Downlink Channel Estimation in FDD Massive MIMO Systems Based on Multi-Resolution Discrimination Dictionary Learning

Zhentao Zou[a] Lin Chen[b] Xue Jiang[c*] Pei Xiao[d]

[a] School of Information Science and Engineering, Lanzhou University, Lanzhou, China

[b] Wireless Network RAN Research Department, Huawei Technologies Co. LTD, Shanghai, China

[c] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

[d] Institute for Communication Systems, University of Surrey, Guildford, U.K

***Abstract*—Compressive sensing (CS) techniques can be used to reduce the pilot overhead, and to improve the performance of channel estimation in massive multiple-input multiple-output (MIMO) systems. Most existing methods adopt the DFT matrix as a basis, which leads to direction mismatch and energy leakage problem in practice. However, the properties of geometry-based stochastic channel model (GSCM) are usually overlooked, but can be exploited to improve the performance of channel estimation. In this paper, a multi-resolution discrimination dictionary learning (MRDDL) method is proposed for downlink sparse channel estimation in frequency-division duplexing (FDD) MIMO systems. By taking into consideration that far scatterers in a specific cell are fixed at a certain position in the space and multipath angle of arrival (AOA) from far scatterers is concentrated in a fixed range, we design a specific dictionary for each far scatterer to reduce the redundant atoms. Simulations are conducted to validate the robustness and effectiveness of the MRDDL method over existing channel estimation methods.**

Keywords-massive MIMO, channel estimation, multi-resolution discrimination dictionary learning, FDD, compressive sensing.

I. Introduction

Massive multiple-input multiple-output (MIMO) has become a key technology in next-generation wireless systems [1]. Accurate acquisition of channel state information (CSI) is essential for large-scale MIMO, which can be utilized for beamforming and precoding. For conventional channel estimation technologies, such as least square (LS), the number of downlink training pilots is directly proportional to the number of base station (BS) antennas, which leads to large overhead and sometimes the pilot contamination problem in large-scale MIMO systems. In a time-division duplex (TDD) massive MIMO system, it is easier to obtain the CSI of the uplink channel on the BS side due to the limited number of users, and downlink CSI can be obtained by exploiting the uplink/downlink channel reciprocity. However, the full channel reciprocity is not hold in a frequency-division duplex (FDD) system [2], and the large antenna array makes it challenging to obtain the downlink CSI in FDD massive MIMO systems.

Compressive sensing (CS) techniques have been widely used in wireless communication systems in recent years [3]. They are based on the premise that the sparse vector can be recovered robustly with a number of measurement vectors, which is only proportional to the number of nonzero entries in a specific domain. It has been demonstrated that the massive MIMO channel has the sparse characteristic in the virtual angular domain, so that the techniques based on CS can be used to estimate sparse channel accurately while alleviating the downlink training pilots overhead. Most existing methods based on CS techniques [4–6] only use the DFT dictionary, which can be considered as uniformly fixed sampling over the angular domain to transform the MIMO channel into a sparse channel vector in the virtual angular domain. Unfortunately, the DFT dictionary matrix may have the problem that the actual azimuth angle of arrival (AOA) or azimuth angle of departure (AOD) does not match the atomic direction of the dictionary, which leads to serious energy leakage. A simple solution to avoid the mismatch problem is to increase the DFT size. To futher improve the performance of channel estimation, the authors in [7] first use dictionary learning algorithm to learn a specific dictionary as a basis to represent massive MIMO channel. Other scholars have further improved the performance of dictionary learning by considering additional conditions, such as partitioning the cell into several sectors in advance [8] and block-structured dictionary learning [9].

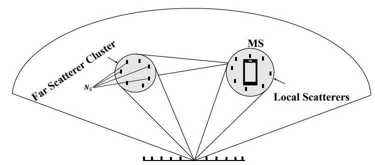


Figure 1. Principle of the GSCM.

The dictionary learning methods [7–9] usually utilize the *Geometry-Based Stochastic Channel Model* (GSCM) [10], as illustrated in Fig 1. This type of channel has been widely used in wireless communications including FDD massive MIMO systems. However, the properties of GSCM have not been fully exploited in existing dictionary learning methods. According to [10], there exist a few far scatterers and local scatterers in the communication system. Those far scatterers such as high-rise buildings and mountains, are far away from both the mobile station (MS) and BS. The position of the far scatterers can be deemed to be fixed in space, which means the range of AOA from far scatterers at the MS is approximately invariant though the position of the MS changes.

