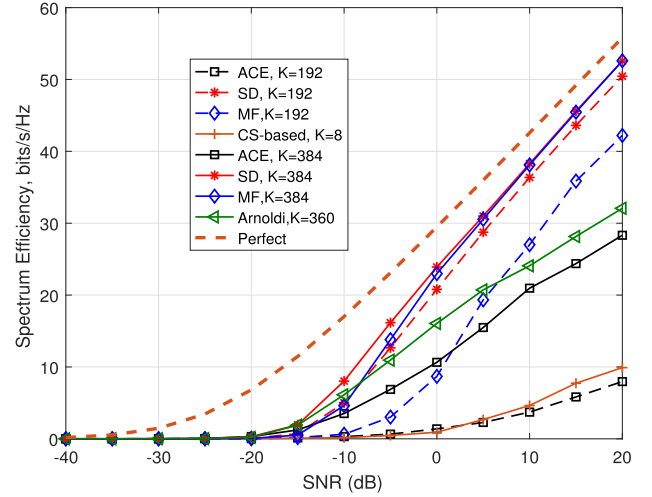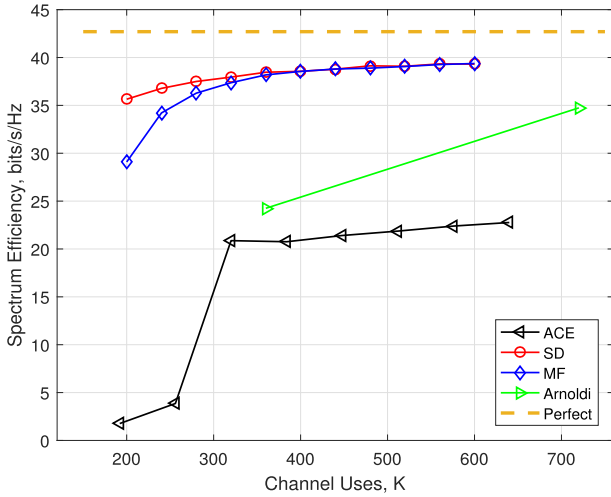
Fig. 4. Spectrum efficiency vs. Channel uses ($N_t = 144, N_r = 36, d = 4, L = 4, N = 4, \text{SNR} = 10$ dB).

Fig. 6. NMSE vs. SNR (dB) ($N_t = 144, N_r = 36, L = 4, d = 4, K = 192, N = 4, 6$).

Fig. 5. NMSE vs. SNR (dB) ($N_t = 144, N_r = 36, d = 4, K = 192, N = 4, L = 4, 5, 6$).

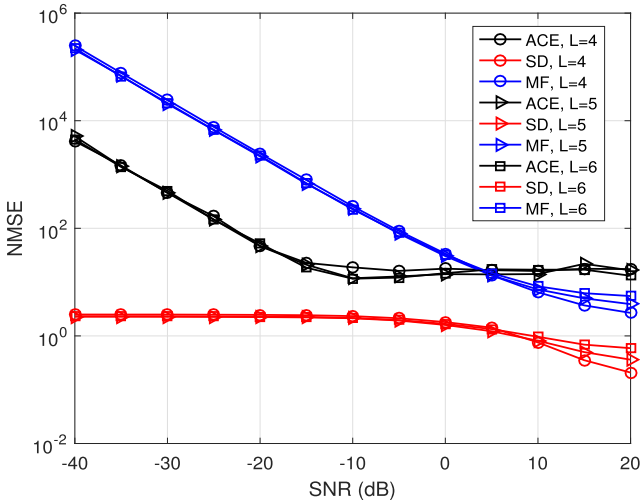Fig. 7. Spectrum efficiency vs. SNR (dB) ($N_t = 144, N_r = 36, L = 4, d = 4, N = 4, 6$).

In Fig. 4, we evaluate the impact of the number of channel uses K on the spectrum efficiency under the same simulation setting as before. The curves are displayed at SNR 10 dB across various K values. There is a clear performance improvement for ACE when $K = 384$ and above. However, ACE requires considerably larger K to further enhance its performance. In the range of K in Fig. 4, the Arnoldi approach [9] can only set $K = 360$ and $K = 720$. It should be noted that the Arnoldi approach becomes ideal if the system can be equipped with a larger number of RF chains (e.g., $N = N_t/8$ in [9]). The MF and SD benefit from increasing K, which is consistent with our analysis in Theorem 4 and (36). The proposed SD achieves the highest spectrum efficiency. It is worth noting that when the number of channel uses is over 440, the performance of MF becomes similar to SD.

