Low-Rank Covariance-Assisted Downlink Training and Channel Estimation for FDD Massive MIMO Systems

Jun Fang, Xingjian Li, Hongbin Li, *Senior Member, IEEE*, and Feifei Gao, *Senior Member, IEEE*

***Abstract*— We consider the problem of downlink training and channel estimation in frequency division duplex (FDD) massive MIMO systems, where the base station (BS) equipped with a large number of antennas serves a number of single-antenna users simultaneously. To obtain the channel state information (CSI) at the BS in FDD systems, the downlink channel has to be estimated by users via downlink training and then fed back to the BS. For FDD large-scale MIMO systems, the overhead for downlink training and CSI uplink feedback could be prohibitively high, which presents a significant challenge. In this paper, we study the behavior of the minimum mean-squared error (MMSE) estimator when the channel covariance matrix has a low rank or an approximate low-rank structure. Our theoretical analysis reveals that the amount of training overhead can be substantially reduced by exploiting the low-rank property of the channel covariance matrix. In particular, we show that the MMSE estimator is able to achieve exact channel recovery in the asymptotic low-noise regime, provided that the number of pilot symbols in time is no less than the rank of the channel covariance matrix. We also present an optimal pilot design for the single-user case, and an asymptotic optimal pilot design for the multi-user scenario. Last, we develop a simple model-based scheme to estimate the channel covariance matrix, based on which the MMSE estimator can be employed to estimate the channel. The proposed scheme does not need any additional training overhead. Simulation results are provided to verify our theoretical results and illustrate the effectiveness of the proposed estimated covariance-assisted MMSE estimator.**

***Index Terms*— Massive MIMO systems, downlink training and channel estimation, channel covariance matrix, low rank structure, MMSE estimator.**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO), also known as large-scale or very-large MIMO, is a promising technology to meet the ever growing demands for higher throughput and better quality-of-service of next-generation wireless communication systems [1], [2]. Massive MIMO systems are those that are equipped with a large number of antennas at the base station (BS) simultaneously serving a much smaller number of single-antenna users sharing the same time-frequency slot. By exploiting the asymptotic orthogonality among channel vectors associated with different users, massive MIMO systems can achieve almost perfect inter-user interference cancelation with a simple linear precoder and receive combiner [3], and thus have the potential to enhance the spectrum efficiency by orders of magnitude. Massive MIMO systems can also improve the energy efficiency and enable the use of inexpensive, low-power components [4]. In addition to these benefits, it was recently shown that massive MIMO can help achieve improved estimation and detection performance in wireless sensor networks [5], [6].

To reach the full potential of massive MIMO, accurate downlink channel state information (CSI) is required at the BS for precoding and other operations. Downlink channel estimation for massive MIMO systems has been extensively studied over the past few years. Most of existing studies, e.g. [1], [3], [7], [8] assume a time division duplex (TDD) mode in which channel reciprocity between opposite links (downlink and uplink) can be exploited to facilitate the acquisition of the downlink CSI at the BS. Nevertheless, it was pointed out that the reciprocity of the wireless channel may not hold exactly due to calibration errors in the downlink/uplink RF chains [9]. Also, it is noted that current wireless cellular systems are still primarily based on the frequency division duplex (FDD). To make the massive MIMO technique backward compatible with current systems, it is of great necessity to study downlink channel estimation for FDD massive MIMO systems.

For FDD systems, the reciprocity between downlink and uplink channels no longer holds. To obtain the CSI at the transmitter, the BS needs to transmit training signals to users, and each user, after acquiring the downlink CSI through the training phase, feeds back the CSI to the BS. This downlink training and uplink feedback strategy may cause several issues. First, the feedback process may lead to delayed or even outdated CSI knowledge [10], i.e. the channel at the time it is measured differs from the channel at the time when the information is used at the transmitter. This delayed CSI knowledge can be addressed by predicting the current channel from delayed CSI knowledge via exploiting the temporal correlation of the channel [10]. Recently, it was shown that

Manuscript received August 21, 2016; revised November 30, 2016; accepted January 18, 2017. Date of publication January 24, 2017; date of current version March 8, 2017. This work was supported in part by the National Science Foundation of China under Grant 61522104 and Grant 61428103, and in part by the National Science Foundation under Grant ECCS-1408182 and Grant ECCS-1609393. The associate editor coordinating the review of this paper and approving it for publication was P. S. Rossi.

J. Fang, and X. Li are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: JunFang@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: Hongbin.Li@stevens.edu).

F. Gao is with the Institute of Information Processing, Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: feifeigao@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2017.2657513

1536-1276 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

even completely outdated CSI may still be useful by means of transmission schemes that code across multiple quasistatic blocks [11]. Another major problem is that the required amount of overhead for downlink training grows linearly with the number of transmit antennas at the BS. This may not be an issue for conventional MIMO scenarios with only a small number of antennas. However, for massive MIMO systems where the number of transmit antennas at the BS is large, the overhead for the downlink training and uplink feedback could become prohibitively high. Therefore reducing the overhead for downlink training and uplink CSIT feedback has been a central issue in FDD massive MIMO systems. A multitude of efforts has been directed towards this goal over the past few years, e.g. [12]–[20]. Specifically, in [12]–[14], the sparsity of the channel on the virtual angular domain has been leveraged to formulate downlink channel estimation as a compressed sensing problem, based on which the overhead for downlink training and uplink feedback can be substantially reduced. Recent experiments and studies (e.g. [7], [21]) show that for a typical cellular configuration with a tower-mounted BS, the angular spread of the incoming/outgoing rays at the BS is usually small, and as a result, the channel has a sparse or an approximate sparse representation on the virtual angular domain.

Besides compressed sensing-based techniques [12], [13], another line of research approaches the overhead reduction issue for FDD massive MIMO by implicitly or explicitly exploiting the low-rank structure of the channel covariance matrix, e.g. [15]–[20]. Low-rank channel covariance matrix also arises as a result of a small angular spread of the incoming/outgoing rays at the BS. Due to the narrow angular spread, different paths between the BS and the user are highly correlated, and consequently, the channel covariance matrix has a low-rank or an approximate low-rank structure with only a few dominant eigenvectors [7], [22]. In [23], it was shown that even for conventional MIMO scenarios, the dimension of the optimal pilot can be reduced if there are only a few dominant eigenvectors associated with the channel covariance matrix. Covariance-aided pilot design was also considered in [17], [18] for FDD massive MIMO systems, where open-loop and closed-loop training strategies were developed to reduce the overhead of the downlink training phase by exploiting the spatial correlation as well as the temporal correlation of the channel. In [15], [16], the dimensionality of the effective channels is reduced via a prebeamforming matrix that depends only on the channel second-order statistics (i.e. channel covariance matrix), based on which a joint spatial division and multiplexing (JSDM) scheme [15] and a beam division multiple access scheme [16] were proposed to achieve significant savings in both the downlink training and the CSIT uplink feedback.

In this paper, we continue the direction of covariance-aided downlink training and channel estimation for FDD massive MIMO systems. Specifically, we study the asymptotic behavior of the minimum mean-squared error (MMSE) estimator when the channel covariance matrix has a low-rank structure. Our theoretical results reveal that with a low-rank channel covariance matrix, the MMSE estimator employing a random

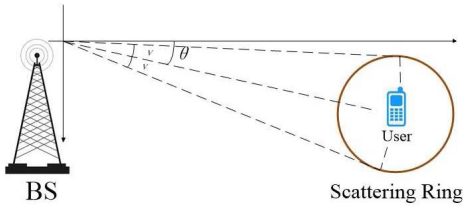


Fig. 1. Schematic for the one-ring model.

(not necessarily optimal) pilot can obtain a perfect channel recovery in the limit of vanishing noise, provided that the length of the pilot (i.e. the number of symbols in time) is no less than the rank of the covariance matrix. We also examine asymptotically optimal pilot design for the multi-user scenario. An overlayed training strategy similar to the JSDM scheme is proposed and shown to be asymptotically optimal in terms of estimation errors when users have mutually non-overlapping angles of arrival (AoAs). The optimal design suggests that the minimum MSE can be achieved as long as the length of pilot is no less than the rank of the channel covariance matrix. In addition, based on the classical one-ring scattering model (see, e.g. [15]), we develop a simple model-based scheme to estimate the channel covariance matrix. The proposed scheme does not require any additional training overhead. Simulation results show that the proposed estimated covariance-assisted MMSE estimator achieves a substantial performance improvement over the compressed sensing-based methods.

The following notations are adopted throughout this paper, where $(\cdot)^T$ and $(\cdot)^H$ represent the transpose and conjugate transpose, respectively, $\mathrm{tr}(\boldsymbol{A})$ denotes the trace of $\boldsymbol{A}$, and $\|\boldsymbol{x}\|_2$ is used to denote the ℓ_2 norm of vector $\boldsymbol{x}$. We let $\mathrm{Range}(\boldsymbol{A})$ denote the column space spanned by the column vectors of $\boldsymbol{A}$, span{} denotes the subspace spanned by the vectors defined in the set {}, $\mathbb{C}^{n\times m}$ and $\mathbb{C}^n$ denote the set of $n \times m$ matrices and the set of n-dimensional column vectors with complex entries, respectively.

The rest of this paper is organized as follows. In Section II, we introduce the system model and basic assumptions. The asymptotic behavior of the MMSE estimator in the limit of vanishing noise is examined in Section III. An optimal pilot design for the single-user scenario and an asymptotic optimal pilot design for the multi-user scenario are studied in Sections IV and V, respectively. In Section VI, we develop a simple model-based scheme to estimate the channel covariance matrix, and construct a MMSE estimator to estimate the channel. Simulation results are provided in Section VII, followed by concluding remarks in Section VIII.