To further exploit the above-mentioned properties of the GSCM, we design a specific dictionary for each far scatterer and propose a new dictionary learning method named *multi-*

*Corresponding author.

This work was supported by the National Natural Science Foundation of China under Grant 62022054 and 61971279.

resolution discrimination dictionary learning (MRDDL) to improve the performance of channel estimation. In addition, we propose an effective method to solve the MRDDL algorithm and further reveal the physical meaning of dictionary learning for channel estimation.

Notations used in this paper are as follows. Uppercase and lowercase boldface letters are used to denote matrices and vectors, respectively. $\|x\|_0$, $\|x\|_1$, $\|x\|_2$, $\|x\|_F$ are used to represent ℓ_0-norm, ℓ_1-norm, ℓ_2-norm and Frobenius norm, respectively. We use $(\cdot)^T$ to denote the transpose and $\boldsymbol{a_j}$ to denote the *j*-th column of $\mathbf{A}$. The notation $\mathbb{C}$ denotes the complex set and diag($\boldsymbol{a}$) is a diagonal matrix containing the elements of $\boldsymbol{a}$ on the diagonal.

II. SYSTEM MODEL

A. Signal Model

We consider a single-cell FDD massive MIMO system where the BS employs N antennas and each MS has a single antenna. The downlink channel is a narrowband block fading channel, there are N_c dominant scattering clusters for each channel response and N_s is the number of propagation subpaths in each cluster. Futhermore, we define the total number of far scattering clusters in the geographic space as N_f. Thus, the downlink channel response can be modeled as

$$\mathbf{h} = \sum_{c=1}^{N_c} \sum_{s=1}^{N_s} \alpha_{c,s} \boldsymbol{a}\left(\theta_{c,s}\right) \tag{1}$$

where $\alpha_{c,s}$ denotes the complex gain of the *s*-th subpath in the *c*-th scattering cluster for the downlink, $\boldsymbol{a}\left(\theta_{c,s}\right)$ is the array response, and $\theta_{c,s}$ is the corresponding AOD for the downlink transmission. Denote the wavelength of downlink as b, the space between antennas is $d = b/2$ and the array response of a uniform linear array (ULA) is denoted as

$$\boldsymbol{a}(\theta_{c,s}) = \left[1, e^{j2\pi \frac{d}{b} \sin(\theta_{c,s})}, \ldots, e^{j2\pi \frac{d}{b} \sin(\theta_{c,s}) \cdot (N-1)}\right]^T \tag{2}$$

The received signal at mobile user (MU) is given by

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{n} \tag{3}$$

where $\mathbf{h} \in \mathbb{C}^{N \times 1}$ denotes the downlink channel vector, $\mathbf{A} \in \mathbb{C}^{T \times N}$ is the matrix corresponding to the pilots transmitted during the T training period, and $\mathbf{n} \in \mathbb{C}^{T \times 1}$ is the Gaussian noise vector during the signal transmission. The downlink channel can be estimated at each MU based on the received signal and corresponding training pilots.

B. CS-Based Donwlink Channel Estimation

Thanks to the sparsity in the virtual angular domain for MIMO channel, the channel can be estimated with much smaller measurements ($T \ll N$) using the CS theory [4–6]. This process can be represented as

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{n} = \mathbf{A}\mathbf{D}\mathbf{x} + \mathbf{n} \tag{4}$$

where $\mathbf{x}$ is a sparse vector in the virtual angular domain. The sparse basis $\mathbf{D}$ can be a DFT matrix, an overcomplete DFT matrix or a learning dictionary. If we use η to represent the modeling error, the channel estimation problem can be formulated as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{D}\mathbf{x}\|_2 \leq \eta \tag{5}$$

The optimization problem (5) can be solved by sparse approximation algorithms efficiently [11].