In Fig. 5, we evaluate the NMSE performance for different values of channel path L. The simulation parameters are the same as before except that we now consider $L \in \{4, 5, 6\}$ for $K = 192$. As can be seen from Fig. 5, as L grows from 4 to 6,
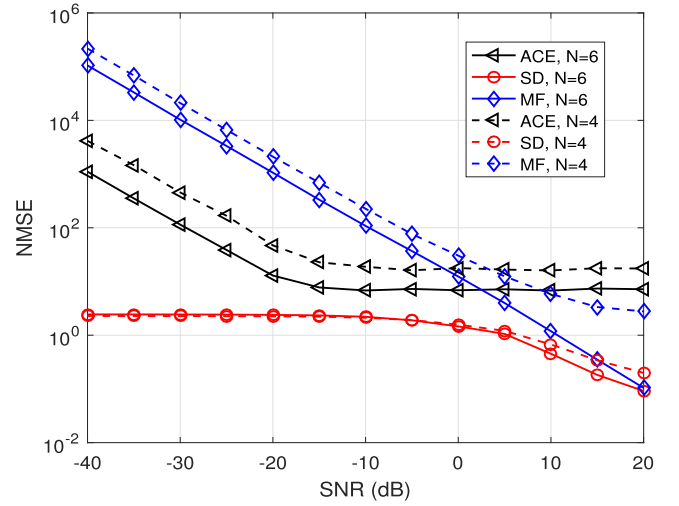
the NMSE performance of SD is deteriorated. This is due to the increased residual error $\left\| \mathbf{H}_{L \setminus d} \right\|_F^2$ in (34) as L increases according to Theorem 4. The same trend can be observed for MF. Nevertheless, the performance gap for different L values is negligible at low SNR for SD. It is observed that the number of paths does not have a considerable effect on the performance of ACE.

In this set of simulations, the effects of the number of RF chains N to the NMSE and spectrum efficiency are evaluated. In Fig. 6, the NMSE curves for $N \in \{4, 6\}$ are displayed. As illustrated in Fig. 6, for fixed $K = 192$, the more RF chains the better NMSE performance is expected. It can also be seen that the proposed SD is less sensitive to the change of N at low SNR, while achieving the lowest NMSE. The spectrum efficiency with the different numbers of RF chains is demonstrated in Fig. 7. In Fig. 7, we set K for MF and SD at $K = 192$ and ACE at $K = 192$ and $K = 384$. For Arnoldi, when $N = 4$, the minimum required channel use is $K = 360$, while when $N = 6$, this is be decreased to $K = 240$. The number

of channel uses for MF and SD are the least among all the benchmarks. Nevertheless, the proposed SD still shows the best performance.

C. Complexity Analysis

In this subsection, we compare the computational complexities of the proposed SD, ACE [8], MF [33], and Arnoldi [9] by counting the number of complex scalar multiplications. Provided the equal channel uses K, the complexity introduced by the precoder and combiner design is common for all the approaches. Therefore, we only need to compare the complexities involved in the alternating minimizations for SD and MF, the two-way echoing vector adaptation for Arnoldi [9], and the training precoder and combiner adaptation for ACE [8]. Specifically, for the proposed SD, the complexity is determined by the maximum number of iterations ℓ_{max} for (S1), (S2), and (S3) and the associated computations in Algorithm 1. Since the subproblems (S1), (S2), and (S3) are all least squares over a sphere, the SVD constitutes the dominant computational overhead. Hence, for a fixed ℓ_{max}, the complexity of SD is given by $\mathcal{O}\big(KNd^2(N_t^2 + N_r^2)\big)$. The same complexity $\mathcal{O}\big(KNd^2(N_t^2 + N_r^2)\big)$ results for MF. The number of AoD and AoA search levels of ACE in [8] is given by $\mathcal{O}(KN/d^3)$. The computational overhead at each search level of ACE is dominated by the complexity of designing the training precoders and combiners, which is given by $\mathcal{O}\big(ND(N_r + N_t)\big)$ [8], where $D\ (\gg \max(N_t, N_r))$ denotes the cardinality of an overcomplete dictionary. Therefore, the complexity of the ACE is given by $\mathcal{O}\big(KN^2D(N_r + N_t)/d^3\big)$. The main computational complexity of Arnoldi is the design of channel echoing vectors, which is given by $\mathcal{O}\big(K^2N^2/(N_r + N_t)\big)$ [9]. In summary, the Arnoldi method has the lowest complexity $\mathcal{O}\big(K^2N^2/(N_r + N_t)\big)$. The proposed SD has the similar complexity with MF, which is about $\mathcal{O}\big(KNd^2(N_t^2 + N_r^2)\big)$, while the ACE reveals relatively lower complexity than the SD and MF.