II. System Model and Problem Formulation

We consider the problem of downlink training and channel estimation in a FDD massive MIMO system, where the BS equipped with a large number of antennas serves a number of single-antenna users simultaneously. To simplify our problem, we consider the single-user scenario. The extension of our results to the multi-user scenario is straightforward, and the pilot design for the multi-user case will be discussed in Section V. We assume the channel $\boldsymbol{h} \in \mathbb{C}^M$ is a flat

Rayleigh fading channel under a narrowband assumption, where M denotes the number of transmit antennas at the BS. The extension to the wideband frequency-selective channel is straightforward when an orthogonal frequency-division multiplexing (OFDM) transmission scheme is adopted. The signal received by the user can be expressed as

$$y_t = \boldsymbol{x}_t^T \boldsymbol{h} + w_t \quad \forall t = 1, \ldots, T \tag{1}$$

where $\boldsymbol{x}_t \in \mathbb{C}^M$ is the transmitted pilot symbol vector at time t, and w_t denotes the additive white Gaussian noise with zero mean and variance σ^2. Define $\boldsymbol{y} \triangleq [y_1 \ y_2 \ \ldots \ y_T]^T$, $\boldsymbol{X} \triangleq [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \ldots \ \boldsymbol{x}_T]^T$, and $\boldsymbol{w} \triangleq [w_1 \ w_2 \ \ldots \ w_T]^T$. The data model (1) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{w} \tag{2}$$

Note that to simplify our problem, we ignore the inter-cell interference resulted from frequency reuse from neighboring cells. Inter-cell interference may pose a problem for channel estimation, and how to cope with inter-cell interference is an important issue worthy of future investigation.

In this paper, we consider the classical one-ring model that has been widely adopted (e.g. [7], [13], [15]) to characterize the massive MIMO channel, where the BS is assumed to be located in an elevated position with few scatterers around, and the propagation between the BS and the user is mainly characterized by rich local scatterers around the user (see Fig. 1). Assuming the propagation consists of P i.i.d. paths, we have

$$\boldsymbol{h} = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} \alpha_p \boldsymbol{a}(\theta_p) \tag{3}$$

where $\alpha_p \sim \mathcal{CN}(0, \xi^2)$ denotes the fading coefficient associated with the pth path, and $\boldsymbol{a}(\theta_p)$ is the steering vector. For a uniform linear array, it is given as

$$\boldsymbol{a}(\theta_p) \triangleq [1 \ \ e^{-j(2\pi/\chi)d\cos(\theta_p)} \ \ \ldots \ \ e^{-j(M-1)(2\pi/\chi)d\cos(\theta_p)}]^T \tag{4}$$

in which χ is the signal wavelength, d denotes the distance between neighboring antenna elements, and $\theta_p \in [0, \pi]$ is the azimuth AoA of the pth path. In the one-ring mode, the user is surrounded by rich local scatterers with a radius r that is relatively small compared to the distance between the BS and the user, D. Thus the angular spread at the BS, approximately given as $\delta = \arctan(r/D)$, is small.

To estimate the channel from the received signal $\boldsymbol{y}$ (c.f. (2)), it is usually required that the number of pilot symbols (in time), T, is no less than the number of transmitted antennas M, i.e. $T \geq M$. When M is large, the overhead for downlink training and uplink CSI feedback becomes prohibitively high. Hopefully, due to the narrow angular spread at the BS, the steering vectors $\{\boldsymbol{a}(\theta_p)\}$ of these P paths are highly correlated, and thus the channel covariance matrix $\boldsymbol{R} = E[\boldsymbol{h}\boldsymbol{h}^H]$ has an approximate low-rank structure. In particular, in the asymptotic regime $M \to \infty$, the rank of the channel covariance matrix is upper bounded by a small quantity that is determined by the angular spread (c.f. (16)). This low-rank structure can be utilized to reduce the overhead for downlink training for FDD systems, see, e.g. [15], [17], [18].

In this paper, we study the behavior of the MMSE estimator when the the channel covariance matrix has a low rank structure. We conduct a quantitative analysis to investigate how much training overhead reduction can be achieved by exploiting the low-rank structure of the channel covariance matrix. Assume $\boldsymbol{h}$ is zero-mean complex Gaussian with covariance matrix $\boldsymbol{R}$, the MMSE estimate of the channel $\boldsymbol{h}$ is given as

$$\hat{\boldsymbol{h}} = \boldsymbol{R}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{R}\boldsymbol{X}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y} \tag{5}$$

Note that the MMSE estimator, with the aid of the statistical information of the channel, does not require an invertible pilot matrix $\boldsymbol{X}$ (i.e. $T \geq M$) for channel estimation. The mean-squared error (MSE) associated with the MMSE estimate is given by

$$\begin{aligned} \text{MSE} &= E\big[\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|_2^2\big] \\ &= \text{tr}\big(\boldsymbol{R} - \boldsymbol{R}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{R}\boldsymbol{X}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{X}\boldsymbol{R}\big) \end{aligned} \tag{6}$$

III. Asymptotic Behavior of The MMSE

A. Main Results

In this section, we first study the behavior of the MMSE estimator in the asymptotic low-noise regime, i.e. $\sigma^2 \to 0$. Our asymptotic analysis shows that a perfect channel recovery from a small number of symbols is possible when the channel covariance matrix has a low-rank structure. Our main results are summarized as follows.

Theorem 1: Consider the channel estimation problem described in (2), where $\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$ and the rank of the channel covariance matrix $\boldsymbol{R}$ is $r = \text{rank}(\boldsymbol{R})$. Define $\boldsymbol{\Phi} \triangleq \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{X}^H\boldsymbol{X}\boldsymbol{R}^{\frac{1}{2}}$. Let $\boldsymbol{\Phi} = \boldsymbol{V}\boldsymbol{\Gamma}\boldsymbol{V}^H$ denote the eigenvalue decomposition (EVD) of $\boldsymbol{\Phi}$, where $\boldsymbol{V} \triangleq [\boldsymbol{v}_1 \ \ldots \ \boldsymbol{v}_M]$ is a unitary matrix consisting of eigenvectors of $\boldsymbol{\Phi}$, and $\boldsymbol{\Gamma} = \text{diag}(\gamma_1, \ldots, \gamma_r, 0, \ldots, 0)$ is a diagonal matrix with $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_r > 0$. Suppose the pilot signal $\boldsymbol{X}$ is randomly generated, and the number of symbols, T, is no less than r, i.e. $T \geq r$, then the MSE of the MMSE estimate of $\boldsymbol{h}$ is given by

$$E\left[\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|_2^2\right] = \sum_{i=1}^{r} \left(1 + \gamma_i/\sigma^2\right)^{-1} \boldsymbol{v}_i^H \boldsymbol{R} \boldsymbol{v}_i \tag{7}$$

and the MSE approaches zero in the limit of vanishing noise, that is,

$$\lim_{\sigma^2 \to 0} E\left[\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|_2^2\right] = 0$$

Proof: Using the Woodbury identity, the MSE (6) can be rewritten as

$$\begin{aligned} &E\left[\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|_2^2\right] \\ &= \text{tr}\left(\boldsymbol{R}^{1/2}(\boldsymbol{I} - \boldsymbol{R}^{1/2}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{R}\boldsymbol{X}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{X}\boldsymbol{R}^{1/2})\boldsymbol{R}^{1/2}\right) \\ &= \text{tr}\left(\boldsymbol{R}^{1/2}(\boldsymbol{I} + \sigma^{-2}\boldsymbol{R}^{1/2}\boldsymbol{X}^H\boldsymbol{X}\boldsymbol{R}^{1/2})^{-1}\boldsymbol{R}^{1/2}\right) \\ &= \text{tr}\left(\boldsymbol{R}\boldsymbol{V}(\sigma^{-2}\boldsymbol{\Gamma} + \boldsymbol{I})^{-1}\boldsymbol{V}^H\right) \\ &= \sum_{i=1}^{r} \left(1 + \gamma_i/\sigma^2\right)^{-1} \boldsymbol{v}_i^H \boldsymbol{R} \boldsymbol{v}_i + \sum_{i=r+1}^{M} \boldsymbol{v}_i^H \boldsymbol{R} \boldsymbol{v}_i \end{aligned} \tag{8}$$

We can see that the first term in (8) vanishes as $\sigma^2 \to 0$.

We now examine under what conditions the second term in (8) reduces to zero. Let

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H \tag{9}$$

denote the reduced EVD of $\boldsymbol{R}$, where $\boldsymbol{U} \in \mathbb{C}^{M\times r}$ and $\boldsymbol{\Lambda} \in \mathbb{C}^{r\times r}$. We can write

$$\boldsymbol{R}^{\frac{1}{2}}\boldsymbol{X}^H = \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{U}^H\boldsymbol{X}^H = \boldsymbol{U}\boldsymbol{C} \tag{10}$$

where $\boldsymbol{C} \triangleq \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{U}^H\boldsymbol{X}^H \in \mathbb{C}^{r\times T}$. When $T \geq r$ and the pilot symbols of $\boldsymbol{X}$ are randomly generated according to some distribution, the matrix $\boldsymbol{C}$ has a full row rank with probability one, i.e. $\text{rank}(\boldsymbol{C}) = r$. Thus we have

$$\text{Range}(\boldsymbol{R}^{\frac{1}{2}}\boldsymbol{X}^H) = \text{Range}(\boldsymbol{\Phi}) = \text{Range}(\boldsymbol{U}) \tag{11}$$

From (11), we can immediately arrive at

$$\boldsymbol{u}_i^H\boldsymbol{R} = \boldsymbol{v}_i^H\boldsymbol{U} = \boldsymbol{v}_i^H\boldsymbol{\Phi} = \boldsymbol{0} \quad \forall i = r+1,\ldots,M \tag{12}$$

Hence the second term in (8) disappears provided that the length of the pilot in time is no less than the rank of the channel covariance matrix, i.e. $T \geq r$, and eventually we reach the conclusion that the MSE of the MMSE estimate of $\boldsymbol{h}$ approaches zero in the limit of vanishing noise, that is,

$$\lim_{\sigma^2\to 0} E\left[\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|_2^2\right] = 0$$

The proof is completed here. ■

B. Discussions

The significance of Theorem 1 lies in that, in the limit of vanishing noise, it establishes sufficient conditions for the MMSE estimator to achieve exact channel recovery from only a small number of pilot symbols. If we write $\boldsymbol{h} = \boldsymbol{R}^{1/2}\boldsymbol{d}$ and treat $\boldsymbol{d}$ (i.e. $\boldsymbol{h}$) as a deterministic vector, a least squares estimator can be used to estimate the channel. It can be easily shown that for the deterministic case, the exact channel recovery result also holds valid in the asymptotic low-noise regime, provided that $T \geq r$ is satisfied. The proof of the exact recovery result for the deterministic case, however, is very different from the proof for Theorem 1 due to the difference in modeling and estimation. After the completion of this work, we noticed similar results were reported in [24] under a different framework for signal reconstruction with compressive measurements.