III. MULTI-RESOLUTION DISCRIMINATION DICTIONARY LEARNING ALGORITHM

A. Overview of Existing Dictionary Learning Algorithms

The dictionary learning methods aim to learn a specific dictionary to represent channel sparsely. In [7], the dictionary learning-based channel model (DLCM) is proposed as

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{x}_m} \quad & \frac{1}{M} \sum_{m=1}^{M} \|\mathbf{x}_m\|_0 \\ \text{s.t.} \quad & \|\mathbf{h}_m - \mathbf{D}\mathbf{x}_m\|_2 \leq \eta, \forall m \end{aligned} \tag{6}$$

where M is the number of training channel samples, the authors in [7–9] rotate the ODFT atoms to match the AOD through the dictionary learning method. In [8], a specific cell is partitioned into several sectors and a discriminative dictionary learning method is proposed to learn a specific sub-dictionary for each sector. However, the number of sub-dictionaries related to sectors in a cell is large, and it is uncertain which sector in the cell a specific user comes from. In addition, it is complex to determine which sector-related sub-dictionary to use for downlink channel estimation. There still exist redundant atoms for these methods and the properties of GSCM have not been fully exploited.

B. The Proposed Algorithm-MRDDL

As discussed in Section I, the multipath AOA from far scatterers is concentrated in a fixed range since far scatterers are fixed at a certain position in space for a specific cell. Thus, we design a specific dictionary for each far scatterer to reduce the redundant atoms in the dictionary.

We propose a MRDDL-based framework for downlink channel estimation in FDD massive MIMO systems. which learns a low-resolution dictionary and a high-resolution discriminative dictionary for each far scatterer with good representation ability. The architecture of proposed dictionary is shown in Fig 2. $\mathbf{D}_0 \in \mathbb{C}^{N \times P_0}$ is the low-resolution dictionary, which provides a sampling grid over the angular domain $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. $\mathbf{D}_i \in \mathbb{C}^{N \times P_i}$ $(i = 1, 2, \cdots, N_f)$ is a high-resolution dictionary with sampling grid over the angular domain $[\phi_i - \Delta\phi, \phi_i + \Delta\phi]$. ($\phi_i$ is the angle of main lobe of AOD for the *i*-th far scatterer and $\Delta\phi$ is the angular spread). The proposed dictionary is formed as $\mathbf{D} = \left[\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2 \ldots, \mathbf{D}_{N_f}\right] \in \mathbb{C}^{N \times P}$ $(P = \sum_{i=0}^{N_f} P_i)$. The dictionary $\mathbf{D}$ and the sparse representation coefficients can be optimized by

$$\begin{aligned} \arg\min_{\mathbf{D}, \mathbf{X}} \|\mathbf{H} - \mathbf{D}\mathbf{X}\|_F^2 + \sum_{m=1}^{M} \Big[& \|\mathbf{h}_m - \mathbf{D}\mathbf{W}_m \mathbf{x}_m\|_2^2 \\ & + \left\|\mathbf{D}\widetilde{\mathbf{W}}_m \mathbf{x}_m\right\|_2^2 \Big] + \lambda \|\mathbf{X}\|_1 \end{aligned} \tag{7}$$

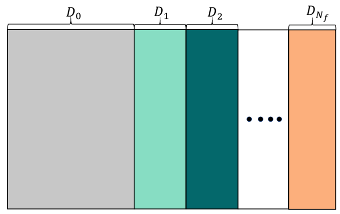


Figure 2. Structure of the proposed dictionary.

The first term constitutes the overall representation ability of the learned dictionary, where $\mathbf{H} \in \mathbb{C}^{N\times M}$ is the downlink training channel samples in a specific cell, and $\mathbf{X} \in \mathbb{C}^{P\times M}$ is a sparse channel coefficient vector in the virtual domain. The last term adds the sparse constraint on the channel coefficient vector, and λ represents the positive regularization parameters. The second term makes up the discriminative ability so that the atoms of each high-resolution dictionary are matched with the corresponding AOA from the far scatterer, where $\mathbf{W}_m = \text{diag}(\mathbf{w}_m)$, $\mathbf{w}_m \in \mathbb{N}^{1\times P}$ is a binary vector, the *j*-th column of $\mathbf{w}_m$ is one if the *j*-th column of $\mathbf{D}$ belongs to the corresponding scatterer index set Ω_m, which denotes the local scatterer and $N_c - 1$ far scatterers index for the *m*-th channel sampling and $\Omega_m \subseteq \{0, 1, 2, \cdots, N_f\}$. The third term sets the coefficients which are not associated with corresponding far scatterers to be zero, $\widetilde{\mathbf{W}}_m$=diag$(\widetilde{\mathbf{w}}_m)$ and $\widetilde{\mathbf{w}}_m$ is also a binary vector whose element is opposite to $\mathbf{w}_m$.

