# Downlink channel estimation for millimeter wave communication combining low-rank and sparse structure characteristics

Jin Zhou[1] 🔘

## Abstract

The acquisition of channel state information (CSI) is essential in millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems. The mmWave channel exhibits sparse scattering characteristics and a meaningful low-rank structure, which can be simultaneously employed to reduce the complexity of channel estimation. Most existing works recover the low-rank structure of channels using nuclear norm theory. However, solving the nuclear norm-based convex problem often leads to a suboptimal solution of the rank minimization problem, thus degrading the accuracy of channel estimation. Previous contributions recover the channel using over-complete dictionary with the assumption that the mmWave channel can be sparsely represented under some dictionary. While over-complete dictionary may increase the computational complexity. To address these problems, we propose a channel estimation framework based on non-convex low-rank approximation and dictionary learning by exploring the joint low-rank and sparse representations of wireless channels. We surrogate the widely used nuclear norm theory with non-convex low-rank approximation method and design a dictionary learning algorithm based on channel feature classification employing deep neural network (DNN). Our simulation results reveal the proposed scheme outperform the conventional dictionary learning algorithm, Bayesian framework algorithm, and compressed sensing-based algorithms.

**Keywords** Sparse representation · Non-convex theory · Low-rank approximation · Channel state information · Deep neural network

## 1 Introduction

With the rapid increase of demand for high-speed wireless transmission communication systems, massive multiple input and multiple output systems (MIMO) have attracted extensive attention in academy and industry due to their outstanding ability to improve system capacity and spectrum utilization rate [1, 2]. MIMO technology has been widely used in advanced communication standards, such as IEEE 802.11 ac [3], IEEE 802.16m [4], and 3GPP Long Term Evolution Networks [5, 6]. Owing to the extremely high attenuation and serious signal absorption at the mmWave frequency bands, mmWave communication systems employ large antenna arrays at the base station. Obtaining accurate channel state information is a prerequisite for gaining optimal system performance. In the aspect of CSI detection, Time Division Duplex (TDD) mode takes advantage of the reciprocity of the uplink and downlink link. In Frequency Division Duplex (FDD) mode where the channel reciprocity condition is no longer satisfied, the base station sends a downlink pilot signal, the mobile station receives and detects the pilot signal and then feeds back CSI to the base station. As for traditional channel estimation method, the length of pilot sequence must be proportional to the number of base station antennas, which makes it difficult to complete channel estimation within coherent time. Moreover, the uplink feedback load is high. Therefore, it is unrealistic to use traditional methods for the channel estimation of massive MIMO systems.

However, recent advances in sparse representation and compressed sensing have inspired new approaches, e.g., distributed compressive sensing (DCS) [7], matrix completion [8], and deep neural network (DNN) [9], aiming at reducing heavy pilot load and high uplink feedback quantization overhead. Distributed compressed sensing channel estimation algorithm based on channel joint sparse structure characteristics in multi-user environment was proposed to reduce pilot load and uplink feedback overhead in Rao and Lau [7]. Z. Gao [10] originally proposed a structured compressed sensing

✉ Jin Zhou
zhoujin@tjufe.edu.cn

1 School of Science and Technology, Tianjin University of Finance and Economics, Tianjin 300222, China

framework based on joint channel estimation of space and time by exploring the common sparse support set of MIMO channels to reduce pilot load. However, this method increases the computational load of the terminal, and the signal recovery probability based on subspace tracking algorithm needs to be improved. In Fang et al [11], the author proposed an adaptive channel estimation and feedback framework based on spatial common sparsity in FDD massive MIMO communication environment. According to the low-rank characteristics of massive MIMO channels, the author in Sun et al. [12] investigated a method for FDD downlink channel estimation which mines the low-rank or near-low-rank characteristics of the channel covariance matrix. However, the computation of the channel covariance matrix is too large to be suitable for the actual communication scenarios of multi-user cells. Downlink channel estimation for FDD was performed by using the low-rank characteristics of massive MIMO channel with iterative optimization and deep learning methods in [2]. However, there exist several major problems with the above methods:

(1) Only the sparse or low-rank characteristics of the channel are utilized, while the joint channel estimation with sparse and low-rank characteristics can further reduce the training and feedback load and overhead.