Another line of research [12], [13] for FDD downlink training and channel estimation exploits the sparsity of the channel on the virtual angular domain and formulates the channel estimation as a compressed sensing problem:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{w} = \boldsymbol{X}\boldsymbol{A}\tilde{\boldsymbol{h}} + \boldsymbol{w} \tag{13}$$

where $\boldsymbol{A}$ is a basis for the virtual angular domain. For the uniform linear array case, the basis $\boldsymbol{A}$ is a discrete Fourier transform (DFT) matrix. $\tilde{\boldsymbol{h}}$ is a sparse vector to be estimated. This class of approaches are justified by compressed sensing theories, which assert that a sparse signal can be perfectly recovered from compressive measurements, provided that the measurement matrix satisfies a certain restricted isometry property (RIP) condition [25]. Our theorem here can be regarded as a counterpart result for the MMSE estimator, and provides a justification for using the MMSE estimator for channel estimation from a small number of pilot symbols.

It is also interesting to compare conditions required by the MMSE estimator and those by compressed sensing techniques to achieve perfect channel recovery. First recall the following lemma that characterizes the number of dimensions of a subspace spanned by a number of steering vectors with a bounded support of AoAs:

Lemma 1: Define

$$\boldsymbol{\alpha}(x) \triangleq [1 \;\; e^{-j\pi x} \;\; \ldots \;\; e^{-j\pi(M-1)x}]^T \tag{14}$$

and $\mathcal{A} \triangleq \text{span}\{\boldsymbol{\alpha}(x), x \in [-1,1)\}$. Given $b_1, b_2 \in [-1,1]$ and $b_1 < b_2$, define $\mathcal{B} \triangleq \text{span}\{\boldsymbol{\alpha}(x), x \in [b_1,b_2]\}$, then

$$\begin{aligned} &\dim(\mathcal{A}) = M \\ &\dim(\mathcal{B}) \sim (b_2-b_1)M/2 \text{ when } M \text{ grows large} \end{aligned} \tag{15}$$

Proof: See [7, Lemma 1]. ■

Consider the one-ring model (3) with the multipath angle of arrival θ distributed on a bounded support, i.e. $\theta \in [\theta_{\min}, \theta_{\max}]$. From Lemma 1, the rank of the channel covariance matrix $\boldsymbol{R}$ is upper bounded by

$$\text{rank}(\boldsymbol{R}) \leq \eta M \quad \text{as } M \to \infty \tag{16}$$

where η is defined as

$$\eta \triangleq |\cos(\theta_{\min}) - \cos(\theta_{\max})|d/\chi \tag{17}$$

in which d denotes the distance between neighboring antennas and χ is the signal wavelength. Another important property from Lemma 1 is that, when $M \to \infty$, the channel $\boldsymbol{h}$ has a sparse representation on a virtual angular domain with $r = \text{rank}(\boldsymbol{R})$ nonzero coefficients.

With the above results, we are now ready to make a fair comparison between conditions required by the MMSE estimator and, respectively, by the compressed sensing methods for exact recovery of the channel. For the MMSE estimator, from Theorem 1, we know that as few as $T = r$ symbols are needed to perfectly recover the channel. On the other hand, for compressed sensing-based methods, it has been shown that the number of required measurements for exact recovery is of order $T = O(r\log(M/r))$ using polynomial-time optimization solvers or greedy algorithms [25]. If the computational complexity is not a concern, then at least $T = 2r$ measurements are required for exact recovery via the ℓ_0-minimization. From the above discussion, we can see that the MMSE estimator requires fewer symbols than compressed sensing techniques for exact channel recovery. This result puts the covariance-aided methods into a favorable position for FDD downlink training and channel estimation.

IV. Optimal Pilot Sequence Design

Our analysis in the previous section reveals that as few as $T = r$ symbols in time are required to guarantee perfect channel recovery in the asymptotic low-noise regime, i.e. $\sigma^2 \to 0$. Nevertheless, assuming a noiseless scenario is unrealistic in practical systems. Therefore it is meaningful to study the behavior of the MMSE estimator for a non-vanishing σ^2.

For the case $\sigma^2 \neq 0$, we would like to examine whether a larger value of T leads to a better estimation accuracy, or if $T = r$ is sufficient to attain a minimum MSE. To answer this question, we first need to impose a power constraint on the pilot signal, i.e. $\mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) \leq P$; otherwise a fair comparison between pilots of different lengths is impossible. Note that different pilots of the same length also result in different MSEs. Hence simply comparing the MSEs attained by two arbitrary pilots of different lengths does not provide any meaningful answers. To make sense, we have to compare the MSEs attained by optimally devised pilots for different values of T, and see if increasing T will result in a lower MSE. This requires us to examine the following optimization problem

$$\begin{aligned} &\min_{\boldsymbol{X}} \ \mathrm{MSE} = \mathrm{tr}\left(\boldsymbol{R} - \boldsymbol{R}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{R}\boldsymbol{X}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{X}\boldsymbol{R}\right) \\ &\text{s.t. } \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) \leq P \end{aligned} \tag{18}$$

The solution of the above optimization problem is summarized as follows.

Theorem 2: Let $\boldsymbol{R} = \boldsymbol{U}_0\boldsymbol{\Lambda}_0\boldsymbol{U}_0^H$ denote the EVD[1] of $\boldsymbol{R}$, where $\boldsymbol{\Lambda}_0 = \mathrm{diag}(\lambda_1, \ldots, \lambda_M)$ is a diagonal matrix with its diagonal entries arranged in a decreasing order and $\boldsymbol{U}_0 \in \mathbb{C}^{M\times M}$ is a unitary matrix. The optimal solution to (18) is then given by

$$\boldsymbol{X} = [\boldsymbol{\Delta} \ \ \boldsymbol{0}]\boldsymbol{U}_0^H \tag{19}$$

where $\boldsymbol{\Delta} = \mathrm{diag}(\delta_1, \ldots, \delta_T)$ with δ_i given as

$$\delta_i = \begin{cases} \sqrt{\mu - \sigma^2\lambda_i^{-1}} & \text{if } \mu \geq \sigma^2\lambda_i^{-1} \text{ and } \lambda_i \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

in which μ is determined by the constraint $\sum_{i=1}^T \delta_i^2 = P$.

Proof: According to [26, Theorem 1], the optimal $\boldsymbol{X}$ has a form of

$$\boldsymbol{X}^H = \boldsymbol{U}_0[:, 1:T]\boldsymbol{\Delta} \tag{21}$$

where $\boldsymbol{U}_0[:, 1:T]$ consists of T eigenvectors of $\boldsymbol{R}$ associated with the first T largest eigenvalues, and $\boldsymbol{\Delta} = \mathrm{diag}(\delta_1, \ldots, \delta_T)$ is a diagonal matrix with its diagonal elements to be determined as follows. Substituting (21) into (18), the optimization (18) can be simplified as

$$\begin{aligned} &\min_{\{\delta_i\}} \ \sum_{i=1}^T \frac{\sigma^2\lambda_i}{\delta_i^2\lambda_i + \sigma^2} \\ &\text{s.t. } \sum_{i=1}^T \delta_i^2 \leq P \\ &\quad \delta_i^2 \geq 0 \quad \forall i = 1, \ldots, T \end{aligned} \tag{22}$$

The above optimization can be solved analytically by resorting to the Lagrangian function and Karush-Kuhn-Tucker (KKT) conditions, which leads to a water-filling type power allocation scheme described by

$$\delta_i = \begin{cases} \sqrt{\mu - \sigma^2\lambda_i^{-1}} & \text{if } \mu \geq \sigma^2\lambda_i^{-1} \text{ and } \lambda_i \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

[1]Here $\boldsymbol{R} = \boldsymbol{U}_0\boldsymbol{\Lambda}_0\boldsymbol{U}_0^H$ is used to distinguish itself from the truncated EVD $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$.

where μ is determined to ensure that the KKT condition $\sum_{i=1}^T \delta_i^2 = P$ is satisfied. The proof is completed here. ■

We now discuss whether a larger value of T would result in a lower MSE. Note that the MSE achieved by the optimal $\boldsymbol{X}$ is given by

$$\mathrm{MSE}(T) = \sum_{i=1}^r \lambda_i - \sum_{i=1}^T \frac{\delta_i^2(T)\lambda_i^2}{\delta_i^2(T)\lambda_i + \sigma^2} \tag{24}$$

where we use $\delta_i(T)$ to denote the dependence of δ_i on T. The r-rank channel covariance matrix $\boldsymbol{R}$ implies $\lambda_i = 0, \forall i > r$. Considering the case $T > r$, from (20), it is easy to verify that for any $T > r$, we have

$$\delta_i(T) = \begin{cases} \delta_i(r) & \forall i = 1, \ldots, r \\ 0 & \forall i = r+1, \ldots, T \end{cases} \tag{25}$$

Therefore we can arrive at

$$\mathrm{MSE}(T) = \mathrm{MSE}(r) \quad \forall T > r \tag{26}$$

On the other hand, from the optimality of the solution (20), it is clear that

$$\mathrm{MSE}(T) \leq \mathrm{MSE}(r) \quad \forall T < r \tag{27}$$

Based on the above results, we know that the minimum MSE can be attained by simply choosing $T = r$, and a larger T beyond the value of r does not lead to a smaller MSE. This result provides an affirmative answer to the question discussed at the beginning of this section, that is, given a transmit power constraint $\mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) = P$, a minimum MSE can be achieved by setting the number of symbols equal to the rank of the channel covariance matrix, i.e. $T = r$.

Remark 1: Note that the pilot constraint considered here is different from that of [17], where unitary training with equal power allocation per pilot symbol is assumed, i.e. $\boldsymbol{X}\boldsymbol{X}^H = \rho\boldsymbol{I}$. Clearly, for the pilot constraint adopted in [17], the total amount of transmit power increases unbounded as T becomes large, more precisely, we have $\mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) = \rho T$. Hence a larger T always leads to an improved channel estimation accuracy. Overall, the pilot constraint considered in our paper exploits the channel structure more thoroughly, and hence makes use of the power more efficiently, while the pilot constraint in [17] leads to simpler implementation and is free from the peak-power problem.

Remark 2: Our results in Theorem 2 are closely related to the results in [23], [27]. In fact, the optimization problem (18) in our paper can be deemed as a special case of the optimization problem (9) in [23], with a single antenna at the receiver and i.i.d. Gaussian noise. Nevertheless, [23], [27] assume a general channel covariance matrix in considering the problem of optimal pilot design. The authors did not explicitly address the case where the channel covariance matrix has a low-rank structure. As indicated in [23], for a general full-rank covariance matrix, training overhead reduction may also be achieved due to the fact that a water-filling solution could assign all power to a few strong eigendirections and no power to other weak eigendirections. But, as pointed out in [23], the minimum length of the optimal pilot required to achieve the minimum MSE cannot be determined analytically for the

general covariance case. Therefore it is unclear exactly how much training overhead reduction can be achieved. Different from [23], [27], our work explicitly assumes a low-rank channel covariance matrix. Our analysis shows that the optimal pilot under this low-rank assumption has a more concise form and the required number of the pilot symbols to attain the minimum MSE is no less than the rank of the covariance matrix. This result, although easy to be deduced, has not been explicitly reported in previous works.