C. *Efficient Solutions for Optimization Problems*

The dictionary update process can be divided into two stages, i.e., we can update sparse coefficients $\mathbf{X}$ while keeping the dictionary $\mathbf{D}$ fixed and then update $\mathbf{D}$ while $\mathbf{X}$ remains fixed [12]. The process may be repeated many times until convergence. The original problem (7) is equivalent to :

$$\arg\min_{\mathbf{D},\mathbf{X}} \quad \mathbf{J_H}(\mathbf{D},\mathbf{X}) = \mathbf{F_H}(\mathbf{D},\mathbf{X}) + \lambda\|\mathbf{X}\|_1 \tag{8}$$

$$\mathbf{F_H}(\mathbf{D},\mathbf{X}) = \|\mathbf{H} - \mathbf{D}\mathbf{X}\|_F + \sum_{m=1}^{M}\left[\|\mathbf{h}_m - \mathbf{D}\mathbf{W}_m\mathbf{x}_m\|_2^2 + \left\|\mathbf{D}\widetilde{\mathbf{W}}_\mathbf{m}\mathbf{x}_m\right\|_2^2\right] \tag{9}$$

1) *MRDDL Sparse Coefficient Update* (MRDDL-X):
When $\mathbf{D}$ is fixed, $\mathbf{X}$ is updated by solving

$$\mathbf{X} = \arg\min_{\mathbf{X}} \quad \mathbf{J_H}(\mathbf{D},\mathbf{X}) = \mathbf{F_H}(\mathbf{D},\mathbf{X}) + \lambda\left\|\mathbf{X}\right\|_1 \tag{10}$$

The problem (10) can be solved by the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) [13]. First, we should calculate the gradient of $\mathbf{F_H}(\mathbf{D},\mathbf{X})$.

We define $\mathbf{D}^{\Omega_m} = \left[\mathcal{T}(\mathbf{D}_0), \mathcal{T}(\mathbf{D}_1)\cdots\mathcal{T}\left(\mathbf{D}_{N_f}\right)\right]$ and

$$\mathcal{T}(\mathbf{D}_n) = \begin{cases}\mathbf{D}_n & n \in \Omega_m \\ 0 & n \notin \Omega_m\end{cases} \tag{11}$$

Similarly, we define $\mathbf{D}^{\widetilde{\Omega}_m} = \left[\mathcal{G}(\mathbf{D}_0), \mathcal{G}(\mathbf{D}_1)\cdots\mathcal{G}\left(\mathbf{D}_{N_f}\right)\right]$ and

$$\mathcal{G}(\mathbf{D}_n) = \begin{cases}0 & n \in \Omega_m \\ \mathbf{D}_n & n \notin \Omega_m\end{cases} \tag{12}$$

$$\frac{\partial\mathbf{F_H}(\mathbf{D},\mathbf{X})}{\partial\mathbf{X}} = \mathbf{D}^T\mathbf{D}\mathbf{X} - \mathbf{D}^T\mathbf{H} + \sum_{m=1}^{M}\left[\left(\mathbf{D}^{\Omega_m}\right)^T\mathbf{D}^{\Omega_m}\mathbf{x}_m - \left(\mathbf{D}^{\Omega_m}\right)^T\mathbf{h}_m + \left(\mathbf{D}^{\widetilde{\Omega}_m}\right)^T\mathbf{D}^{\widetilde{\Omega}_m}\mathbf{x}_m\right] \tag{13}$$

We can use FISTA [13] to update X once $\nabla\mathbf{F_H}(\mathbf{D},\mathbf{X})$ has been calculated, and the Lipschitz coefficient L of $\nabla\mathbf{J_H}(\mathbf{D},\mathbf{X})$ should be calculated to control the iteration step. The algorithm is summarized in Algorithm 1.