(2) In many practical application scenarios, the channel state information only shows sparsity under the conversion of suitable bases. Therefore, the above algorithm based on channel sparsity will degrade the accuracy of channel state estimation in some actual communication scenario. According to the channel sparsity in the angular domain, Wei [13] introduced the learning dictionary instead of the predefined dictionary to estimate the downlink channel of FDD massive MIMO system and obtains better channel sparse representation performance. However, the learning dictionary of this method comes from the actual measured values of the channel, and the experimental operation is complex, which is not suitable for industrial information transmission systems with high real-time requirements.

In addition to the sparse scattering nature, mmWave channels may exhibit a low-rank structure. The rank minimization problem is challenging to solve. Thus, rank function is usually substituted by the convex nuclear norm, leading to a relaxed convex formulation of channel estimation problem [14–19]. However, the obtained recovery of channels by convex nuclear norm is usually suboptimal because the nuclear norm is a loose approximation of the rank function [16]. In Lu and Tang [16], the authors proposed a non-convex approximation of $l_0$ norm to approximate the rank function, which is solved by iterative reweighted nuclear norm algorithm.

In this study, a novel channel estimation framework is proposed, which utilizes the joint low-rank and block sparsity

feature of mmWave channels in the angle domain and deep learning network-based sparsity representation. For the low-rank characteristics of the channel, a non-convex method is used for low-rank approximation, with its iterative optimization algorithm designed. Besides, we present a deep learning network-based dictionary learning method such that a dictionary is learned from the output of deep learning network which extracts key characteristics of mmWave channel measurements. The learned dictionary adapts specifically to the key characteristic of the cell and promotes a more efficient and robust channel sparse representation, which in turn boosts the performance of the channel estimation.

Notations: $l_0$ norm is the number of non-zero elements in a vector. $l_1$ norm is the sum of the absolute values of elements in the vector.

## 2 System model

We consider a single-cell massive MIMO system working in FDD mode. The BS is equipped with $N$ antennas, and each mobile station employs $M$ antennas. To realize downlink channel estimation, the base station sends training pilot sequence to the mobile station. The mobile station feeds back CSI to the base station. The pilot signal received at the mobile station at the $j$th time slot ($j = 1, 2, \cdots, T$) is expressed as

$$\boldsymbol{y}_j = \sqrt{\rho}\boldsymbol{h}^T\boldsymbol{x}_j + \boldsymbol{n}_j \tag{1}$$

where $\boldsymbol{h} \in \mathbb{C}^{N \times 1}$ is the downlink channel response vector between the BS and the mobile user and $\boldsymbol{n} \in \mathbb{C}^{T_d \times 1}$ is the noise vector at the receiver such that $\boldsymbol{n} \sim \mathcal{CN}(0, \mathbf{I})$. $\boldsymbol{x}_j \in \mathbb{C}^{1 \times N}$ represents the downlink pilot sequence vector transmitted during the training period of $T_d$ symbols, where $\|\boldsymbol{p}\|_F^2 = \rho_d T_d$ such that $\rho_d$ is the training SNR. The concatenated received signals in T time slots $\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_T] \in \mathbb{C}^{1 \times T}$ is expressed as

$$\boldsymbol{y} = \sqrt{\rho}\boldsymbol{h}^T\boldsymbol{X} + \boldsymbol{n} \tag{2}$$

where $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T] \in \mathbb{C}^{N \times T}$ and $N = [\boldsymbol{n}_1, \boldsymbol{n}_2, \cdots, \boldsymbol{n}_T] \in \mathbb{C}^{N \times T}$ are the aggregated signal and noise, respectively. Using a conventional channel estimation method such as the least square method, the channel estimation is given by the following formula:

$$\hat{\boldsymbol{h}}^{LS} = \boldsymbol{X}^\dagger\boldsymbol{y} \tag{3}$$

where $\boldsymbol{X}^\dagger$ is the Moore-Penrose pseudoinverse. Accurate acquisition of channel state information by conventional method requires $T_d \geq N$; the length of pilot sequence must be greater than or equal to the number of antennas. For a massive MIMO system, $N$ is large, making conventional algorithm infeasible. It takes a long training period to complete channel estimation, which makes it impossible to complete channel estimation

_Springer

within the channel coherence time. In addition, the terminal needs to feed back the channel state information to the base station, which also requires the feedback load proportional to the dimension of the channel matrix.

In order to complete the downlink channel estimation with limited training overhead, the channel estimation technology based on compressed sensing has attracted much attention. In the compressed sensing framework, as long as the original signal is sparse on some bases, fewer measurements ($T_d < N$) can be used to represent high-dimensional signals (in this paper, high-dimensional signals refer to Massive MIMO channels).

Assume there exists a matrix $\boldsymbol{D} \in \mathbb{C}^{N \times M} (N \geq M)$ such that $\boldsymbol{h} = \boldsymbol{D}\boldsymbol{\beta}$, where the sparse vector $\boldsymbol{\beta} \in \mathbb{C}^{M \times 1}$ is sparse, that is $S = \|\boldsymbol{\beta}\|_0 \leq N$, $\|n\|_2 \leq \varepsilon$. Therefore, the downlink channel estimation problem can be converted into the following mathematical problem:

$$\boldsymbol{y} = \sqrt{\rho}\boldsymbol{h}^T\boldsymbol{X} + \boldsymbol{n} = \sqrt{\rho}\boldsymbol{D}\boldsymbol{\beta}\boldsymbol{X} + \boldsymbol{n} \tag{4}$$

Provided $y$, $X$ and $D$, if we are able to solve for $\beta$, then the channel state information could be gained as $\widehat{h} = \boldsymbol{D}\boldsymbol{\beta}$. At that time, if the channel is sparse under certain conditions, the high-dimensional channel matrix can be recovered. Note the sparsity of the channel and assume that the channel estimation can be converted into the following mathematical problem:

$$\widehat{\boldsymbol{\beta}} = \arg\min\|\boldsymbol{\beta}\|_0, \text{s.t.}\left\|\boldsymbol{y}\text{-}\sqrt{\rho}\boldsymbol{D}\boldsymbol{\beta}\boldsymbol{X}\right\|_0 \leq \varepsilon \tag{5}$$

According to existing theories, when $T_d \geq cS\log(N/S)$ ($c$ is constant) is satisfied, $l_1$ norm is used to replace $l_0$ norm for relaxation solution, which can be recovered with high precision. Therefore, accurate channel estimation can be realized only by using training symbols proportional to channel sparsity, and the training period length is no longer required to be proportional to the number of base station antennas. However, (4) is an undetermined equation if we intend to employ a small number of training samples $T_d < N$. Thus, the system has an infinite number of solutions for $\beta$, which must be solved by sparse constraints. Therefore, it is necessary to study the constraint conditions of minimum sparsity. The mmWave channel could be characterized in matrix form by the following model

$$\mathbf{H} = \sum_{l=1}^{L} \alpha_l \boldsymbol{\alpha}_{BS}(\theta_l)\boldsymbol{\alpha}_{MS}^H(\phi_l) \tag{6}$$

where $L$ is the number of paths; $\alpha_l$ is the complex gain associated with the $l$th path; $\theta_l \in [0, 2\pi]$ and $\phi_l \in [0, 2\pi]$ are the associated azimuth AoA and azimuth AoD, respectively; and $\boldsymbol{\alpha}_{BS} \in \mathbb{C}^N$, $\boldsymbol{\alpha}_{MS} \in \mathbb{C}^M$ are the array response vector associated with the BS and mobile station, respectively. Assume a uniform linear (ULA) antenna array is employed. Thus, the steering vectors at the BS and the MS can be expressed as

