PILOT DESIGN FOR GAUSSIAN MIXTURE CHANNEL ESTIMATION IN MASSIVE MIMO

Yujie Gu and Yimin D. Zhang

Department of Electrical & Computer Engineering, Temple University, Philadelphia, PA 19122, USA

ABSTRACT

Massive multiple-input multiple-output (MIMO) is a promising technique for 5G communications due to its superior spectrum and energy efficiencies. Despite its many advantages, the high number of antennas used in massive MIMO brings many challenges in practical implementations. Among them, the pilot overhead for downlink channel estimation becomes unaffordable in frequency division duplex (FDD) massive MIMO systems. In this paper, we exploit the available *a priori* knowledge of the channel to optimize the pilot design. By utilizing the low-rank nature of the channel matrix, we first derive the minimum number of pilot symbols required for perfect channel recovery. Further, under the general Gaussian mixture model for the channel vector, the pilot symbols are optimized to maximize the mutual information between the measurements of the user and the corresponding channel vector. Simulation results demonstrate the effectiveness of the proposed optimal pilot design for the downlink channel estimation in FDD massive MIMO systems.

***Index Terms*—** Channel estimation, Gaussian mixture, massive MIMO, pilot optimization.

1. INTRODUCTION

With the development of millimeter wave technology, massive multiple-input multiple-output (MIMO) has been widely expected to be the key enabling technology in next generation wireless communications standards. The significant benefits of massive MIMO in system capacity, spectrum efficiency, energy efficiency, security, and robustness stem from the high number of antennas exploited at the base station [1–7]. This fact, on the other hand, also brings a number of challenges, such as the pilot design for downlink channel estimation in frequency division duplex (FDD) massive MIMO.

In wireless communications, the channel state information is essential for reliable data transmission and efficient resource allocation. In general, the number of pilot symbols in MIMO communications should not be less than the number of antennas at the base station in order to effectively identify the channel states. This results in a huge pilot overhead for the downlink channel estimation in FDD massive MIMO, thereby reducing the spectrum efficiency. In the FDD operation, a user estimates its downlink channel from the received pilot symbols, and feeds the estimated channel state information back to the base station. Hence, it is a critical task to identify the downlink channel in massive MIMO with a significantly reduced number of pilot symbols.

In recent years, reducing the pilot overhead in massive MIMO systems has been the subject of extensive studies. Different design criteria have been developed, such as maximizing the system spectral efficiency [8], maximizing the average received signal-to-noise ratio (SNR) [9], maximizing the summation of the conditional mutual information [10], maximizing the sum-rate upper bound [11], minimizing the mean squared error (MSE) [9, 12], minimizing the sum MSE [12], and minimizing the weighted sum MSE [13]. These techniques are developed by exploiting the sparsity of the channel vector due to the narrow angular spread of the incoming/outgoing rays at the base station in typical cellular systems. All these techniques assume the channel vector to be modeled as a single smooth Gaussian variable. Such assumption is very strict and may suffer from performance loss when the actual channel deviates from the assumed model.

In this paper, we perform optimal pilot design by exploiting the low-rank feature of the downlink channels in FDD massive MIMO, and modeling the channels to follow a general Gaussian mixture distribution. First, we study the asymptotic behavior of the minimum mean-squared error (MMSE) estimator. It reveals that a perfect channel recovery can be achieved in the asymptotic regime, provided that the number of pilot symbols is not less than the maximum rank of the channel covariance matrices of all Gaussian components. Second, we adopt the mutual information maximization criterion to optimize the pilot symbols for the downlink channel estimation. The optimization problem is solved using a gradient-based search method based on the gradient of the approximated information with respect to the pilot matrix. Simulation results demonstrate the effectiveness of the proposed pilot design for the downlink channel estimation in massive MIMO systems.