Algorithm 1 MRDDL Sparse Coefficients Update by FISTA

Input: Channel Samples $\mathbf{H}$, Scatter Index $\mathbf{W}_m$, Dictionary $\mathbf{D}$, Regularization parameter λ, Sparse Coefficients $\mathbf{X}$.
Output: Sparse Coefficients $\mathbf{X}$.

function $\mathbf{X} = \mathbf{MRDDL\_X}(\mathbf{H}, \mathbf{W}_m, \mathbf{D}, \lambda, \mathbf{X})$
 Calculate:

$$\mathbf{A} = \mathbf{D^T D}$$
$$\mathbf{B}^m = \left(\mathbf{D}^{\Omega_m}\right)^T\mathbf{D}^{\Omega_m}$$
$$\mathbf{C}^m = \left(\mathbf{D}^{\widetilde{\Omega}_m}\right)^T\mathbf{D}^{\widetilde{\Omega}_m}$$
$$L = \lambda_{max}(\mathbf{A}) + \lambda_{max}\left[\frac{1}{M}\sum_{i=1}^{M}(\mathbf{B}^m + \mathbf{C}^m)\right] + 6\lambda$$

 Initialize: $\mathbf{P_1} = \mathbf{Q_0} = \mathbf{X}$, t_1=1, k=1
 while not converge and $k < k_{max}$ **do**
  Calculate gradient:

$$\mathbf{G} = \mathbf{A}\mathbf{X} - \mathbf{D^T H} + \sum_{i=1}^{M}[(\mathbf{B}_m + \mathbf{C}_m)\mathbf{x}_m - (\mathbf{D}^{\Omega_m})^T\mathbf{h}_m]$$

  $\mathbf{Q}_k = \mathcal{S}_{\lambda/L}\left(\mathbf{P_k} - \mathbf{G}/L\right)$ ($\mathcal{S}_\alpha()$ is the soft thresholding function. $\mathcal{S}_\alpha(x) = \text{sgn}(\mathbf{x})\left(|\mathbf{x}| - \alpha\right)_+$).
  t_{k+1}=$(1 + \sqrt{1 + 4t_k^2})/2$
  $\mathbf{P}_{k+1} = \mathbf{Q}_k + \frac{t_k - 1}{t_k + 1}\left(\mathbf{Q}_k - \mathbf{Q}_{k-1}\right)$
  $k = k + 1$
 end while
end function

---

2) *MRDDL Dictionary Update* (MRDDL-D):
When $\mathbf{X}$ is fixed, $\mathbf{D}$ can be updated by solving

$$\mathbf{D} = \arg\min_{\mathbf{D}} \quad \mathbf{F_H}(\mathbf{D},\mathbf{X}) \tag{14}$$

The dictionary update problem (14) is equivalent to

$$\mathbf{D} = \arg\min_{\mathbf{D}}\left\{-2\,\text{trace}\left(\mathbf{E}\mathbf{D}^T\right) + \text{trace}\left(\mathbf{F}\mathbf{D}^T\mathbf{D}\right)\right\} \tag{15}$$

where $\mathbf{E} = \mathbf{H}(\mathbf{X} + \mathcal{M}(\mathbf{X}))^T$, $\mathbf{F} = \mathbf{X}\mathbf{X^T} + \mathcal{M}(\mathbf{X})\mathcal{M}(\mathbf{X})^\mathbf{T} + \mathcal{N}(\mathbf{X})\mathcal{N}(\mathbf{X})^\mathbf{T}$. Refer to Appendix for the proof. We can use Online Dictionary Learning (ODL) [14] to update D efficiently.

In order to speed up convergence, we need to initialize $\mathbf{D}$ by setting the low-resolution dictionary to the ODFT matrix and each high-resolution dictionary to the DFT matrix. The algorithm converges faster when we modify sparse coefficients

$\mathbf{X}$ via equation (16) after updating D. We summarize the whole process in Algorithm 2.

Algorithm 2 MRDDL Algorithm

Input: Channel Samples $\mathbf{H}$, Scatter Index $\mathbf{W}_m$, Regularization parameter λ.

Output: Dictionary $\mathbf{D}$.

function $\mathbf{D} = \mathbf{MRDDL}(\mathbf{H}, \mathbf{W}_m, \lambda)$

  Initialization $\mathbf{D}$, and:

$$\mathbf{X} = \arg\min_{\mathbf{X}} \|\mathbf{H} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_1$$

  while not converge **do**

    Update $\mathbf{X}$ by Algorithm 1.