VII. CONCLUSIONS

In this paper, we have proposed an open-loop low-rank subspace decomposition technique for millimeter wave MIMO channel estimation. The RIP of the channel subspace sampling signals that are randomly generated by the hybrid analog and digital arrays was established. When the number of channel uses is around $\frac{L}{N}(N_t + N_r)$, the considered random subspace sampling signals satisfy the RIP with high probability. The established RIP was applied to analyze the channel estimation error performance. This analysis showed the reliability and resilience of the proposed low-rank subspace decomposition technique, especially, at low SNR. Moreover, we devised an alternating optimization algorithm that effectively finds a suboptimal, but stationary, solution to the problem. The simulation studies corroborate our analysis and showed that the proposed low-rank subspace decomposition technique can achieve near optimal performance at high SNR, while offering robust performance at low SNR with channel use overhead less than $\frac{L}{N}(N_r + N_t)$. It outperforms other adaptive closed-loop and two-way channel estimation techniques with considerably reduced channel use overhead. This demonstrates that a non-adaptive open-loop channel estimation scheme, if it is carefully designed, can still be used to provide accurate CSI for large-scale millimeter wave MIMO channels. The proposed techniques does not assume any explicit channel model and statistics, making it extendable to a wide range of sparse scenarios.

APPENDIX A
PROOF OF LEMMA 2

Any phase rotation applied to $[\mathbf{A}_\mathcal{S}]_{m,n}$ does not change its distribution, i.e., for any $\varphi \in [0, 2\pi)$,

$$[\mathbf{A}_\mathcal{S}]_{m,n} \overset{d}{=} [\mathbf{A}_\mathcal{S}]_{m,n} e^{j\varphi}.$$

Hence the following holds

$$E\left[|[\mathbf{A}_\mathcal{S}]_{m,:}\mathbf{x}|^{2u}\right] = E\left[|[\mathbf{A}_\mathcal{S}]_{m,:}\bar{\mathbf{x}}|^{2u}\right].$$

Both sides of (26) can be expanded as

$$E\left[|[\mathbf{A}_\mathcal{S}]_{m,:}\bar{\mathbf{x}}|^{2u}\right]$$
$$= \sum_{n_1=1}^{N_rN_t}\sum_{n_2=1}^{N_rN_t}\cdots\sum_{n_{2u-1}=1}^{N_rN_t}\sum_{n_{2u}=1}^{N_rN_t} E\big[[\mathbf{A}_\mathcal{S}]_{m,n_1}[\bar{\mathbf{x}}]_{n_1}$$
$$\times [\mathbf{A}_\mathcal{S}]_{m,n_2}^H[\bar{\mathbf{x}}]_{n_2}\cdots[\mathbf{A}_\mathcal{S}]_{m,n_{2u-1}}[\bar{\mathbf{x}}]_{n_{2u-1}}[\mathbf{A}_\mathcal{S}]_{m,n_{2u}}^H[\bar{\mathbf{x}}]_{n_{2u}}\big]$$

and

$$E\left[|[\mathbf{B}]_{m,:}\bar{\mathbf{x}}|^{2u}\right]$$
$$= \sum_{n_1=1}^{N_rN_t}\cdots\sum_{n_{2u}=1}^{N_rN_t} E[[\mathbf{B}]_{m,n_1}[\bar{\mathbf{x}}]_{n_1}\cdots[\mathbf{B}]_{m,n_{2u}}[\bar{\mathbf{x}}]_{n_{2u}}],$$

respectively. Since $[\bar{\mathbf{x}}]_n \geq 0$, $n = 1, \ldots, N_rN_t$, in order to verify (26), it is sufficient to show

$$E\left[[\mathbf{A}_\mathcal{S}]_{m,n_1}\ldots[\mathbf{A}_\mathcal{S}]_{m,n_{2u}}^H\right] \leq E\left[[\mathbf{B}]_{m,n_1}\ldots[\mathbf{B}]_{m,n_{2u}}\right] \tag{47}$$

for all $2u$-tuples $\{(n_1, \ldots, n_{2u})\}$.