2. SIGNAL MODEL IN MASSIVE MIMO

Assume that a massive MIMO base station uses $N \gg 1$ antennas to transmit a set of pilot symbols $\{\boldsymbol{\phi}(l) \in \mathbb{C}^N, l = 1, 2, \cdots, L\}$. The baseband received signal at the singleantenna user terminal is expressed as

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{h} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}(1), \boldsymbol{\phi}(2), \cdots, \boldsymbol{\phi}(L)]^{\mathrm{T}} \in \mathbb{C}^{L\times N}$ is the L-symbol pilot matrix, $\boldsymbol{h} \in \mathbb{C}^{N}$ is the frequency-flat fading downlink channel vector, and $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I})$ is the zero-mean additive Gaussian white noise with variance σ_n^2. Here, $(\cdot)^{\mathrm{T}}$ denotes the transpose operation, and $\boldsymbol{I}$ represents an identity matrix.

The least squares (LS) estimate of $\boldsymbol{h}$ is given by

$$\hat{\boldsymbol{h}}^{\mathrm{LS}} = \left[\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\right]^{-1} \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{y}, \tag{2}$$

where $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transpose. In order to perform the above matrix inversion, the number of pilot symbols must not be less than the number of antennas, i.e., $L \geq N$. In massive MIMO systems, as the number of antennas at the base station, N, is typically very large, such a pilot overhead becomes unaffordable.

In order to guarantee the spectrum efficiency, therefore, the number of pilot symbols in FDD massive MIMO systems must be much less than the number of antennas at the base station, i.e., $L \ll N$. In such a case, the LS solution is no longer applicable. Hence, we exploit the *a prior* knowledge of the channel to significantly reduce the number of pilot symbols required for channel recovery.

3. PILOT OPTIMIZATION IN MASSIVE MIMO

3.1. Gaussian mixture channel model

In a massive MIMO system, the channels are usually sparse and low-rank due to the small number of users and the narrow angular spread of the beam. The latter leads to a high correlation between different paths that link the base station and the user [8–10]. We further model the channel vector to follow a Gaussian mixture distribution. This distribution model is well verified in practice to describe the real environment with a high flexibility and tractability (see, e.g., [14–16] and the references therein).

Let the probability density function (pdf) of the channel vector $\boldsymbol{h}$ be modeled by a Gaussian mixture distribution as

$$f(\boldsymbol{h}) = \sum_{k\in\mathcal{K}} p_k f^{(k)}(\boldsymbol{h}), \tag{3}$$

which implies that it contains $K = |\mathcal{K}|$ Gaussian components, and the k-th component is activated with probability $p_k > 0$ ($\sum_{k\in\mathcal{K}} p_k = 1$) and, when activated, that component generates a complex-valued Gaussian vector with distribution $f^{(k)}(\boldsymbol{h}) = \mathcal{CN}\left(\boldsymbol{u}_{\boldsymbol{h}}^{(k)}, \boldsymbol{R}_{\boldsymbol{hh}}^{(k)}\right)$.

Under the channel distribution model (3), the user measurement $\boldsymbol{y}$ also follows a Gaussian mixture distribution as

$$f(\boldsymbol{y}) = \sum_{k\in\mathcal{K}} p_k \mathcal{CN}\left(\boldsymbol{u}_{\boldsymbol{y}}^{(k)}, \boldsymbol{R}_{\boldsymbol{yy}}^{(k)}\right), \tag{4}$$

where the mean vector and covariance matrix of the k-th component are given by

$$\begin{aligned} \boldsymbol{u}_{\boldsymbol{y}}^{(k)} &= \boldsymbol{\Phi}\boldsymbol{u}_{\boldsymbol{h}}^{(k)}, \\ \boldsymbol{R}_{\boldsymbol{yy}}^{(k)} &= \boldsymbol{\Phi}\boldsymbol{R}_{\boldsymbol{hh}}^{(k)}\boldsymbol{\Phi}^{\mathrm{H}} + \sigma_n^2\boldsymbol{I}. \end{aligned} \tag{5}$$