    Update $\mathbf{D}$ by ODL:

$$\mathbf{E} = \mathbf{H}(\mathbf{X} + \mathcal{M}(\mathbf{X}))^{\mathbf{T}}$$

$$\mathbf{F} = \mathbf{X}\mathbf{X}^T + \mathcal{M}(\mathbf{X})\mathcal{M}(\mathbf{X})^T + \mathcal{N}(\mathbf{X})\mathcal{N}(\mathbf{X})^T$$

$$\mathbf{D} = \arg\min_{\mathbf{D}} \left\{-2\,\mathrm{trace}\left(\mathbf{E}\mathbf{D}^T\right) + \mathrm{trace}\left(\mathbf{F}\mathbf{D}^T\mathbf{D}\right)\right\}$$

    Modify Sparse Coefficients:

$$\mathbf{X} = \arg\min_{\mathbf{X}} \|\mathbf{H} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_1 \tag{16}$$

  end while

end function

Once obtaining the dictionary $\mathbf{D}$ through the MRDDL algorithm, the sparse channel coefficient vector can be calculated by

$$\widetilde{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{D}\mathbf{x}\|_2 \leq \eta \tag{17}$$

where $\widetilde{\mathbf{x}} = \left[\widetilde{\mathbf{x}}_0, \widetilde{\mathbf{x}}_1, \ldots \widetilde{\mathbf{x}}_{N_f}\right]^T \in \mathbb{C}^{P\times 1}$, and $\widetilde{\mathbf{x}}_k$ is the coefficient vector related to sub-dictionary $\mathbf{D}_k$ $(k = 0, 1, \ldots N_f)$. The estimated channel is obtained as $\widetilde{\mathbf{h}} = \mathbf{D}\widetilde{\mathbf{x}}$.

IV. Experimental Results

In the simulation, the BS employs 100 antennas with ULA and each MS has a single antenna, the cell specific scattering clusters are generated following the principle of GSCM [10]. Each scatter has 20 subpaths with 4 angular spread. Far scatter clusters are generated with radius 1500m and $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, and then kept constant in the whole process of dictionary learning and channel estimation. Each channel response consists of multipath AOA from three far scatterers which are closest to MS and one local scatterer which is generated randomly according to the location of MS. The other related channel parameters, such as delay spread, angular spread, and path power are generated according to the 3GPP standard [15].

The recovered sparse channel coefficient vector by the learning dictionary is depicted in Fig 3. The nonzero entries of coefficients $\widetilde{x}$ are only located at the corresponding position of sub-dictionary $\{\mathbf{D}_0, \mathbf{D}_2, \mathbf{D}_4, \mathbf{D}_5\}$, which is consistent with the scatterer index set $\Omega = \{0, 2, 4, 5\}$. It also shows the discriminative and representative ability of the learning dictionary.

To evaluate the performance of our proposed model MRDDL, we calculate the performance of downlink channel

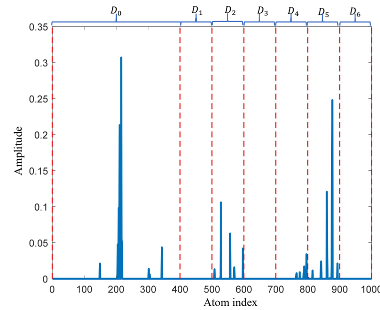
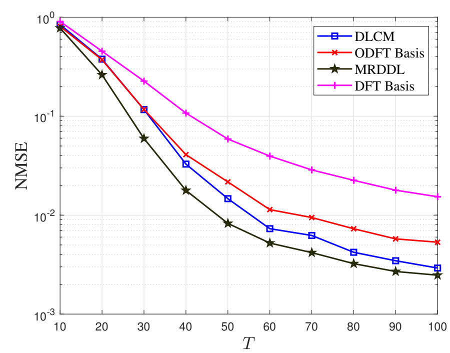