$$\boldsymbol{\alpha}_{BS}(\theta_l) = \frac{1}{\sqrt{N}}\left[1, e^{j\frac{2\pi}{\lambda}dsin(\theta_l)}, \cdots, e^{j(N-1)\frac{2\pi}{\lambda}dsin(\theta_l)}\right]^T$$
$$\boldsymbol{\alpha}_{MS}(\phi_l) = \frac{1}{\sqrt{M}}\left[1, e^{j\frac{2\pi}{\lambda}dsin(\phi_l)}, \cdots, e^{j(M-1)\frac{2\pi}{\lambda}dsin(\phi_l)}\right]^T \tag{7}$$

where $\lambda$ is the signal wavelength and $d$ is the distance between adjacent antenna elements.

# 3 Block-sparse and low-rank features of mmWave channels

## 3.1 Block-sparse property

To formulate the channel estimation as a sparse signal recovery problem, we first express the channel as a beam space MIMO representation as follows

$$\boldsymbol{H} = \boldsymbol{A}_{BS}\boldsymbol{H}_\nu\boldsymbol{A}_{MS}^H \tag{8}$$

where $\boldsymbol{A}_{BS} \triangleq [\boldsymbol{\alpha}_{BS}(\psi_1), \cdots, \boldsymbol{\alpha}_{BS}(\psi_{N1})]$ is an over-complete matrix ($N_1 \geq N$) with each column a steering vector parameterized by a pre-discretized AoA, $\boldsymbol{A}_{MS} \triangleq [\boldsymbol{\alpha}_{MS}(\omega_1), \cdots, \boldsymbol{\alpha}_{MS}(\omega_{N2})]$ is an over-complete matrix ($N_2 \geq M$) with each column a steering vector parameterized by a pre-discretized AoD, and $\boldsymbol{H}_\nu \in \mathbb{C}^{N_1 \times N_2}$ is a sparse matrix with L non-zero entries corresponding to the channel path gains $\{\alpha_l\}$.

By substituting (7) into (3), it yields $\boldsymbol{y} = \sqrt{\rho}\boldsymbol{h}^T\boldsymbol{X} + \boldsymbol{n} = \sqrt{\rho}\boldsymbol{D}\boldsymbol{\beta}\boldsymbol{X} + \boldsymbol{n}$

$$\begin{aligned}\boldsymbol{y} &= \sqrt{\rho}A_{BS}H_\nu A_{MS}^H\boldsymbol{X} + \mathbf{N} \\ &= \left[\left(A_{MS}^H\sqrt{\rho}\right)^T \otimes (\mathbf{X}A_{BS})\right]\widetilde{h} + \mathbf{N} \\ &= \left((\boldsymbol{X})\left(A_{MS}^* \otimes A_{BS}\right)\right)\widetilde{h} + \mathbf{N}\end{aligned} \tag{9}$$

where $\otimes$ denotes the Kronecker product, $(\ )^*$ is the complex conjugate, and $\widetilde{h}\underline{\triangle}vec(\mathbf{Hv})$. Collecting all measurements $\{y(t)\}$ and staking them into a vector $y\underline{\triangle}\left[y_1 \cdots y_{T_d}\right]^T$, we arrive at

$$\begin{aligned}y &= \begin{bmatrix}(X_1) \\ \vdots \\ (X_T)\end{bmatrix}\left(A_{MS}^* \otimes A_{BS}\right)\widetilde{h} + n \\ &\underline{\triangle}\psi\widetilde{h} + n\end{aligned} \tag{10}$$

Since $\widetilde{h}$ is sparse, channel estimation now has become a sparse signal recovery problem. To estimate $\widetilde{h}$, we can resort to corresponding algorithms referring to sparse signal recovery.