The MMSE estimate of the channel vector $\boldsymbol{h}$, defined as

$$\min_{\hat{\boldsymbol{h}}} E\left\{\left\|\boldsymbol{h} - \hat{\boldsymbol{h}}\right\|_2^2\right\}, \tag{6}$$

is given by [17]

$$\hat{\boldsymbol{h}}^{\mathrm{MMSE}} = E\{\boldsymbol{h}|\boldsymbol{y}\} = \sum_{k\in\mathcal{K}} p_{k|\boldsymbol{y}} \boldsymbol{u}_{\boldsymbol{h}|\boldsymbol{y}}^{(k)}, \tag{7}$$

where $E\{\cdot\}$ denotes the statistical expectation,

$$\boldsymbol{u}_{\boldsymbol{h}|\boldsymbol{y}}^{(k)} = \boldsymbol{u}_{\boldsymbol{h}}^{(k)} + \boldsymbol{R}_{\boldsymbol{hh}}^{(k)}\boldsymbol{\Phi}^{\mathrm{H}}\left[\boldsymbol{R}_{\boldsymbol{yy}}^{(k)}\right]^{-1}\left(\boldsymbol{y} - \boldsymbol{u}_{\boldsymbol{y}}^{(k)}\right) \tag{8}$$

is the k-th component of the MMSE estimator of $\boldsymbol{h}$ given the measurement $\boldsymbol{y}$, and

$$p_{k|\boldsymbol{y}} = \frac{p_k f^{(k)}(\boldsymbol{y})}{f(\boldsymbol{y})} \tag{9}$$

is the corresponding posterior probability [18]. When the downlink channel is estimated, it will be sent back to the base station.

3.2. Asymptotic behavior of the MMSE estimator

Similar to the asymptotic analysis in [12], in this section, we analyze the behavior of the MMSE estimator in the high-SNR asymptotic regime. The asymptotic analysis verifies the possibility of perfect channel recovery from a small number of pilot symbols, because the channel matrix in massive MIMO systems is low rank.

Theorem: Let $r^{(k)}$ be the rank of the channel covariance matrix of the k-th Gaussian component in the Gaussian mixture distribution, $\boldsymbol{R}_{\boldsymbol{hh}}^{(k)}$, i.e., $r^{(k)} = \mathrm{rank}(\boldsymbol{R}_{\boldsymbol{hh}}^{(k)})$, where $\mathrm{rank}(\cdot)$ denotes the rank of a matrix. Let $\boldsymbol{V}^{(k)}\boldsymbol{\Gamma}^{(k)}(\boldsymbol{V}^{(k)})^{\mathrm{H}}$ denote the eigen-decomposition of $(\boldsymbol{R}_{\boldsymbol{hh}}^{(k)})^{\frac{1}{2}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}(\boldsymbol{R}_{\boldsymbol{hh}}^{(k)})^{\frac{1}{2}}$, where $\boldsymbol{V}^{(k)} = [\boldsymbol{v}_1^{(k)}, \cdots, \boldsymbol{v}_N^{(k)}]$ is a unitary matrix consisting of the eigenvectors, and $\boldsymbol{\Gamma}^{(k)} = \mathrm{diag}(\gamma_1, \gamma_2, \cdots, \gamma_{r^{(k)}}, 0, \cdots, 0)$ is a diagonal matrix consisting of the eigenvalues with $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{r^{(k)}} > 0$. Assume that the number of randomly generated pilot symbols, L, is no less than the maximum of $r^{(k)}$ of all Gaussian components, i.e., $L \geq \max_{k\in\mathcal{K}} r^{(k)}$, then the *lower* bound of the MSE of the MMSE estimate of $\boldsymbol{h}$ is given by

$$\varepsilon_{\mathrm{Lower}} \leq \sum_{k\in\mathcal{K}} p_k \sum_{i=1}^{r^{(k)}} \left(1 + \frac{\gamma_i}{\sigma_n^2}\right)^{-1} \left(\boldsymbol{v}_i^{(k)}\right)^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{hh}}^{(k)} \boldsymbol{v}_i^{(k)}, \tag{10}$$

which approaches zero in the high-SNR asymptotic regime, i.e., when $\sigma_n^2 \rightarrow 0$. The upper and lower bounds approach each other with an increasing SNR, and they coincide as the SNR tends to infinity [17]. Therefore, perfect channel recovery is possible because

$$\lim_{\text{SNR}\rightarrow\infty} E\left\{\left\|\boldsymbol{h}-\hat{\boldsymbol{h}}^{\text{MMSE}}\right\|_2^2\right\}=0, \tag{11}$$

where $\text{SNR}=\|\boldsymbol{h}\|^2/\sigma_n^2$ denotes the SNR of the user's channel.