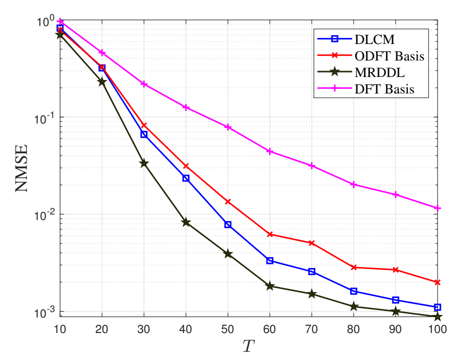
Figure 3. Recovered sparse channel coefficient vector for k-th user

estimation based on CS techniques by applying different sparsifying matrices. We compare our algorithm to 100×100 DFT basis, 100× 400 ODFT basis, and dictionary learning algorithm [7]. We generate training pilots A as i.i.d. $\mathcal{CN}(0, \rho/N)$, so that $\mathbb{E}\|\mathbf{A}\|_F^2 = \rho T$. We adopt normalized mean square error (NMSE) as the performance metric, which is defined as $\mathrm{NMSE} = \mathbb{E}\left\{\|\mathbf{h} - \hat{\mathbf{h}}\|_2^2 / \|\mathbf{h}\|_2^2\right\}$.

Fig 4 shows the NMSE performance with respect to the number of downlink training pilots. Our proposed algorithm MRDDL achieves superior performance over the existing algorithms. The required training pilots of MRDDL are much less than the DFT basis, ODFT basis and DLCM. To achieve the same NMSE performance, MRDDL saves nearly 20% pilot overhead in comparison to DLCM. The figure also shows that the performance of the four algorithms has improved with the increase in SNR.



(a) 25dB



(b) 30dB

Figure 4. NMSE comparison of different sparsifying matrices for downlink channel estimation in FDD massive estimation

V. CONCLUSIONS

In this paper, we developed an MRDDL algorithm based downlink channel estimation in FDD massive MIMO systems. By taking into consideration that far scatterers in a specific cell are fixed at a certain position in space and the range of multipath AOA from far scatterers remains unchanged, we design a specific dictionary for each far scatterer to reduce the redundant atoms. Moreover, we evaluated the performance of channel estimation based on the learned dictionary. Simulation results demonstrated the robustness and effectiveness of our algorithm compared to the existing CS-based algorithm. Consequently, our model can be expanded to a joint uplink/-downlink channel estimation frame and this MRDDL model can be used as an efficient framework in future massive MIMO systems.

APPENDIX

$\mathbf{F_H}(\mathbf{D}, \mathbf{X})$ can be rewritten as

$$\|\mathbf{H}-\mathbf{D}\mathbf{X}\|_F^2+\sum_{m=1}^{M}\left[\|\mathbf{h}_m-\mathbf{D}\mathbf{W}_m\mathbf{x}_m\|_2^2+\left\|\mathbf{D}\widetilde{\mathbf{W}}_m\mathbf{x}_m\right\|_F^2\right]$$
$$=\textbf{trace}\left[\left[\mathbf{X}\mathbf{X}^T+\sum_{i=1}^{M}\left[\mathbf{W}_m\mathbf{x}_m\mathbf{x}_m^T\mathbf{W}_m^T+\widetilde{\mathbf{W}}_m\mathbf{x}_m\mathbf{x}_m^T\widetilde{\mathbf{W}}_m\right]\right]\right.$$
$$\left.\mathbf{D^T}\mathbf{D}\right]-2\textbf{trace}\left[\left[\mathbf{H}\mathbf{X^T}+\sum_{i=1}^{M}\mathbf{h}_m\mathbf{x}_m^T\mathbf{W}_m\right]\mathbf{D^T}\right]+\textbf{constant}$$
$$=-2\textbf{trace}(\mathbf{E}\mathbf{D^T})+\textbf{trace}(\mathbf{F}\mathbf{D^T}\mathbf{D})+\textbf{constant}$$

where we have defined

$$\begin{aligned}\mathbf{E}&=\mathbf{H}\mathbf{X^T}+\sum_{i=1}^{M}\mathbf{h}_m\mathbf{x}_m^T\mathbf{W}_m\\&=\mathbf{H}\mathbf{X}^T+\left[\mathbf{h}_1\left(\mathbf{x}_1^{\Omega_1}\right)^T\dots\mathbf{h}_m\left(\mathbf{x}_m^{\Omega_m}\right)^T\right]\\&=\mathbf{H}\left\{\mathbf{X}^T+\left[\left(\mathbf{x}_1^{\Omega_1}\right)^T\cdots\left(\mathbf{x}_m^{\Omega_m}\right)^T\right]\right\}\\&=\mathbf{H}(\mathbf{X}+\mathcal{M}(\mathbf{X}))^T\end{aligned}$$