## 3.2 Low-rank property

Due to the spatial correlation and the unsymmetric angular spreads over different domains, mmWave channels may exhibit a meaningful low-rank structure that can be utilized to

improve the sample complexity. Such a low-rank structure can be observed from recent real-world mmWave channel measurements in [20].

## 4 Overall framework of our proposed algorithm

The overall algorithm framework of this paper is shown in Fig. 1. Our proposed algorithm is divided into three stages:

(1)  Preprocessing: in terms of joint low-rank and sparse characteristics of massive MIMO channels, a mathematical model for channel state acquisition is established and a non-convex weighted low-rank approximation method is used to surrogate the traditional nuclear norm low-rank approximation model.
(2)  Construction of incomplete learning dictionary: obtain channel measurements as training data set from a specific cell, convert the training data set from one-dimensional data to two-dimensional images, construct channel response samples, and construct the deep neural network to identify the transmission features of the mmWave channels.
(3)  Block-sparse compressed sensing [20] method is utilized to obtain the channel state information and complete high-precision estimation of channel state information in FDD downlink massive MIMO.

## 5 Channel estimation algorithm combining low-rank approximation and dictionary learning

In this section, we divide our proposed framework into two separate stages. In the first stage, we investigate the received downlink pilot signals in (2) as a low-rank matrix completion process. In the second stage, we measure channel responses in a specific cell and construct an incomplete learning dictionary such that the dictionary adapts well to the key features of mmWave channels.

### 5.1 Low-rank approximation based on non-convex theory

The received model in (2) can be formulated as sampling from a low-rank matrix:

$$Y_{ij} \triangleq \left(\sqrt{\rho} H^T X\right)_{ij} \tag{11}$$

where $Y \triangleq \sqrt{\rho} H^T X$ is a low-rank matrix with rankrank$(Y)$ = L and $Y_{ij}$ represents the $(i,j)$th element of matrix $Y$. Since mmWave channels exhibit low-rank structure, the matrix $Y$ satisfies the low-rank property as well [19]. Therefore, under the joint conditions of low-rank and sparse constraints, the channel state acquisition model can be expressed as the following model:

$$\underset{\widetilde{h},\beta}{\operatorname{argmin}} \|Y\|_* + \lambda\|\beta\|_1 + \mu \left\|Y - \psi\widetilde{h}\right\|_F^2, subject\ to\ Hv = D\beta \tag{12}$$