V. Asymptotically Optimal Pilot for Multi-User Scenarios

In the previous section, we derived the optimal pilot sequence for the single-user case. For massive MIMO systems where the BS aims to simultaneously serve a number of users, the pilot sequence has to be shared by multiple users. Unfortunately, the channels associated with these users may not have the same channel covariance matrix. In this case, it is impossible to find an optimal pilot sequence $\boldsymbol{X}$ to simultaneously minimize the MSEs associated with all users. To address this difficulty, in [19], the pilot sequence is designed to maximize a summation of the conditional mutual information associated with all users, and an iterative algorithm was developed to solve the maximization problem. In this section, a different criterion is considered, where the objective is to minimize the sum of MSEs associated with all users, i.e.

$$\begin{aligned}\min_{\boldsymbol{X}} \ & \sum_{k=1}^{K} \mathrm{MSE}_k \\ &= \sum_{k=1}^{K} \mathrm{tr}\left(\boldsymbol{R}_k - \boldsymbol{R}_k \boldsymbol{X}^H (\boldsymbol{X}\boldsymbol{R}_k\boldsymbol{X}^H + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{X}\boldsymbol{R}_k\right) \\ &\text{s.t. } \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) \le P \end{aligned} \tag{28}$$

where $\boldsymbol{R}_k$ and MSE_k denote the channel covariance matrix and the MSE associated with the kth user, respectively. Also, to simplify the problem, we assume the noise variances across different users are identical, i.e. $\sigma_1^2 = \ldots = \sigma_K^2 = \sigma^2$. Finding an analytical solution to the above optimization is difficult. Nevertheless, we will show that an asymptotically optimal training sequence can be devised given that *users have mutually non-overlapping AoAs*. Here the asymptotic optimality means that the solution approaches the optimal one as the number of antennas at the BS goes to infinity. Note that non-overlapping AoAs among users imply spatial separation in the bearing space. As a result, subspaces spanned by different users' channel covariance matrices are approximately orthogonal to each other. We will show that this approximate orthogonality property plays a vital role in decoupling and solving the optimization (28).

Before proceeding, we first introduce the following properties which were proved in [7], [28] and reveal the eigenstructure properties of the channel covariance matrices. Consider the channel $\boldsymbol{h}$ generated by the one ring model with a bounded support of angle of arrival $\theta \in [\theta_{\min}, \theta_{\max}]$. Let $\boldsymbol{R}$ denote the channel covariance matrix. We have the following properties regarding the channel covariance matrix.

Property 1 [7, Lemma 3]: In the asymptotic regime of large number of antennas, steering vectors $\boldsymbol{a}(\vartheta)$ with $\vartheta \notin [\theta_{\min}, \theta_{\max}]$ fall in the null space of the covariance matrix $\boldsymbol{R}$, i.e.

$$\mathrm{null}(\boldsymbol{R}) \supset \mathrm{span}\{\boldsymbol{a}(\vartheta)/\sqrt{M}, \forall \vartheta \notin [\theta_{\min}, \theta_{\max}]\}, \quad \text{as } M \to \infty \tag{29}$$

Property 2 [28, Lemma 1]: For a uniform linear array, when $M \to \infty$, the eigenvector matrix of the channel covariance matrix $\boldsymbol{R}$ can be well approximated by a unitary DFT matrix.

Property 3: From the above two properties, we naturally arrive at the following property: The column vectors in the DFT matrix whose angular coordinates are located outside the support of angle of arrival form an orthonormal basis for the null space of $\boldsymbol{R}$. Meanwhile, those column vectors in the DFT matrix whose angular coordinates lie within the support of angle of arrival form an orthonormal basis for $\boldsymbol{R}$. More precisely, let

$$\boldsymbol{F} \triangleq \frac{1}{\sqrt{M}}[\boldsymbol{\alpha}(\omega_1) \ \ \boldsymbol{\alpha}(\omega_2) \ \ \ldots \ \ \boldsymbol{\alpha}(\omega_M)] \tag{30}$$

denote the DFT matrix, in which $\omega_m = -1 + 2(m-1)/M, \forall m$, and $\boldsymbol{\alpha}(\omega_m)$ is defined in (14). Let $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$ denote the truncated eigenvalue decomposition, where $\boldsymbol{U} \in \mathbb{C}^{M\times r}$, and $\boldsymbol{\Lambda} \in \mathbb{C}^{r\times r}$. Then as $M \to \infty$, $\boldsymbol{U}$ is composed of column vectors of $\boldsymbol{F}$ whose angular coordinates $\{\omega_i\}$ lie within the support of AoA, i.e.

$$\boldsymbol{U} = [\boldsymbol{\alpha}(\omega_{i_1}) \ \ \ldots \ \ \boldsymbol{\alpha}(\omega_{i_r})] \tag{31}$$

where $\omega_i \in [2d\cos(\theta_{\min})/\chi, \ 2d\cos(\theta_{\max})/\chi]$ for $i = i_1, \ldots, i_r$.

We now discuss how to devise an asymptotically optimal pilot sequence for (28). Let

$$\boldsymbol{R}_k = \boldsymbol{U}_k \boldsymbol{\Lambda}_k \boldsymbol{U}_k^H \tag{32}$$

denote the truncated eigenvalue decomposition, where $\boldsymbol{U}_k \in \mathbb{C}^{M\times r_k}$, and $\boldsymbol{\Lambda}_k \in \mathbb{C}^{r_k\times r_k}$. r_k denotes the rank of $\boldsymbol{R}_k$. For simplicity, we assume $r_k = r, \forall k$. Inspired by the above properties, we propose an overlayed pilot sequence that is a superposition of a set of pilot sequences $\{\boldsymbol{X}_k\}$

$$\boldsymbol{X} = \sum_{k=1}^{K} \boldsymbol{X}_k \tag{33}$$

where $\boldsymbol{X}_k$ denotes the pilot sequence optimally designed for user k, i.e. given a power constraint $\mathrm{tr}(\boldsymbol{X}_k\boldsymbol{X}_k^H) = P_k^*$, $\boldsymbol{X}_k$ is given by Theorem 2, i.e.

$$\boldsymbol{X}_k = \boldsymbol{\Delta}_k^H \boldsymbol{U}_k^H \tag{34}$$

in which $\boldsymbol{\Delta}_k$ is a diagonal matrix with its diagonal elements optimized according to a water-filling power allocation scheme as described in Theorem 2. As to be shown in the following, the overlayed solution (33) arises as a result of the fact that non-overlapping bearing space enables the objective function (28) to be decoupled into K independent sub-problems.

We now show that the asymptotically optimal solution to (28) has a form of (33). Note that any pilot sequence $\boldsymbol{X}$ can be expressed in terms of the DFT matrix as follows

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{F}^H \tag{35}$$

where $\boldsymbol{Z} \in \mathbb{C}^{T\times M}$ is a matrix to be optimized. Recalling Properties 2 and 3, we have

$$\begin{aligned}\boldsymbol{X}\boldsymbol{R}_k &= \boldsymbol{Z}\boldsymbol{F}^H\boldsymbol{R}_k \overset{(a)}{=} (\boldsymbol{Z}_k\boldsymbol{U}_k^H + \bar{\boldsymbol{Z}}_k\bar{\boldsymbol{U}}_k^H)\boldsymbol{R}_k \\ &= \boldsymbol{Z}_k\boldsymbol{U}_k^H\boldsymbol{R}_k \end{aligned} \tag{36}$$

where (a) comes from the fact that we can partition the DFT matrix into two parts $\boldsymbol{F} = [\boldsymbol{U}_k\ \bar{\boldsymbol{U}}_k]$, in which $\boldsymbol{U}_k$ is an orthonormal basis of $\boldsymbol{R}_k$ and $\bar{\boldsymbol{U}}_k$ is an orthonormal basis for the null space of $\boldsymbol{R}_k$. Accordingly, $\boldsymbol{Z}$ can be partitioned into two parts: $\boldsymbol{Z} = [\boldsymbol{Z}_k\ \bar{\boldsymbol{Z}}_k]$, where $\boldsymbol{Z}_k \in \mathbb{C}^{T\times r}$ is a submatrix of $\boldsymbol{Z}$ consisting of r column vectors. Substituting (36) into the objective function (28), we have

$$\begin{aligned}&\sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{R}_k - \boldsymbol{R}_k\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{R}_k\boldsymbol{X}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{X}\boldsymbol{R}_k\right) \\ &\quad= \sum_{k=1}^{K}\mathrm{tr}\Big(\boldsymbol{R}_k - \boldsymbol{R}_k\boldsymbol{U}_k\boldsymbol{Z}_k^H(\boldsymbol{Z}_k\boldsymbol{U}_k^H\boldsymbol{R}_k\boldsymbol{U}_k\boldsymbol{Z}_k^H + \sigma^2\boldsymbol{I})^{-1} \\ &\qquad\qquad \boldsymbol{Z}_k\boldsymbol{U}_k^H\boldsymbol{R}_k\Big)\end{aligned} \tag{37}$$

Since users have mutually non-overlapping AoAs, each matrix $\boldsymbol{Z}_k$ is constructed by r unique columns of $\boldsymbol{Z}$ that are not shared by other matrices $\boldsymbol{Z}_{\bar{k}}, \forall \bar{k} \neq k$. Therefore the optimization (28) can be decomposed into K independent problems, with $\boldsymbol{Z}_k$ optimized in each individual problem

$$\begin{aligned}\min_{\boldsymbol{Z}_k}\ &\mathrm{tr}\Big(\boldsymbol{R}_k - \boldsymbol{R}_k\boldsymbol{U}_k\boldsymbol{Z}_k^H(\boldsymbol{Z}_k\boldsymbol{U}_k^H\boldsymbol{R}_k\boldsymbol{U}_k\boldsymbol{Z}_k^H + \sigma^2\boldsymbol{I})^{-1} \\ &\qquad \boldsymbol{Z}_k\boldsymbol{U}_k^H\boldsymbol{R}_k\Big) \\ \text{s.t. }&\mathrm{tr}(\boldsymbol{Z}_k\boldsymbol{Z}_k^H) = P_k^*\end{aligned} \tag{38}$$

where P_k^* is the optimal power allocated to the kth user. From Theorem 2, we know that setting $T = r$ is sufficient to achieve a minimum MSE and the optimal $\boldsymbol{Z}_k$ is a diagonal matrix

$$\boldsymbol{Z}_k = \boldsymbol{\Delta}_k^H \tag{39}$$

with its diagonal elements determined according to a water-filling power allocation scheme (see Theorem 2) such that the constraint $\mathrm{tr}(\boldsymbol{Z}_k\boldsymbol{Z}_k^H) = P_k^*$ is satisfied. For those columns of $\boldsymbol{Z}$ that are not included in $\{\boldsymbol{Z}_k\}_{k=1}^K$, since they make no difference to the objective function value, they should be set to zero in order to save the transmit power. Therefore the asymptotically optimal pilot signal $\boldsymbol{X}$ can be written as

$$\begin{aligned}\boldsymbol{X} &= \sum_{k=1}^{K}\boldsymbol{Z}_k\boldsymbol{U}_k^H = \sum_{k=1}^{K}\boldsymbol{\Delta}_k^H\boldsymbol{U}_k^H \\ &= \sum_{k=1}^{K}\boldsymbol{X}_k\end{aligned} \tag{40}$$

which is a superposition of a set of pilot sequences, with each pilot sequence optimally designed for each individual user.