3.3. Optimal pilot design

According to [19], a channel estimated under the maximum mutual information criterion is equivalent to the one estimated under the MMSE criterion. Considering that there is no closed-form MSE expression for the Gaussian mixture variable [17], we alternate the mutual information maximization criterion to optimize the pilot symbols for the downlink channel estimation.

The pilot optimization problem is formulated to maximize the mutual information between the measurement vector $\boldsymbol{y}$ and the channel vector $\boldsymbol{h}$ [20], i.e.,

$$\begin{aligned} \max_{\boldsymbol{\Phi}} \quad & I(\boldsymbol{y};\boldsymbol{h})=h(\boldsymbol{y})-h(\boldsymbol{y}|\boldsymbol{h}) \\ \text{subject to} \quad & \boldsymbol{\Phi}\boldsymbol{\Phi}^{\text{H}}=\boldsymbol{I}, \end{aligned} \tag{12}$$

where $h(\boldsymbol{y})$ denotes the differential entropy of the measurement $\boldsymbol{y}$, and $h(\boldsymbol{y}|\boldsymbol{h})$ represents the conditional differential entropy of the measurement $\boldsymbol{y}$ given the channel $\boldsymbol{h}$. The orthonormal constraint $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\text{H}}=\boldsymbol{I}$ is introduced to avoid increasing the mutual information by simply scaling $\boldsymbol{\Phi}$ to be larger because scaling $\boldsymbol{\Phi}$ only affects the channel rather than the noise. Another common constraint is the total transmit power budget, i.e., $\text{Tr}\{\boldsymbol{\Phi}\boldsymbol{\Phi}^{\text{H}}\} \leq L$.

It is difficult to analytically derive the differential entropy even for a simple estimation problem, let alone the parameter estimation problem with the high dimensionality and non-Gaussianity. By performing the first-order Taylor series expansion of the logarithm of the Gaussian mixture distribution in the definition of the differential entropy of the measurement vector $\boldsymbol{y}$, we obtain the approximated differential entropy as [14]

$$h(\boldsymbol{y}) \approx -\log\left[\sum_{k\in\mathcal{K}} p_k f^{(k)}\left(\boldsymbol{y}_0\right)\right], \tag{13}$$

where $\boldsymbol{y}_0=E\{\boldsymbol{y}\}$ is the mean value of the measurement. Following the zero mean assumption of the channel vector [8, 10], we have $\boldsymbol{u}_{\boldsymbol{y}}^{(k)}=\boldsymbol{\Phi}\boldsymbol{u}_{\boldsymbol{h}}^{(k)}=\boldsymbol{0}$ for all individual Gaussian components of $\boldsymbol{y}$. In this case, it is natural to set the Taylor series expansion point to $\boldsymbol{y}_0=\boldsymbol{0}$, resulting in

$$f^{(k)}\left(\boldsymbol{y}_0\right)=\frac{1}{\pi^L\left|\boldsymbol{R}_{\boldsymbol{yy}}^{(k)}\right|}, \tag{14}$$

where $|\cdot|$ denotes the determinant of a matrix. We now have the approximated differential entropy of $\boldsymbol{y}$ as

$$h(\boldsymbol{y}) \approx -\log\left[\sum_{k\in\mathcal{K}} p_k\left|\boldsymbol{R}_{\boldsymbol{yy}}^{(k)}\right|^{-1}\right]+L\log\pi, \tag{15}$$

where the second term is a constant independent of the pilot matrix $\boldsymbol{\Phi}$.