**Fig. 1** Block diagram of the proposed method
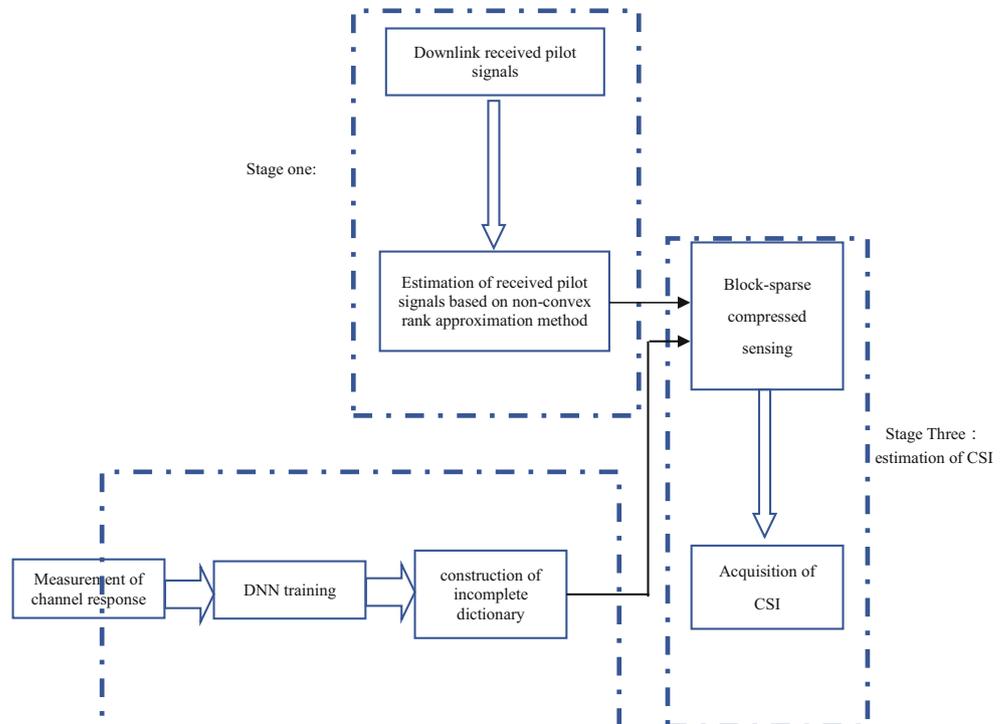
The first term of formula (12) is the nuclear norm approximation of the low-rank matrix; i.e., it represents the singular value of the matrix. The second term is $l_1$ norm of sparse components, which aims to constrain sparsity. The third term is Frobenius norm (defined as $\|A\|_F^2 = \sqrt{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{i,j}|^2} = \sqrt{tr(A^H A)}$ ), which is used to ensure consistent continuity of observed signals. Equation (12) is regarded as a convex optimization problem and can be effectively solved by various methods. However, its solution is always suboptimal because the nuclear norm is only a loose approximation of the rank function. Regarding norm relaxation function, nonconvex penalty function outperforms traditional norm relaxation in accuracy of solution. The nonconvex penalty function is defined as $f(t) = \eta(|t|)$, where $\eta(t)$ is a monotonically non-decreasing nonconvex function in the real field $\mathbb{R}$ and $\eta(0) = 0$.

The rank of the matrix is equal to the norm of the singular value vector of the matrix, and the nuclear norm of the matrix is equal to the norm of the singular value vector of the matrix, so the nonconvex penalty function of the vector is naturally extended to the singular value vector of the matrix, thus obtaining the nonconvex weighted kernel norm of the matrix.

*Definition*: The nonconvex weighted nuclear norm of any matrix $X$ is defined as the sum of singular values of the matrix after nonconvex function mapping; that is, the nonconvex weighted kernel norm of matrix $X$ is defined as $|X|_{mw} = \sum\limits_{i=1}^{n} g(\sigma_i^X)$, where $n \leq m$ and $\sigma_i^X$ is $i$th singular value of matrix $X$. $g(t)$ is a non-convex penalty function acting on the singular value of the matrix, so the real function $g(t)$ only needs to satisfy two conditions: $g(0) = 0$ and $g(t)$ is a non-monotonically decreasing non-convex function in interval [0,1].

Based on the above theory, in this section we propose a non-convex algorithm for channel low-rank function approximation, thus converting the downlink channel estimation of Eq. (12) into the solution of the following function:

$$\min_{\widetilde{h},\beta} \alpha \sum_i g(\sigma_i(Y)) + \lambda\|\beta\|_1 + \mu\left\|Y - \psi\widetilde{h}\right\|_F^2 \qquad (13)$$

It can been seen from Eq. (13) that the nuclear norm in Eq. (12) is surrogated by a non-convex function as an approximation of the low-rank matrix and that the third term is a loss function to constrain the estimation error. Eq. (13) includes continuous variables together with discrete variables, combined with nonconvex functions. However, the solution to the weighted nuclear norm low-rank approximation of non-convex function is a tricky as it cannot be directly solved. Thus, we utilize the Alternating Direction Method of Multipliers (ADMM) to address the problem.