To determine the optimal power allocation $\{P_k^*\}$, we substitute (40) back into the original optimization (28). After some simplifications, we arrive at

$$\begin{aligned}\min_{\{\boldsymbol{\Delta}_k\}}\ &\sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{\Lambda}_k - \boldsymbol{\Lambda}_k\boldsymbol{\Delta}_k(\boldsymbol{\Delta}_k^H\boldsymbol{\Lambda}_k\boldsymbol{\Delta}_k + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{\Delta}_k^H\boldsymbol{\Lambda}_k\right) \\ \text{s.t. }&\sum_{k=1}^{K}\mathrm{tr}(\boldsymbol{\Delta}_k\boldsymbol{\Delta}_k^H) \le P\end{aligned} \tag{41}$$

Let $\boldsymbol{\Delta}_k = \mathrm{diag}(\delta_{k,1}, \ldots, \delta_{k,r})$. The above optimization can be further simplified as

$$\begin{aligned}\min_{\delta_{k,i}}\ &\sum_{k=1}^{K}\sum_{i=1}^{r}\frac{\sigma^2\lambda_{k,i}}{\delta_{k,i}^2\lambda_{k,i} + \sigma^2} \\ \text{s.t. }&\sum_{k=1}^{K}\sum_{i=1}^{r}\delta_{k,i}^2 \le P \\ &\delta_{k,i}^2 \ge 0 \quad \forall k = 1, \ldots, K\ \forall i = 1, \ldots, r\end{aligned} \tag{42}$$

where $\lambda_{k,i}$ denotes the ith eigenvalue of $\boldsymbol{R}_k$. Similar to (22), (42) can be analytically solved by resorting to KKT conditions, based on which the optimal power allocation can be obtained.

Remark 1: The above overlayed pilot design has an intuitive explanation. Given that the AoAs of all users are distinct, from Property 1, we know that the channel of each user is asymptotically orthogonal to the channel covariance matrices associated with other users as $M \to \infty$, i.e. $\boldsymbol{h}_k^H\boldsymbol{R}_{k'} = \boldsymbol{0}, \forall k \neq k'$. As a result, we have $\boldsymbol{X}_{k'}\boldsymbol{h}_k = \boldsymbol{0}, \forall k \neq k'$ for the pilot sequence $\{\boldsymbol{X}_{k'}\}$ devised in (34). Hence from the user's perspective, only the optimal pilot signal will be received, while other non-optimal pilot signals are filtered when propagating through the channel.

Remark 2: The proposed overlayed downlink training scheme bears a resemblance to the JSDM strategy [15], where a prebeamforming matrix is employed to reduce the dimension of the channel to be estimated. In particular, the prebeamforming matrix suggested by [15] is a concatenation of $\{\boldsymbol{U}_k\}_{k=1}^K$. Although both the proposed overlayed training scheme and the JSDM scheme use the eigenvectors of the channel covariance matrices for downlink training, the rationale behind these two schemes are different. The JSDM scheme is shown to be asymptotically optimal in terms of the achievable capacity, whereas the asymptotic optimality of the proposed overlayed training scheme is established from the channel estimation perspective. Finally, we remark that a coordination strategy can be used to make sure that users to be served in the same time-frequency slot are well separated in the AoA domain, similarly as discussed in [7], [15].

VI. Estimated Covariance-Assisted MMSE

The MMSE estimator assumes perfect knowledge of the downlink channel covariance matrix. This knowledge, however, is unavailable and needs to be estimated in practice. If the covariance matrix is estimated by the user, it needs to be fed back to the BS through some control channel, which involves a significant amount of overhead. One way to overcome this difficulty is to estimate the downlink channel covariance matrix from the uplink covariance matrix, e.g. [29], [30]. This approach, however, still requires a certain amount of specific uplink training. In this section, we develop a simple scheme to estimate the channel covariance matrix based on the one ring model. A MMSE estimator is then constructed based on the estimated covariance matrix. Our simulation results indicate that the covariance estimation scheme is effective and can obtain notable improvement in estimation performance.

According to the one-ring model (3), the covariance matrix of $\boldsymbol{h}$ can be written as

$$\boldsymbol{R} = \frac{\xi^2}{P}\sum_{i=1}^{P} E[\boldsymbol{a}(\theta_p)\boldsymbol{a}(\theta_p)^H] = \xi^2 E[\boldsymbol{a}(\theta)\boldsymbol{a}(\theta)^H] \quad (43)$$

To calculate $E[\boldsymbol{a}(\theta)\boldsymbol{a}(\theta)^H]$, we need to know the distribution of θ. Here we assume θ is uniformly distributed with mean angle $\bar{\theta}$ and angular spread ν. Thus the (m,n)th entry of $\boldsymbol{R}$ can be expressed as

$$R_{mn} = \frac{\xi^2}{2\nu}\int_{\bar{\theta}-\nu}^{\bar{\theta}+\nu} e^{-j2\pi\frac{(m-n)d}{\chi}\cos(\theta)}d\theta \quad (44)$$

The above integration, however, is difficult to calculate. Noting that the angular spread ν is usually small, we can use the Taylor expansion of $\cos(\theta)$ to approximate the integral. We have

$$\cos(\theta) \approx \cos(\bar{\theta}) - \sin(\bar{\theta})(\theta - \bar{\theta}) \quad (45)$$

Substituting (45) into (44), we arrive at

$$\begin{aligned} R_{mn} &\approx \frac{\xi^2}{2\nu}e^{jA_{mn}\cos(\bar{\theta})}\int_{-\nu}^{\nu} e^{-jA_{mn}\sin(\bar{\theta})\theta}d\theta \\ &= \xi^2 e^{jA_{mn}\cos(\bar{\theta})}\mathrm{sinc}(A_{mn}\sin(\bar{\theta})\nu) \end{aligned} \quad (46)$$

where $A_{mn} \triangleq 2\pi(m-n)d/\chi$, and $\mathrm{sinc}(x) \triangleq \sin(x)/x$ is the sinc function. Therefore, the covariance matrix $\boldsymbol{R}$ can be approximated as a parametric matrix with parameters $\bar{\theta}$ and ν. Note that the parameter ξ^2 in (46) can be ignored since as a scaling factor, it is independent of the signal subspace of $\boldsymbol{R}$. Thus the channel covariance estimation problem is simplified to find the mean angle $\bar{\theta}$ and the angular spread ν. There are several ways to estimate these two parameters. Here we introduce a compressed sensing-based method. Recalling that the channel with a narrow angular spread has an approximate sparse representation on the angular domain, i.e.

$$\boldsymbol{h} = \boldsymbol{A}\tilde{\boldsymbol{h}} \quad (47)$$

where $\boldsymbol{A}$ is an $M \times M$ unitary matrix determined by the array geometry at the base station. For the uniform linear array, $\boldsymbol{A}$ becomes the DFT matrix consisting of columns characterized by different angular coordinates. $\tilde{\boldsymbol{h}}$ is an approximately sparse vector, of which the mth element is contributed by the paths around the mth angular coordinate. Due to the narrow angular spread, a majority of the channel energy is concentrated on a few consecutive angular coordinates. Hence the mean angle and angular spread can be coarsely estimated from the sparse signal $\tilde{\boldsymbol{h}}$. More precisely, the angular coordinate which has the largest magnitude can be estimated as the mean angle, i.e.

$$\hat{\bar{\theta}} = \begin{cases} \arccos\left[\frac{\chi}{d}\left(\frac{s-1}{M}\right)\right] & \text{if } s \le \frac{M}{2}+1 \\ \arccos\left[\frac{\chi}{d}\left(\frac{s-1}{M}-1\right)\right] & \text{otherwise} \end{cases} \quad (48)$$

where s is the index of the angular coordinate which has the largest magnitude, i.e. the sth element of $\tilde{\boldsymbol{h}}$ has the largest magnitude. The angular spread can be estimated as a symmetric interval around the mean angle, say, $[\hat{\bar{\theta}}-\hat{\nu}, \hat{\bar{\theta}}+\hat{\nu}]$, with a majority of the channel energy (say, 90%) included in this interval. Now it remains to estimate the sparse vector $\tilde{\boldsymbol{h}}$.

Algorithm 1 Estimated Covariance-Assisted MMSE (EC-MMSE)

Given the received signal $\boldsymbol{y} \in \mathbb{C}^T$ and the pilot signal $\boldsymbol{X} \in \mathbb{C}^{T\times M}$.

1. Recover $\tilde{\boldsymbol{h}}$ from $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{A}\tilde{\boldsymbol{h}} + \boldsymbol{w}$ via compressed sensing techniques, where $\boldsymbol{A}$ is a DFT matrix for the uniform linear array case.
2. Estimate the mean angle $\hat{\bar{\theta}}$ and angular spread $\hat{\nu}$ based on $\tilde{\boldsymbol{h}}$, then obtain an estimate of the channel covariance matrix, $\hat{\boldsymbol{R}}$, via (46).
3. Construct a MMSE estimator $\hat{\boldsymbol{h}} = \hat{\boldsymbol{R}}\boldsymbol{X}^H(\boldsymbol{X}\hat{\boldsymbol{R}}\boldsymbol{X}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y}$ to estimate the channel $\hat{\boldsymbol{h}}$.