Because the additive noise $\boldsymbol{n}$ is independent of the channel $\boldsymbol{h}$, the conditional differential entropy of $\boldsymbol{y}$ given $\boldsymbol{h}$ becomes

$$h(\boldsymbol{y}|\boldsymbol{h})=h(\boldsymbol{n})=L\log(e\pi\sigma_n^2), \tag{16}$$

which is independent of the pilot matrix $\boldsymbol{\Phi}$.

Note that the mutual information $I(\boldsymbol{y};\boldsymbol{h})$ is *invariant* when the pilot matrix $\boldsymbol{\Phi}$ undergoes a unitary rotation, i.e.,

$$\begin{aligned} I(\boldsymbol{y};\boldsymbol{h})|_{\boldsymbol{Q\Phi}} &= I(\boldsymbol{y};\boldsymbol{h})|_{\boldsymbol{\Phi}} \\ &\approx -\log\left[\sum_{k\in\mathcal{K}} p_k\left|\boldsymbol{R}_{\boldsymbol{yy}}^{(k)}\right|^{-1}\right]-L\log(e\sigma_n^2), \end{aligned} \tag{17}$$

where $\boldsymbol{Q}\in\mathbb{C}^{L\times L}$ is a unitary matrix satisfying $\boldsymbol{Q}^{\text{H}}\boldsymbol{Q}=\boldsymbol{Q}\boldsymbol{Q}^{\text{H}}=\boldsymbol{I}$.

Taking the gradient of the approximated mutual information in (17) with respect to the pilot matrix $\boldsymbol{\Phi}$, we have

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\boldsymbol{h}) \approx \frac{\sum\limits_{k\in\mathcal{K}} p_k\left|\boldsymbol{R}_{\boldsymbol{yy}}^{(k)}\right|^{-1}\left[\boldsymbol{R}_{\boldsymbol{yy}}^{(k)}\right]^{-1}\boldsymbol{\Phi}\boldsymbol{R}_{\boldsymbol{hh}}^{(k)}}{\sum\limits_{k\in\mathcal{K}} p_k\left|\boldsymbol{R}_{\boldsymbol{yy}}^{(k)}\right|^{-1}}, \tag{18}$$

which is then used in a gradient ascent search for the optimal pilot design according to

$$\hat{\boldsymbol{\Phi}}=\boldsymbol{\Phi}+\gamma\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\boldsymbol{h}), \tag{19}$$

where $\gamma>0$ is a small step size. The updated pilot matrix is a linear combination of the current pilot matrix and the approximated mutual information gradient with respect to the pilot matrix.

The orthonormal constraint in (12) can be enforced by seeking the closest row-orthonormal matrix to the updated pilot matrix $\hat{\boldsymbol{\Phi}}$, which is an orthogonal Procrustes problem [21]. The orthonormal pilot matrix closest to $\hat{\boldsymbol{\Phi}}$ is given by

$$\tilde{\boldsymbol{\Phi}}=\boldsymbol{D}\boldsymbol{\Sigma}^{'}\boldsymbol{G}^{\text{H}}, \tag{20}$$

where $\hat{\boldsymbol{\Phi}}=\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{G}^{\text{H}}$ is the singular value decomposition (SVD) of $\hat{\boldsymbol{\Phi}}$, and $\boldsymbol{\Sigma}^{'}$ is a modified $\boldsymbol{\Sigma}$ with all singular values replaced by one. Then, $\tilde{\boldsymbol{\Phi}}$ is used to substitute $\boldsymbol{\Phi}$ in the next iteration in calculating (18) and then updated using (19) to achieve an iterative search procedure until convergence. Experience shows that fast convergence is achieved benefited from the sparsity of the channel.