$$\begin{aligned}\mathbf{F}&=\mathbf{X}\mathbf{X^T}+\sum_{i=1}^{M}\left[\mathbf{W}_m\mathbf{x}_m\mathbf{x}_m^T\mathbf{W}_m^T+\widetilde{\mathbf{W}}_m\mathbf{x}_m\mathbf{x}_m^T\widetilde{\mathbf{W}}_m\right]\\&=\mathbf{X}\mathbf{X^T}+\left[\left(\mathbf{x}_1^{\Omega_1}\right)\left(\mathbf{x}_1^{\Omega_1}\right)^T\cdots\left(\mathbf{x}_m^{\Omega_m}\right)\left(\mathbf{x}_m^{\Omega_m}\right)^T\right]+\\&\quad\left[\left(\mathbf{x}_1^{\widetilde{\Omega}_1}\right)\left(\mathbf{x}_1^{\widetilde{\Omega}_1}\right)^T\cdots\left(\mathbf{x}_m^{\widetilde{\Omega}_m}\right)\left(\mathbf{x}_m^{\widetilde{\Omega}_m}\right)^T\right]\\&=\mathbf{X}\mathbf{X^T}+\mathcal{M}(\mathbf{X})\mathcal{M}(\mathbf{X})^\mathbf{T}+\mathcal{N}(\mathbf{X})\mathcal{N}(\mathbf{X})^\mathbf{T}\end{aligned}$$

$$\begin{aligned}\mathcal{M}(\mathbf{X})&=\left[\left(\mathbf{x}_1^{\Omega_1}\right)\cdots\left(\mathbf{x}_m^{\Omega_m}\right)\right]\\\mathcal{N}(\mathbf{X})&=\left[\left(\mathbf{x}_1^{\widetilde{\Omega}_1}\right)\cdots\left(\mathbf{x}_m^{\widetilde{\Omega}_m}\right)\right]\end{aligned}$$

the definition of $\mathbf{x}_m^{\Omega_m}$, $\mathbf{x}_m^{\widetilde{\Omega}_m}$ is similar to $\mathbf{D}_m^{\Omega_m}$, $\mathbf{D}_m^{\widetilde{\Omega}_m}$, as expressed in equations (11), (12).

REFERENCES

[1] E. G. Larsson and L. Van der Perre, "Massive MIMO for 5G," *IEEE 5G Tech Focus*, vol. 1, no. 1, pp. 1–4, 2017.

[2] L. Chen, X. Jiang, Z. Zhong, X. Liu, and M. Haardt, "Tensor-based downlink channel reconstruction for FDD massive MIMO," in *Proc. IEEE Asilomar Conf. Signals. Syst. Comput.*, 2021, pp. 1453–1458.

[3] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, 2017.

[4] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, 2015.

[5] 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, 2015.

[6] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2852–2862, 2017.

[7] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, 2018.

[8] J. Chen, X. Zhang, and P. Zhang, "DDL-based sparse channel representation and estimation for downlink FDD massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–6.

[9] Y.-D. Huang, Y.-C. Liang, and F. Gao, "Channel estimation in FDD massive MIMO systems based on block-structured dictionary learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.

[10] A. F. Molisch, A. Kuchar, J. Laurila, K. Hugl, and R. Schmalenberger, "Geometry-based directional model for mobile radio channels-principles and implementation," *Eur. Trans. Telecommun.*, vol. 14, no. 4, pp. 351–359, 2003.

[11] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[12] T. H. Vu and V. Monga, "Fast low-rank shared dictionary learning for image classification," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5160–5175, 2017.

[13] 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.

[14] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding." *J. Mach. Learn. Res.*, vol. 11, no. 1, 2010.

[15] "Spatial channel model for multiple input multiple output (MIMO) simulations," document 25.996 Version 12.0.0 Release 12, 3GPP, 2014.