As indicated earlier in this paper, the estimation of $\tilde{\boldsymbol{h}}$ can be formulated into a sparse signal recovery problem:

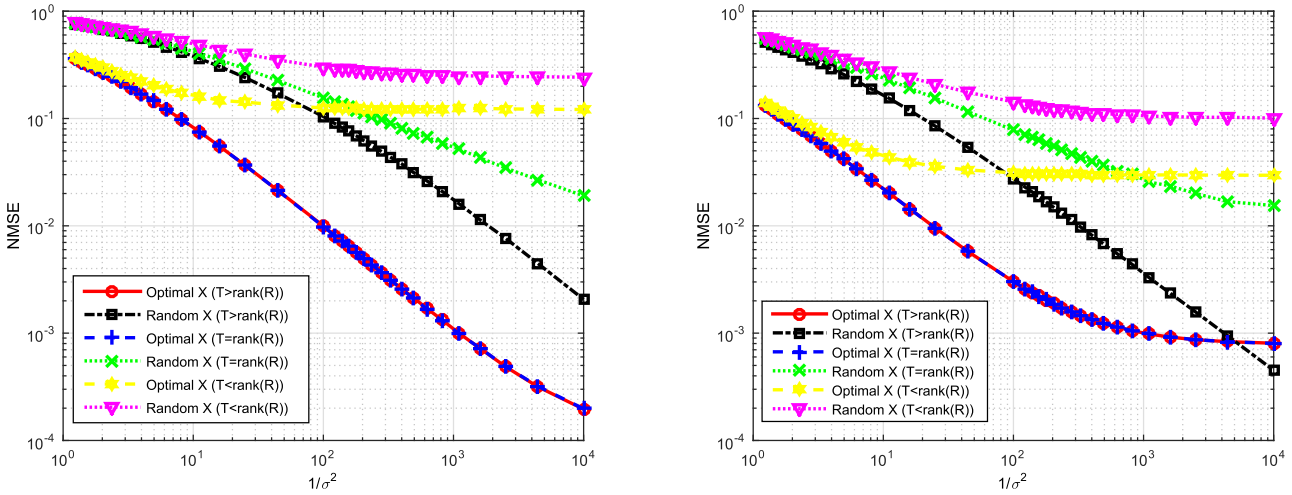
$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{w} = \boldsymbol{X}\boldsymbol{A}\tilde{\boldsymbol{h}} + \boldsymbol{w} \quad (49)$$

and can be efficiently solved via greedy or convex optimization methods. After $\tilde{\boldsymbol{h}}$ is recovered, the mean angle and the angular spread can be obtained by using the aforementioned procedure, and an estimate of the channel covariance matrix can be computed by substituting the estimated mean angle and angular spread into (46). Finally, a MMSE estimate of $\boldsymbol{h}$ can be obtained.

For clarity, we summarize our proposed estimated covariance-assisted MMSE scheme in Algorithm 1.

Remark 1: Although $\boldsymbol{h}$ can be directly estimated from (49) via compressed sensing techniques, the MMSE estimator with the help of the estimated channel covariance matrix can provide a better estimation accuracy, as demonstrated by our simulation results. Our proposed MMSE estimator can be employed either at the mobile station (i.e. user) or at the BS to estimate the channel. If the channel is estimated by the mobile station, the full CSI needs to be fed back to the BS, which causes a large amount of uplink overhead when M is large. An alternative approach is to let the mobile station simply feed back the received signal $\boldsymbol{y}$ to the BS, and let the BS form an estimate of the channel based on $\boldsymbol{y}$. This approach requires less uplink overhead since the dimension of $\boldsymbol{y}$ is usually smaller than the dimension of the channel $\boldsymbol{h}$. It should be noted for our proposed method, the received pilot signal $\boldsymbol{y}$ is used for both the channel covariance matrix and the channel estimation. Thus no additional overhead is required.

Remark 2: Our scheme assumes a uniform AoA distribution when estimating the channel covariance matrix. In practice, the AoA may not strictly follow a uniform distribution. Nevertheless, note that the eigenvectors of the channel covariance matrix are more closely related to the location of the interval over which the AoA is distributed, but less dependent on the specific distribution of the AoA. Hence our estimation scheme which assumes a uniform AoA distribution can still reliably estimate the true dominant eigenvectors when there is a mismatch between the presumed AoA distribution and the true distribution. As a result, the proposed MMSE estimator still deliver superior performance, as verified by our simulation results.



(a) Exact low-rank covariance matrix

(b) Approximate low-rank covariance matrix

Fig. 2. NMSE versus $1/\sigma^2$ for different choices of T.

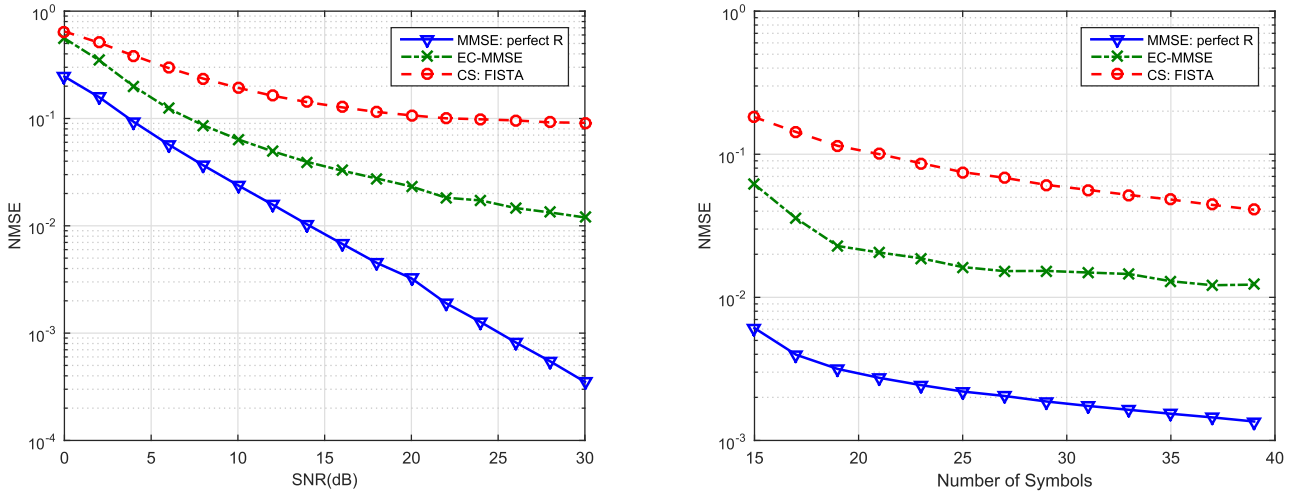
Remark 3: The estimation of the channel covariance matrix based on the one-ring model was also considered in [18]. Specifically, the work [18] suggested to estimate the channel covariance matrix as $\hat{\boldsymbol{R}} = \boldsymbol{F}\boldsymbol{D}\boldsymbol{F}^H$, where $\boldsymbol{F}$ is a DFT matrix and $\boldsymbol{D}$ is a diagonal matrix that contains the angular power spectral values. Our simulation results, however, show that, for a finite number of antennas, this covariance estimation approximation is not accurate enough and a MMSE estimator based on this covariance approximation even leads to deteriorated estimation performance. In addition, as indicated in [18], the estimation of the angular power spectrum requires additional training overhead and computational cost.

VII. Simulation Results

We now carry out experiments to validate our theoretical results and to illustrate the performance of the estimated covariance-assisted MMSE estimator (referred to as EC-MMSE) proposed in Section VI. Throughout our simulations, unless otherwise explicitly specified, we assume a uniform linear array with $M = 64$ antennas, and the distance between neighboring antenna elements is set to a half of the wavelength of the signal.

We first examine the behavior of the MMSE estimator in the asymptotic low-noise regime when the channel covariance matrix has a low-rank or an approximate low-rank structure. The channel covariance matrix is assumed perfectly known by the MMSE estimator. Fig. 2 depicts the normalized mean-squared errors (NMSEs) of the MMSE estimator vs. the reciprocal of the noise variance, where we consider both the optimal pilot sequence devised according to Theorem 2 and a random pilot sequence whose entries are i.i.d. normal random variables. Note that the random pilot sequence has to be multiplied by a scaling factor to satisfy a power constraint $\mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) \leq P$ that is also imposed on the optimal pilot. In Fig. 2(a), we randomly generate an exact low-rank channel covariance matrix $\boldsymbol{R}$ whose rank is set equal to 15. While for Fig. 2(b), the channel covariance matrix is generated according to the one-ring model, where the AoAs are assumed to be uniformly distributed over an interval $[\bar{\theta} - \nu, \bar{\theta} + \nu]$, with the mean angle and the angular spread given respectively by $\bar{\theta} = \pi/6$ and $\nu = \pi/10$, the total number of i.i.d. paths is set to $P = 100$, and α_p follows a complex Gaussian distribution with zero mean and variance $\xi^2 = 1$. A numerical average is utilized to compute (43) and obtain the channel covariance matrix for the one-ring model. Numerical results show that the covariance matrix has an approximate low-rank structure with about 12 dominant eigenvalues. To examine the impact of the number of pilot symbols on the estimation performance, we consider three different choices of T in our simulations, namely, $T = 20 > \mathrm{rank}(\boldsymbol{R})$, $T = \mathrm{rank}(\boldsymbol{R})$, and $T = 10 < \mathrm{rank}(\boldsymbol{R})$. From Fig. 2, we observe that when the number of symbols T is no less than the rank of the channel covariance matrix, the NMSE of the MMSE estimator approaches zero in the limit of vanishing noise, i.e. $\sigma^2 \to 0$, whatever an optimal pilot sequence or a random pilot sequence is employed. On the other hand, when $T < \mathrm{rank}(\boldsymbol{R})$, there exists an error floor for both the optimal and random pilots, that is, once the error floor is reached, a decrease in the noise power does not bring any additional estimation performance improvement. This result corroborates our theoretical analysis in Section III. Given a power constraint, the optimal pilot sequences for $T > \mathrm{rank}(\boldsymbol{R})$ and $T = \mathrm{rank}(\boldsymbol{R})$ are identical. Thus the NMSEs achieved by optimal pilot sequences remain unaltered for these two cases. We also observe that for the approximate low-rank case, it seems that an error floor exists even for an optimal pilot design with $T \geq \mathrm{rank}(\boldsymbol{R})$. This is because we only retain the largest 12 eigenvalues of the approximate low-rank channel covariance matrix and ignore the rest small eigenvalues when devising the optimal pilot design.