4. SIMULATION RESULTS

In the simulation, we assume that the base station in a massive MIMO is equipped with a uniform linear array (ULA) with $N = 100$ omnidirectional antennas spaced a half wavelength apart (i.e., $d = \lambda/2$). A Gaussian mixture model is learned from the *a priori* power azimuth spread of the user channel. Specifically, the channel covariance matrix is generated as

$$\boldsymbol{R}_{\boldsymbol{hh}}^{(k)} = \int_{\mathcal{A}^{(k)}} \sigma_{\boldsymbol{h}}^2 \boldsymbol{a}(\theta)\boldsymbol{a}^{\mathrm{H}}(\theta)d\theta \tag{21}$$

according to the piecewise-Gaussian approximation, where $\boldsymbol{a}(\theta)$ is the steering vector of the ULA, $\sigma_{\boldsymbol{h}}^2$ is the channel power, and $\mathcal{A}^{(k)}$ denotes the k-th observation region at the base station (e.g., $\mathcal{A}^{(k)} \bigcap \mathcal{A}^{(k')} = \emptyset, \forall\, k, k' \in \mathcal{K}, k \neq k'$ and $\bigcup_{k\in\mathcal{K}} \mathcal{A}^{(k)} = (-\pi/2, \pi/2]$ for the ULA). The corresponding probability of the k-th Gaussian component, which reflects the power azimuth spread, can be modeled by a Laplacian distribution as [22]

$$p_k = \frac{1}{\sqrt{2}\sigma_{AS}} e^{-\frac{\sqrt{2}|\theta_k - \bar{\theta}|}{\sigma_{AS}}}, \tag{22}$$

where $\bar{\theta}$ and σ_{AS} respectively denote the mean direction-of-arrival (DOA) and the azimuth spread of the downlink channel. Although the Laplacian distribution is the most popular one in typical outdoor propagations, other classes of distributions are also applicable in certain circumstances [22].

In the simulation, the mean DOA is uniformly distributed as $\bar{\theta} \sim \mathcal{U}[-90^\circ, 90^\circ]$, and the azimuth spread is set as $\sigma_{AS} = 3^\circ$. Both the mean DOA and the azimuth spread are assumed to be known at the base station. In order to avoid the effects of user channel SNR scaling on the absolute error levels, we utilize the normalized MSE (NMSE), defined as

$$\mathrm{NMSE}(\boldsymbol{h}) = \frac{1}{N_{MC}} \sum_{q=1}^{N_{MC}} \frac{\left\| \boldsymbol{h}(q) - \hat{\boldsymbol{h}}^{\mathrm{MMSE}}(q) \right\|^2}{\|\boldsymbol{h}(q)\|^2}, \tag{23}$$

to evaluate the performance of channel estimation, where $N_{MC} = 1,000$ is the number of Monte-Carlo trials, and $\hat{\boldsymbol{h}}^{\mathrm{MMSE}}(q)$ is the MMSE estimate of $\boldsymbol{h}(q)$, i.e., the user channel obtained in the q-th Monte Carlo trial. The step size for the iterative pilot optimization is set as $\gamma = 0.1$.

Numerical results show that the maximum rank of the channel covariance matrices over different Gaussian components is 8, i.e., $\max_{k\in\mathcal{K}} r^{(k)} = 8$. In Fig. 1, we depict the NMSEs versus the input SNR of the channel with different pilot lengths. The NMSE performance is clearly a function of the input SNR for both the optimized and random pilot symbols, where the channels are scaled by $\sqrt{\mathrm{SNR}}$ to model the varying quality of the channel. From Fig. 1, it is observed that the channel estimation performance can be greatly improved by using the proposed pilot symbols as compared to the random pilot symbols. The performance advantage becomes more significant as the input SNR increases. It is also observed that, when the number of pilot symbols reaches the maximum rank of the covariance matrices over different Gaussian components, the channel estimation performance cannot be further improved by increasing the number of pilot symbols. On the contrary, when the number of pilot symbols is less than the maximum rank of the covariance matrices, the channel estimation performance can be further improved by increasing the number of pilot symbols. The simulation result is consistent with the asymptotic behavior analyzed in the Theorem.