We rewrite the inequality in (47) as

$$E\left[\prod_{n=1}^{N_tN_r}([\mathbf{A}_\mathcal{S}]_{m,n})^{p_n}([\mathbf{A}_\mathcal{S}]_{m,n}^H)^{q_n}\right] \leq E\left[\prod_{n=1}^{N_tN_r}([\mathbf{B}]_{m,n})^{p_n+q_n}\right],$$

where p_n and q_n are non-negative integers such that $\sum_{n=1}^{N_rN_t}p_n = \sum_{n=1}^{N_rN_t}q_n = u$. Since the entries of $\mathbf{A}_\mathcal{S}$ and $\mathbf{B}$ are i.i.d., the latter is equivalent to

$$\prod_{n=1}^{N_tN_r} E[([\mathbf{A}_\mathcal{S}]_{m,n})^{p_n}([\mathbf{A}_\mathcal{S}]_{m,n}^H)^{q_n}] \leq \prod_{n=1}^{N_tN_r} E[([\mathbf{B}]_{m,n})^{p_n+q_n}]. \tag{48}$$

Note that both sides of (48) can be either $\frac{1}{(N_rN_t)^u}$ or 0 depending on the p_n and q_n values. When $\prod_{n=1}^{N_tN_r} E[([\mathbf{A}_\mathcal{S}]_{m,n})^{p_n}([\mathbf{A}_\mathcal{S}]_{m,n}^H)^{q_n}] = \frac{1}{(N_rN_t)^u}$, we must have $p_n = q_n$, $\forall n$. This is to say that the left hand side (l.h.s.) of (48) is $E\big[\prod_{n=1}^{N_tN_r}|[\mathbf{A}_\mathcal{S}]_{m,n}|^{2p_n}\big] = \frac{1}{(N_rN_t)^u}$, if and only if $p_n = q_n$, $\forall n$. When $p_n = q_n$, it is easy to see that the right hand side (r.h.s.) of (48) is also $\mathbb{E}[\prod_{n=1}^{N_rN_t}|[\mathbf{B}]_{m,n}|^{2p_n}] = \frac{1}{(N_rN_t)^u}$.

On the other hand, when $p_n \neq q_n$, the l.h.s. of (48) is $\prod_{n=1}^{N_t N_r} E[([\mathbf{A}_\mathcal{S}]_{m,n})^{p_n} ([\mathbf{A}_\mathcal{S}]_{m,n}^H)^{q_n}] = 0$ and the inequality in (48) always holds regardless of the value on the r.h.s. This concludes the proof.

APPENDIX B
PROOF OF THEOREM 2

Since a part of the proof is based on a similar technique in [37], we first introduce the following lemma, which will be used in the proof.

Lemma 5 [37]: Suppose $\mathbf{b} \in \mathbb{R}^{N_r N_t \times 1}$ is a random vector where $[\mathbf{b}]_i \in \{-1/\sqrt{N_r N_t}, 1/\sqrt{N_r N_t}\}$ with equal probability. Then for any $h \in [0, N_r N_t/2]$, and any unit norm real vector $\mathbf{a} \in \mathbb{R}^{N_r N_t \times 1}$, $\|\mathbf{a}\|_2^2 = 1$,

$$E\left[\exp\left(h|\mathbf{b}^T \mathbf{a}|^2\right)\right] \leq \frac{1}{\sqrt{1-2h/(N_r N_t)}} \quad (49)$$

and

$$E\left[|\mathbf{b}^T \mathbf{a}|^4\right] \leq \frac{3}{(N_r N_t)^2}. \quad (50)$$