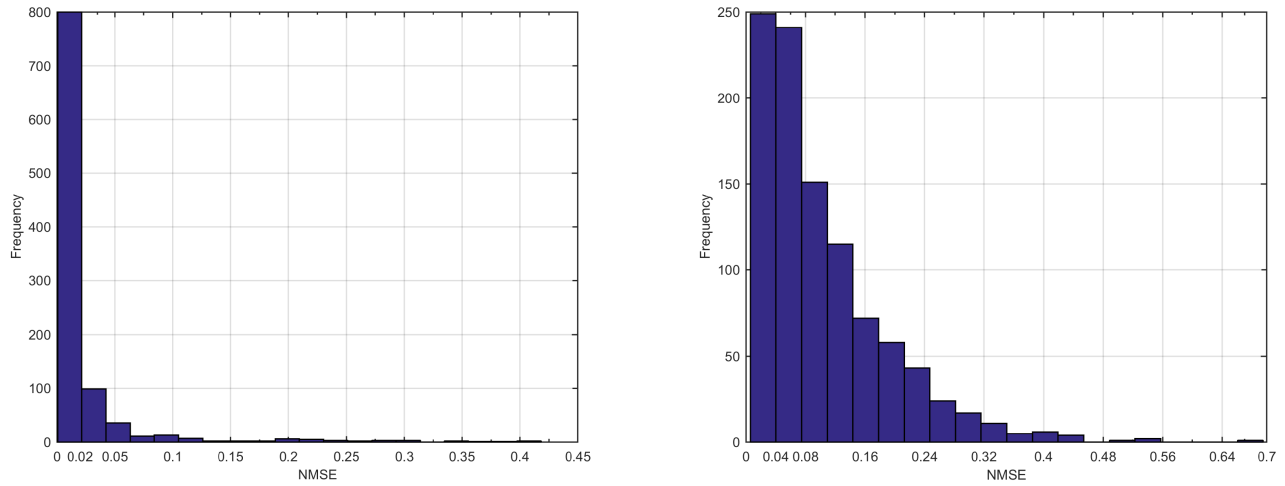
Next, we evaluate the performance of the EC-MMSE estimator proposed in Section VI. In our simulations, channels are randomly generated according to the one-ring model described above, where we set $\bar{\theta} = \pi/6$ and $\nu = 5°$. Fig. 3(a) depicts



(a) NMSE vs. SNR

(b) NMSE vs. number of symbols T

Fig. 3. NMSEs of respective schemes vs. SNR and number of symbols T.



(a) EC-MMSE

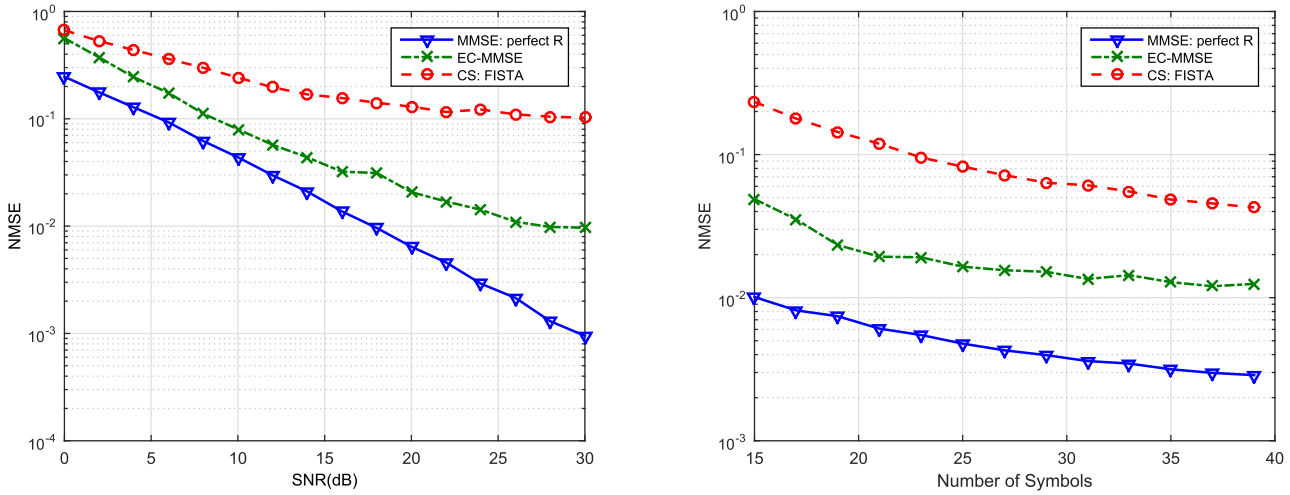
(b) CS:FISTA

Fig. 4. Histogram of the NMSE associated with the EC-MMSE estimator and the compressed sensing method.

the NMSEs of respective methods as a function of the signal-to-noise ratio (SNR), where we set $T = 20$ and the SNR is defined as $10\log(\|\boldsymbol{X}\boldsymbol{h}\|_2^2/T\sigma^2)$. Results are averaged over 1000 independent runs, with the pilot sequence $\boldsymbol{X}$ and the channel $\boldsymbol{h}$ randomly generated for each run. In each run, the noise variance σ^2 is adjusted to meet a pre-specified SNR. A compressed sensing method and a MMSE estimator which has access to the true covariance matrix[2] are also included for comparison. For the compressed sensing method, a fast iterative shrinkage-thresholding algorithm (FISTA) [31] is employed to estimate the channel based on (49). The EC-MMSE estimator is built on the compressed sensing method: after the virtual channel $\tilde{\boldsymbol{h}}$ is estimated via the FISTA, we estimate the mean angle and the angular spread, then obtain an estimate of the channel covariance matrix, and finally construct the MMSE estimator. In our simulations, the angular spread is estimated as a symmetric interval around the estimated mean angle, with 95% of the channel energy concentrated on the interval. From Fig. 3(a), we see that our proposed scheme achieves a notably higher accuracy compared to the compressed sensing method. This result shows that the estimated covariance matrix, although imperfect, can still provide a substantial performance improvement. Fig. 3(b) plots the NMSEs of respective schemes vs. the number of symbols T, where we set SNR $=$ 20dB. This result again demonstrates the advantage of the proposed EC-MMSE estimator over the compressed sensing method. To better illustrate the performance, we plot the histogram in Fig. 4 to show the distribution of the NMSE for the EC-MMSE and the compressed sensing method, respectively. From Fig. 4, we see that the proposed EC-MMSE yields an accurate channel estimate (with an NMSE within the range $[0, 0.02]$) with a high probability, whereas the NMSEs associated with the compressed sensing method spread across the range $[0.04, 0.4]$ with a high probability.

Also, to examine the robustness of the proposed EC-MMSE estimator against the model mismatch, in our simulations, we assume that the angle of arrival follows

[2]The true covariance matrix is calculated according to (43) via numerical average.



(a) NMSE vs. SNR

(b) NMSE vs. number of symbols T

Fig. 5. Gaussian AoA: NMSEs of respective schemes vs. SNR and number of symbols T.

a Gaussian distribution, with the mean and the standard deviation set to be $\bar{\theta} = \pi/6$ and $\sigma_\theta = \pi/30$, respectively. Note that in the proposed EC-MMSE estimator, a uniform AoA distribution is assumed to estimate the channel covariance matrix. In Fig. 5(a) and Fig. 5(b), we plot the NMSEs of respective schemes as a function of the SNR and the number of symbols, respectively, where we set $T = 20$ for Fig. 5(a) and SNR = 20dB for Fig. 5(b). Results are averaged over 1000 independent runs, with the pilot sequence and the channel randomly generated for each run. In each run, the noise variance is adjusted to meet a pre-defined SNR. From Fig. 5, we see that the proposed EC-MMSE estimator achieves superior performance even the presumed AoA distribution is different from the true one. The reason, as already explained in the previous section, is that the eigenvectors of the channel covariance matrix are less dependent on the AoA distribution. Therefore our scheme which assumes a uniform AoA distribution can still reliably estimate the signal subspace spanned by dominant eigenvectors, and as a result, the EC-MMSE estimator still outperforms the compressed sensing method by a big margin.

Next, we examine the robustness of our proposed scheme against the one-ring model mismatch. It is known that the one-ring model is characterized by a small angular spread (typically around 5°–10°). Thus a wider angular spread implies that the channel deviates more from the one-ring model. Fig. 6 plots the NMSEs of respective schemes as a function of the angular spread ν, where we set $T = 30$ and SNR = 20dB. Channels are randomly generated according to the one-ring model, with $\bar{\theta} = \pi/6$ and the angular spread ν varying from 5° to 25°. From Fig. 6, we can see that the performance gain achieved by the EC-MMSE estimator shrinks as the angular spread becomes wider and wider. When the angular spread ν increases up to 20°–25°, our proposed scheme achieves performance similar to that of the compressed sensing technique. This is due to the fact that for our scheme, the covariance matrix approximation based on the Taylor expansion is no longer accurate when the angular spread is large, and as a consequence, the use of inaccurate channel statistical

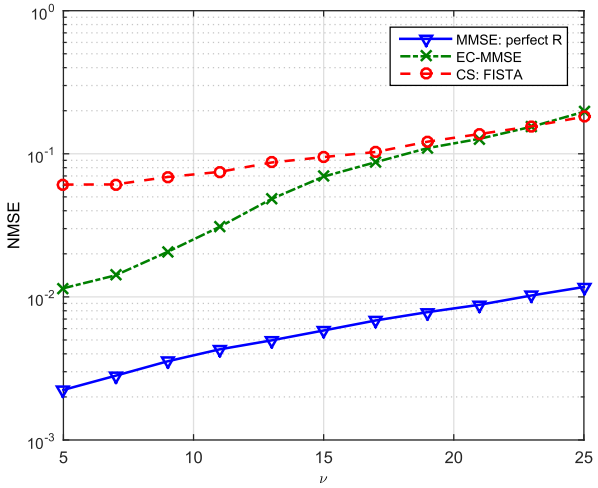
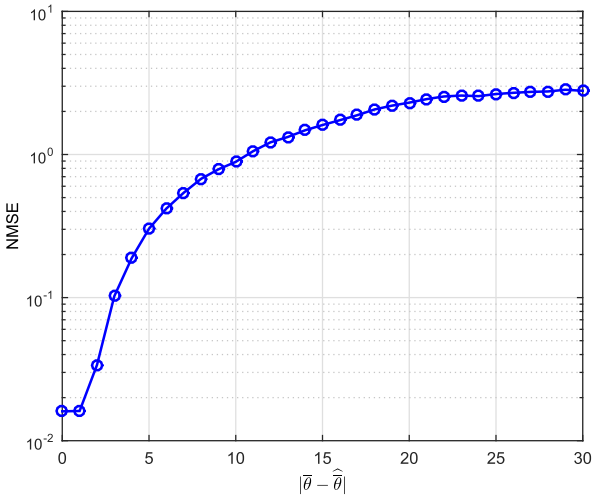