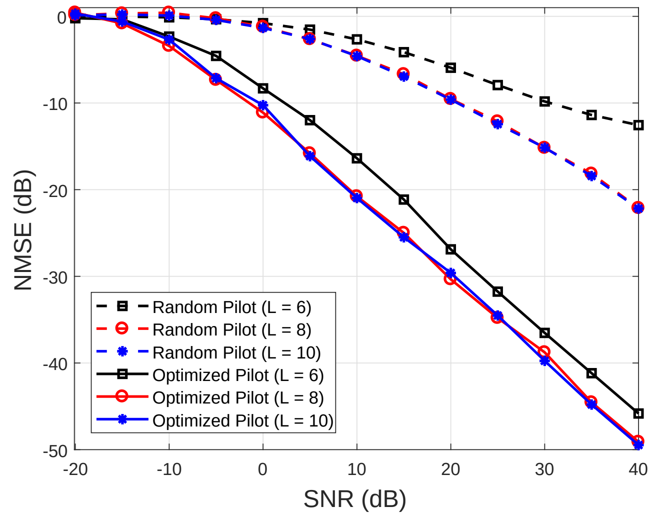

Fig. 1. NMSE performance comparison of the channel estimation versus the input SNR of the channel for different numbers of the pilot symbols.

5. CONCLUSION

In this paper, by modeling the channel vector in massive MIMO as a flexible and tractable Gaussian mixture distribution, we first proved that the channel can be perfectly recovered in the high-SNR asymptotic regime, provided that the number of pilot symbols is not less than the maximum rank of the channel covariance matrices of all Gaussian components. Then, we proposed an optimal pilot design by maximizing the mutual information between the measurements of the user and its corresponding channel vector. The proposed pilot optimization method can be extended to serve an arbitrary number of users in FDD massive MIMO systems by exploiting the *a priori* knowledge of the channel vectors to be estimated. With the available Gaussian mixture distribution, there is a closed-form solution to the underdetermined channel estimation problem under the MMSE criterion. Simulation results demonstrated that the proposed optimal pilot outperforms the random pilot in terms of the NMSE of channel estimation.

6. REFERENCES

[1] F. Rusek, D. Persson, Buon K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

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

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[4] Y. Gu and A. Leshem, "Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3881–3885, July 2012.

[5] Y. Gu, N. A. Goodman, S. Hong, and Y. Li, "Robust adaptive beamforming based on interference covariance matrix sparse reconstruction," *Signal Process.*, vol. 96, Part B, pp. 375–381, Mar. 2014.

[6] Y. Gu, Y. D. Zhang, and N. A. Goodman, "Optimized compressive sensing-based direction-of-arrival estimation in massive MIMO," in *Proc. IEEE ICASSP*, New Orleans, LA, Mar. 2017, pp. 3181–3185.

[7] C. Zhou, Y. Gu, S. He, and Z. Shi, "A robust and efficient algorithm for coprime array adaptive beamforming," *IEEE Trans. Veh. Technol.*, DOI: 10.1109/TVT.2017.2704610.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] J. Fang, X. Li, H. Li, and F. Gao, "Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1935–1947, Mar. 2017.

[13] S. Bazzi and W. Xu, "Downlink training sequence design for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Proc.*, vol. 65, no. 18, pp. 4732–4744, Sept. 2017.

[14] Y. Gu, N. A. Goodman, and A. Ashok, "Radar target profiling and recognition based on TSI-optimized compressive sensing kernel," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3194–3207, June 2014.

[15] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, "Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1356–1368, Mar. 2015.

[16] Y. Gu and N. A. Goodman, "Information-theoretic compressive sensing kernel optimization and Bayesian Cramér-Rao bound for time delay estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4525–4537, Sept. 2017.

[17] J. T. Flåm, S. Chatterjee, K. Kansanen, and T. Ekman, "On MMSE: A linear model under Gaussian mixture statistics," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3840–3844, Jul. 2012.

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[19] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Edition. Hoboken, NJ, USA: John Wiley & Sons, 2006.

[21] P. H. Schönemann, "A generalized solution of the orthogonal Procrustes problem", *Psychometrika*, vol. 31, no. 1, pp. 1–10, Mar. 1966.

[22] S. Saunders and A. Aragón-Zavala, *Antennas and Propagation for Wireless Communication Systems*, 2nd Edition. Hoboken, NJ, USA: John Wiley & Sons, 2007.