We first focus on the upper tail probability in (27), i.e., $\Pr(\|\sqrt{N_t N_r/M} \mathbf{A}_\mathcal{S} \mathbf{x}\|_2^2 > 1 + \alpha)$. For any $h > 0$, applying Markov's inequality to the upper tail probability yields

$$\Pr\left(\|\sqrt{N_t N_r/M} \mathbf{A}_\mathcal{S} \mathbf{x}\|_2^2 > 1 + \alpha\right) \leq E\left[e^{h\|\mathbf{A}_\mathcal{S} \mathbf{x}\|_2^2}\right] e^{-h\frac{M(1+\alpha)}{N_r N_t}}$$
$$= \left(E\left[e^{h|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2}\right]\right)^M e^{-h\frac{M(1+\alpha)}{N_r N_t}}, \quad (51)$$

where the last step follows from the fact that the entries of $\mathbf{A}_\mathcal{S}$ are i.i.d. The term $E[e^{h|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2}]$ in (51) can be further upper bounded by

$$E\left[e^{h|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2}\right] \leq E\left[e^{h|[\mathbf{B}]_{1,:} \bar{\mathbf{x}}|^2}\right]$$
$$\leq \frac{1}{\sqrt{1-2h/(N_r N_t)}}, \quad (52)$$

where the first inequality is due to the Taylor series expansion of $E\left[e^{h|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2}\right] = \sum_{k=0}^{\infty} \frac{h^k}{k!} E[|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^{2k}]$ and the inequality $E[|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^{2k}] \leq E[|[\mathbf{B}]_{1,:} \bar{\mathbf{x}}|^{2k}]$ in (26) of Lemma 2, with $\mathbf{B}$ following the definition in Lemma 2. The last step in (52) is due to (49). Inserting (52) into (51) yields

$$\Pr\left(\|\sqrt{N_t N_r/M} \mathbf{A}_\mathcal{S} \mathbf{x}\|_2^2 > 1 + \alpha\right)$$
$$\leq (1-2h/(N_r N_t))^{-\frac{M}{2}} e^{-h\frac{M(1+\alpha)}{N_r N_t}}. \quad (53)$$

Since the inequality holds for any $h > 0$, insearting $h = \frac{N_r N_t \alpha}{2(1+\alpha)}$ in (53) gives

$$\Pr\left(\|\sqrt{N_t N_r/M} \mathbf{A}_\mathcal{S} \mathbf{x}\|_2^2 > 1 + \alpha\right) \leq (1+\alpha)^{\frac{M}{2}} e^{-\alpha \frac{M}{2}}$$
$$\leq e^{-\frac{M}{2}\left(\frac{\alpha^2}{2} - \frac{\alpha^3}{3}\right)},$$

where the last inequality follows from $(1 + \alpha)^{\frac{M}{2}} e^{-\alpha \frac{M}{2}} = e^{\frac{M}{2}(\ln(1+\alpha) - \alpha)}$ and $\ln(1+\alpha) \leq \alpha - \alpha^2/2 + \alpha^3/3, \forall \alpha \in (0, 1)$. This proves the upper tail bound in (27).

Using the same approaches in (51), the lower tail bound of (27) is upper bounded by

$$\Pr\left(\|\sqrt{N_t N_r/M} \mathbf{A}_\mathcal{S} \mathbf{x}\|_2^2 < 1 - \alpha\right)$$
$$\leq \left(E\left[e^{-h|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2}\right]\right)^M e^{h\frac{M(1-\alpha)}{N_r N_t}}. \quad (54)$$

Applying the Taylor series expansion of $E\left[e^{-h|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2}\right]$ and taking the first three dominant terms yields

$$E\left[e^{-h|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2}\right] \leq E\left[1 - h|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2 + \frac{h^2 |[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^4}{2}\right].$$

Now, using the fact that $E[|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2] = \frac{1}{N_r N_t}$ and applying a sequence of bounds $E[|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^4] \leq E[|[\mathbf{B}]_{1,:} \bar{\mathbf{x}}|^4] \leq \frac{3}{(N_r N_t)^2}$, where the first bound is due to Lemma 2 and the second one is due to (50), yields

$$E\left[e^{-h|[\mathbf{A}_\mathcal{S}]_{1,:} \mathbf{x}|^2}\right] \leq 1 - \frac{h}{N_r N_t} + \frac{3h^2}{2(N_r N_t)^2}. \quad (55)$$