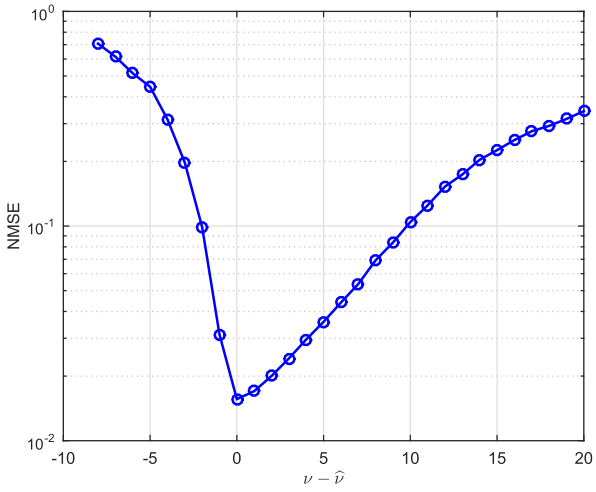
Fig. 6. NMSEs of respective algorithms vs. the angular spread ν (in angular degrees).

information leads to no performance improvement. Nevertheless, our proposed EC-MMSE scheme achieves a performance improvement over the compressed sensing technique for a moderately large angular spread, say, $\nu \in [10°, 20°]$. Thus our proposed estimator exhibits some robustness against the one-ring model mismatch.

Lastly, to more thoroughly evaluate the performance of the proposed EC-MMSE estimator, we examine its robustness against estimation errors of the mean angle and the angular spread. Since in the EC-MMSE scheme, the channel covariance matrix is obtained based on the estimated mean angle and angular spread, estimation errors of the mean angle and the angular spread will impair the estimation quality of the covariance matrix, which, in turn, affects the estimation accuracy of the EC-MMSE estimator. In Fig. 7(a), we plot the NMSE of the EC-MMSE estimator as the estimated mean angle deviates from the true one, where we set $T = 15$, SNR = 20dB, and the angular spread is assumed perfectly known. Results are averaged over 10^3 independent runs, and for each run, the pilot sequence is randomly generated to meet a pre-specified power constraint, and the channel is randomly



(a) NMSE vs. the deviation (in angular degrees) of the estimated mean angle from the true one



(b) NMSE vs. the deviation (in angular degrees) of the estimated angle spread from the true one

Fig. 7. NMSE of EC-MMSE estimator vs. estimation errors of the mean angle and the angular spread.

generated according to the one-ring model, with $\bar{\theta} = \pi/6$ and $\nu = 10°$. We see that the EC-MMSE estimator exhibits some robustness against the mean angle mismatch: the EC-MMSE estimator incurs mild performance degradation if the deviation of the estimated mean angle from the true one is small, say, $|\bar{\theta} - \hat{\bar{\theta}}| < 3°$. Nevertheless, a large deviation would result in a significant performance degradation. Fig. 7(b) depicts the behavior of the proposed EC-MMSE estimator when the estimated angular spread deviates from the true angular spread, where the mean angle is assumed perfectly estimated. From Fig. 7(b), it can be observed that the EC-MMSE estimator is robust to an overestimation of the angular spread, but is sensitive to the underestimation errors: it suffers from a substantial performance loss when the estimated angular spread is smaller than the true one. Hence it is safer to overestimate than to underestimate the angular spread.

VIII. Conclusions

We considered the problem of downlink training and channel estimation for FDD massive MIMO systems. Since the required amount of overhead for downlink training grows linearly with the number of transmit antennas at the BS, reducing the overhead for downlink training and uplink feedback has been a central issue in FDD massive MIMO systems. In this paper, we exploited the low-rank structure of the channel covariance matrix to reduce the overhead for downlink training. We studied the asymptotic behavior of the MMSE estimator when the channel covariance matrix has a low-rank structure. Our analysis shows that the MMSE estimator can achieve an exact channel recovery in the asymptotic low-noise regime, provided that the number of pilot symbols in time is no smaller than the rank of the channel covariance matrix. We also examined the optimal pilot sequence design for the single-user case, and an asymptotic optimal pilot sequence design for the multi-user scenario. We also develop a training-free scheme to estimate the channel covariance matrix. Simulation results show that a MMSE estimator based on the estimated covariance matrix achieves a substantial performance improvement as compared with the compressed sensing method, and is robust against the AoA distribution mismatch and the angular spread estimation error.

References

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

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

[3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[4] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[5] D. Ciuonzo, P. S. Rossi, and S. Dey, "Massive MIMO channel-aware decision fusion," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 604–619, Feb. 2015.

[6] A. Shirazinia, S. Dey, D. Ciuonzo, and P. Salvo Rossi, "Massive MIMO for decentralized estimation of a correlated source," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2499–2512, May 2016.

[7] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 2, pp. 264–273, Feb. 2013.

[8] R. R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[9] J.-C. Guey and L. D. Larsson, "Modeling and evaluation of MIMO systems exploiting channel reciprocity in TDD mode," in *Proc. IEEE 60th Veh. Technol. Conf. (VTC2004-Fall)*, Los Angeles, CA, USA, Sep. 2004, pp. 4265–4269.

[10] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with Downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[11] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, Jul. 2012.

[12] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[13] 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.

[14] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, to be published.

[15] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
[16] C. Sun, X. Q. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.
[17] 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.
[18] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.
[19] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.
[20] W. Shen, L. Dai, B. Shim, S. Mumtaz, and Z. Wang, "Joint CSIT acquisition based on low rank matrix completion for FDD massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2178–2181, Dec. 2015.
[21] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via virtual channel representation," Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep., 2007.
[22] C. Shepard *et al.*, "Argos: Practical many-antenna base stations," in *Proc. MobiCom*, Istanbul, Turkey, Aug. 2012, pp. 53–64.
[23] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.
[24] F. Renna, R. Calderbank, L. Carin, and M. R. D. Rodrigues, "Reconstruction of signals drawn from a Gaussian mixture via noisy compressive measurements," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2265–2277, May 2014.
[25] E. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[26] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.
[27] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.
[28] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.
[29] Y.-C. Liang and F. P. S. Chin, "Downlink channel covariance matrix estimation and its applications in wireless DS-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222–232, Feb. 2001.
[30] B. M. Hochwald and T. L. Maretta, "Adapting a downlink array from uplink measurements," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 642–653, Mar. 2001.
[31] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

Jun Fang (M'08) received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 1998 and 2001, respectively, and the Ph.D. degree from the National University of Singapore, Singapore, in 2006, all in electrical engineering.

During 2006, he was a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University. From 2007 to 2010, he was a Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology. Since 2011, he has been with the University of Electronic of Science and Technology of China. His current research interests include sparse theory and compressed sensing, statistical learning, millimeter Wave, and massive MIMO communications.

Dr. Fang received the IEEE Jack Neubauer Memorial Award in 2013 for the best systems paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is an Associate Technical Editor for the *IEEE Communications Magazine*, and an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.

Xingjian Li received the B.Sc. degree from the University of Electronic Science and Technology in 2015. He is currently pursuing the master's degree with the University of Electronic of Science and Technology of China. His current research interests include compressed sensing, millimeter Wave, and massive MIMO communications.

Hongbin Li (M'99–SM'08) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 1999, all in electrical engineering.

From 1996 to 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Florida. Since 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, where he is a Professor. He was a Summer Visiting Faculty Member with the Air Force Research Laboratory, in 2003, 2004, and 2009. His general research interests include statistical signal processing, wireless communications, and radars.

Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He has been a member of the IEEE SPS Signal Processing Theory and Methods Technical Committee, since 2011, and was a member of the IEEE SPS Sensor Array and Multichannel Technical Committee from 2006 to 2012. He received the IEEE Jack Neubauer Memorial Award in 2013 for the best systems paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the Outstanding Paper Award at the IEEE AFICON Conference in 2011, the Harvey N. Davis Teaching Award in 2003 and the Jess H. Davis Memorial Award for excellence in research in 2001 from the Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He has been an Associate Editor for *Signal Processing* (Elsevier) since 2013, the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2006 to 2009 and since 2014), the IEEE SIGNAL PROCESSING LETTERS (2005 to 2006), and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2003 to 2006), as well as a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and *EURASIP Journal on Applied Signal Processing*. He has been involved in various conference organization activities, including serving as a General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop, held in Hoboken, NJ, USA, in 2012.

Feifei Gao (M'09–SM'14) received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China, in 2002, the M.Sc. degree from McMaster University, Hamilton, ON, Canada, in 2004, and the Ph.D. degree from the National University of Singapore, Singapore, in 2007.

He was a Research Fellow with the Institute for Infocomm Research, A*STAR, Singapore, in 2008, and an Assistant Professor with the School of Engineering and Science, Jacobs University, Bremen, Germany, from 2009 to 2010. In 2011, he joined the Department of Automation, Tsinghua University, Beijing, China, where he is currently an Associate Professor. His research areas include communication theory, signal processing for communications, array signal processing, and convex optimizations, with particular interests in MIMO techniques, multi-carrier communications, cooperative communication, and cognitive radio networks. He has authored/co-authored over 90 refereed IEEE journal papers and over 120 IEEE conference proceeding papers, which have been cited over 3500 times according to Google Scholar.

Professor Gao has served as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE COMMUNICATIONS LETTERS, the IEEE SIGNAL PROCESSING LETTERS, the IEEE WIRELESS COMMUNICATIONS LETTERS, *International Journal on Antennas and Propagations*, and *China Communications*. He also served as the Symposium Co-Chair at the 2015 IEEE Conference on Communications, the 2014 IEEE Global Communications Conference, the 2014 IEEE Vehicular Technology Conference Fall, and as a technical committee member for many other IEEE conferences.