Then, plugging in (55) into (54) with the substitution $h = \frac{N_r N_t \alpha}{2(1+\alpha)}$ results in

$$\Pr\left(\|\sqrt{N_t N_r/M} \mathbf{A}_\mathcal{S} \mathbf{x}\|_2^2 \leq 1 - \alpha\right)$$
$$\leq \left(1 - \frac{\alpha(4+\alpha)}{8(1+\alpha)^2}\right)^M e^{\frac{\alpha(1-\alpha)M}{2(1+\alpha)}}$$
$$\leq e^{-\frac{M}{2}\left(\frac{\alpha^2}{2} - \frac{\alpha^3}{3}\right)}, \quad (56)$$

where the last inequality follows from $\left(1 - \frac{\alpha(4+\alpha)}{8(1+\alpha)^2}\right)^M e^{\frac{\alpha(1-\alpha)M}{2(1+\alpha)}} = e^{M\left(\ln\left(1 - \frac{\alpha(4+\alpha)}{8(1+\alpha)^2}\right) + \frac{\alpha(1-\alpha)}{2(1+\alpha)}\right)}$ and the inequality $\ln\left(1 - \frac{\alpha(4+\alpha)}{8(1+\alpha)^2}\right) \leq -\frac{\alpha(1-\alpha)}{2(1+\alpha)} - \frac{\alpha^2}{4} + \frac{\alpha^3}{6}$. This concludes the proof.

APPENDIX C
PROOF OF THEOREM 4

We start by realizing that the desired channel $\mathbf{H}_d = \sum_{l=1}^{d} \lambda_l \psi_l \phi_l^H$ is feasible to (13). Thus, one should have

$$\left\|\mathbf{y} - \mathcal{A}_\mathcal{S}(\widehat{\mathbf{H}})\right\|_2 \leq \left\|\mathbf{y} - \mathcal{A}_\mathcal{S}(\mathbf{H}_d)\right\|_2. \quad (57)$$

Inserting $\mathbf{y} = \mathcal{A}_\mathcal{S}(\mathbf{H}_d) + \mathcal{A}_\mathcal{S}(\mathbf{H}_{L \backslash d}) + \tilde{\mathbf{n}}$ in (57) yields

$$\left\|\mathcal{A}_\mathcal{S}(\mathbf{H}_d - \widehat{\mathbf{H}}) + \mathcal{A}_\mathcal{S}(\mathbf{H}_{L \backslash d}) + \tilde{\mathbf{n}}\right\|_2 \leq \left\|\mathcal{A}_\mathcal{S}(\mathbf{H}_{L \backslash d}) + \tilde{\mathbf{n}}\right\|_2. \quad (58)$$

Applying the triangular inequality to the l.h.s. of (58) gives

$$\left\|\mathcal{A}_\mathcal{S}(\mathbf{H}_d - \widehat{\mathbf{H}}) + \mathcal{A}_\mathcal{S}(\mathbf{H}_{L \backslash d}) + \tilde{\mathbf{n}}\right\|_2$$
$$\geq \left\|\mathcal{A}_\mathcal{S}(\widehat{\mathbf{H}} - \mathbf{H}_d)\right\|_2 - \left\|\mathcal{A}_\mathcal{S}(\mathbf{H}_{L \backslash d}) + \tilde{\mathbf{n}}\right\|_2. \quad (59)$$

Thus, combining (58) and (59) results in

$$\left\|\mathcal{A}_\mathcal{S}(\widehat{\mathbf{H}} - \mathbf{H}_d)\right\|_2 \leq 2 \left\|\mathcal{A}_\mathcal{S}(\mathbf{H}_{L \backslash d}) + \tilde{\mathbf{n}}\right\|_2. \quad (60)$$

Note that $\text{rank}(\widehat{\mathbf{H}} - \mathbf{H}_d) \leq 2d$. Then, based on the RIP of $\sqrt{\frac{N_r N_t}{M}} \mathcal{A}_\mathcal{S}(\cdot)$ by Theorem 3, we have

$$(1 - \delta_{2d}) \frac{M}{N_t N_r} \left\|\mathbf{H}_d - \widehat{\mathbf{H}}\right\|_F^2 \leq \left\|\mathcal{A}_\mathcal{S}(\widehat{\mathbf{H}} - \mathbf{H}_d)\right\|_2^2, \quad (61)$$

where δ_{2d} is $2d$-RIP constant. Hence, combining (60) and (61) results in the error bound

$$\left\|\mathbf{H}_d - \widehat{\mathbf{H}}\right\|_F^2 \leq \frac{4N_t N_r}{(1 - \delta_{2d})M} \left\|\mathcal{A}_\mathcal{S}(\mathbf{H}_{L \backslash d}) + \tilde{\mathbf{n}}\right\|_2^2. \quad (62)$$

The inequality constraint in (13) is equivalent to $\|\widehat{\boldsymbol{\Sigma}}\|_F^2 \leq \beta$. Moreover, since $\mathbf{H}_d$ is feasible to (13), one should have by the triangular inequality

$$\left\|\mathbf{H}_d - \widehat{\mathbf{H}}\right\|_F^2 \leq 2\beta. \tag{63}$$

Finally, combining (62) and (63) leads to

$$\left\|\mathbf{H}_d - \widehat{\mathbf{H}}\right\|_F^2 \leq \min\left\{\frac{4N_t N_r \|\mathcal{A}_\mathcal{S}(\mathbf{H}_{L\setminus d}) + \tilde{\mathbf{n}}\|_2^2}{(1-\delta_{2d})M}, 2\beta\right\}.$$

This bound holds, by Theorem 3, with probability greater than or equal to $1 - \exp(-qM)$ when $M \geq 4d(N_t + N_r + 1)$ for $q = \frac{\alpha^2}{4} - \frac{\alpha^3}{6} - \frac{\ln(36\sqrt{2}/\delta_{2d})}{2}$ and $0 < \alpha < 1$.

References

[1] W. Zhang, T. Kim, and D. J. Love, "Sparse subspace decomposition for millimeter wave MIMO channel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[2] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[3] E. Torkildson, U. Madhow, and M. Rodwell, "Indoor millimeter wave MIMO: Feasibility and performance," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4150–4160, Dec. 2011.

[4] R. W. Heath, Jr., 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.

[5] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[6] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

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

[8] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, Jr., "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.

[9] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.

[10] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[11] J. Song, J. Choi, S. G. Larew, D. J. Love, T. A. Thomas, and A. A. Ghosh, "Adaptive millimeter wave beam alignment for dual-polarized MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6283–6296, Nov. 2015.

[12] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution code-book and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[13] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun./Jul. 2015, pp. 146–150.

[14] A. J. Duly, T. Kim, D. J. Love, and J. V. Krogmeier, "Closed-loop beam alignment for massive MIMO channel estimation," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1439–1442, Aug. 2014.

[15] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[16] S. Noh, M. D. Zoltowski, and D. J. Love, "Training sequence design for feedback assisted hybrid beamforming in massive MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 187–200, Jan. 2016.

[17] T. Dahl, N. Christophersen, and D. Gesbert, "Blind MIMO eigenmode transmission based on the algebraic power method," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2424–2431, Sep. 2004.

[18] Y. Tang, B. Vucetic, and Y. Li, "An iterative singular vectors estimation scheme for beamforming transmission and detection in MIMO systems," *IEEE Commun. Lett.*, vol. 9, no. 6, pp. 505–507, Jun. 2005.

[19] L. P. Withers, R. M. Taylor, and D. M. Warme, "Echo-MIMO: A two-way channel training method for matched cooperative beamforming," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4419–4432, Sep. 2008.

[20] D. Ogbe, D. J. Love, and V. Raghavan, "Noisy beam alignment techniques for reciprocal MIMO channels," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5092–5107, Oct. 2017.

[21] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[22] D. S. Baum and H. Bolcskei, "Information-theoretic analysis of MIMO channel sounding," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7555–7577, Nov. 2011.

[23] "5G channel model for bands up to 100 GHz," 5GCM, White Paper, Mar. 2016, vol. 2. [Online]. Available: http://www.5gworkshops.com/5G_Channel_Model_for_bands_up_to100_GHz(2015-12-6).pdf

[24] S. Hur *et al.*, "Proposal on millimeter-wave channel modeling for 5G cellular system," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 454–469, Apr. 2016.

[25] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 608–622, Jun. 2016.

[26] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[27] Y. Chen, Y. Chi, and A. J. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 4034–4059, Jul. 2015.

[28] P. Jain, R. Meka, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2010, pp. 937–945.

[29] T. T. Cai and A. Zhang, "Compressed sensing and affine rank minimization under restricted isometry," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3279–3290, Jul. 2013.

[30] T. T. Cai and A. Zhang, "Rop: Matrix recovery via rank-one projections," *Ann. Stat.*, vol. 43, no. 1, pp. 102–138, 2015.

[31] E. J. Candés and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2011.

[32] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[33] J. P. Haldar and D. Hernando, "Rank-constrained solutions to linear matrix equations using powerfactorization," *IEEE Signal Process. Lett.*, vol. 16, no. 7, pp. 584–587, Jul. 2009.

[34] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. 45th Annu. ACM Symp. Theory Comput.*, 2013, pp. 665–674.

[35] D. J. Love, R. W. Heath, Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[36] S. Dasgupta and A. Gupta, "An elementary proof of a theorem of Johnson and Lindenstrauss," *Random Struct. Algorithms*, vol. 22, no. 1, pp. 60–65, 2003.

[37] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, 2003.

[38] R. Durrett, *Probability: Theory and Examples*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[39] S. M. Kay, *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*, vol. 3. London, U.K.: Pearson, 2013.

[40] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[41] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[42] Y. Peng, Y. Li, and P. Wang, "An enhanced channel estimation method for millimeter wave systems with massive antenna arrays," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1592–1595, Sep. 2015.

[43] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

Wei Zhang (S'17) received the B.S. and M.S. degrees from the Harbin Institute of Technology, Harbin, China, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong. His research interests include channel estimation, signal processing in millimeter wave MIMO systems, and low-rank matrix reconstruction.

Taejoon Kim (S'08–M'11) received the B.S. degree (Hons.) from Sogang University in 2002, the M.S. degree from KAIST in 2004, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2011. From 2011 to 2012, he was with Nokia Research Center, Berkeley, CA, USA, as a Senior Researcher. Before joining The University of Kansas as an Assistant Professor, he was a Post-Doctoral Researcher at KTH Royal Institute of Technology, Stockholm, Sweden, and an Assistant Professor with the City University of Hong Kong from 2013 to 2017. His research interests are in the design and analysis of communication systems and statistical signal processing.

He holds 26 issued U.S. patents. He received The President's Award of City University of Hong Kong in 2017. He was a recipient of the IEEE PIMRC Best Paper Award in 2012, the Nokia Research Center Kudos Award in 2012, and the IEEE Communications Society Stephen O. Rice Prize in 2016. Since 2016, he has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.

David J. Love (S'98–M'05–SM'09–F'15) received the B.S. (Hons.), M.S.E., and Ph.D. degrees in electrical engineering from The University of Texas at Austin in 2000, 2002, and 2004, respectively. Since 2004, he has been with the School of Electrical and Computer Engineering, Purdue University, where he is currently the Reilly Professor of Electrical and Computer Engineering and leads the College of Engineering Preeminent Team on Efficient Spectrum Usage. His research interests are in the design and analysis of broadband wireless communication systems, 5G wireless systems, multiple-input multiple-output (MIMO) communications, millimeter wave wireless, software-defined radios and wireless networks, coding theory, and MIMO array processing. He has held a number of editorial positions with journals such as the *IEEE Signal Processing Magazine*, the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.

He has authored over 200 technical papers and holds 27 issued U.S. patents. Along with his co-authors, he received the Best Paper Awards from the IEEE Vehicular Technology Society (2009 IEEE Transactions on Vehicular Technology Jack Neubauer Memorial Award), the IEEE Signal Processing Society (2015 IEEE Signal Processing Society Best Paper Award), and the IEEE Communications Society (2016 IEEE Communications Society Stephen O. Rice Prize). He has received multiple IEEE Global Communications Conference (GLOBECOM) Best Paper Awards.

Erik Perrins (S'96–M'05–SM'06) received the B.S. *(magna cum laude)*, M.S., and Ph.D. degrees from BYU, Provo, UT, USA, in 1997, 1998, and 2005, respectively, all in electrical engineering. From 1998 to 2004, he was with Motorola, Inc., Schaumburg, IL, USA, where he was engaged in research on land mobile radio products. Since 2004, he has also been an industry consultant on problems such as reduced-complexity receiver design, receiver synchronization, and error control coding. Since 2005, he has been with the Department of Electrical Engineering and Computer Science, The University of Kansas, Lawrence, KS, USA, where he is currently a Professor. His current research interests include digital communication theory, synchronization, channel coding, energy efficient communications, and complexity reduction in receivers.

Dr. Perrins is a member of the IEEE Communications Society. He is currently the Chair of the Communication Theory Technical Committee with the IEEE Communications Society. He was an Editor and an Area Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2007 to 2018.