In terms of ADMM algorithm, formula (13) is obtained by fixing one variable while solving another variable. Accordingly,

when the dictionary matrix $D$ is fixed, the acquisition in Eq. (13) can be regarded as a matrix completion problem. Therefore, the low-rank approximation of the received signals $Y$ can be converted into the following equation:

$$\min_Y \alpha \sum_i g(\sigma_i(Y)) + \mu\left\|Y - \psi\widetilde{h}\right\|_F^2 \qquad (14)$$

The function $g(t)$ is a non-convex on the interval and satisfies the following formula:

$$g(\sigma_i^k) + \omega_i^k(\sigma_i - \sigma_i^k) \geq g(\sigma_i) \qquad (15)$$

The $k$th iteration of matrix $Y$ is denoted as $Y^k$, and $\sigma_i^k$ is $i$th singular value of $Y^k$ in the $k$th iteration, denoted as $\sigma_i^k = \sigma_i(Y^k)$. Since $g(t)$ is a nonconvex function, it is difficult to directly solve Eq. (14). Equation (15) gives a linear approximation of non-convex functions, so Eq. (14) can be iteratively solved employing the following methods:

$$
\begin{aligned}
Y^{k+1} &= \arg\min_{\widetilde{h}} \alpha \sum_i \left[g(\sigma_i^k) + \omega_i^k(\sigma_i - \sigma_i^k)\right] + \frac{\mu}{2}\left\|Y^k - \psi\widetilde{h}\right\|_F^2 \\
&= \arg\min_{\widetilde{h}} \frac{\mu}{2}\left\|Y^k - \psi\widetilde{h}\right\|_F^2 + \alpha \sum_i \omega_i^k \sigma_i^k
\end{aligned}
$$
$$(16)$$

where $\omega_i^k$ is the sub-gradient of $g(t)$ acting on the interval $[0, +\infty)$. To solve Eq. (16) more easily, we utilize first-order Taylor expansion for function (16). Thus, the first-order Taylor expansion at the matrix $K \in \mathbf{R}^{M \times N}$ is denoted as

$$
\begin{aligned}
\Gamma(Y) &\approx \Gamma(K) + \langle\nabla\Gamma(K), Y - K\rangle + \frac{\mu}{2}\|Y - K\|_F^2 \\
&= \langle 2(Y - XDK - XD\beta), Y - K\rangle + \frac{\mu}{2}\|Y - K\|_F^2
\end{aligned}
\qquad (17)
$$

Therefore, the iterative function (16) can be converted to

$$
\begin{aligned}
Y^{k+1} = \arg\min_{\widetilde{h}} \alpha \sum_i \omega_i^k \sigma_i(Y^k) \\
+ \frac{\mu}{2}\left\|Y^k - \left(Y^k + \frac{1}{\mu}\left(Y^k - \Psi\widetilde{h}\right)\right)\right\|_F^2
\end{aligned}
\qquad (18)
$$

Equation (18) remains still a nonconvex optimization problem, but according to the following theorem, there is an analytical solution to this problem.

As was proved in Menglu [14]: for any arbitrary $\lambda > 0$ and $Y \in \mathbb{R}^{m \times n}$, if $0 \leq w_1 \leq w_2 \cdots \leq w_n$, then the optimal solution to the minimization problem $\min_X \lambda \sum\limits_{i=1}^{n} w_i \sigma_i^Y + \frac{1}{2}\|X - Y\|_F^2$ can be obtained by weighted singular value operators, denoted as $X^* = US_{\lambda w}(\Sigma)V^T$, where $Y = U\Sigma V^T$ is the singular value decomposition of matrix $Y$ and the singular values are expressed as $\sigma_1^Y \geq \sigma_2^Y \geq \cdots \sigma_n^Y \geq 0$. $S_{\lambda w}(\Sigma)$ is the diagonal matrix with

diagonal elements $(\Sigma_{ii} - \lambda w_i)_+$, $i \in \{1, 2, \cdots, n\}$, where $(t-a)_+ = \{t-a \ t > a \ \ 0 \ \ t \leq a$. Based on the above derivation, we develop Algorithm 1 to separate and extract low-rank received signals matrix $Y$, which is shown in Table 1.

## 5.2 Channel measurements extraction utilizing DNN

Under normal circumstances, over-complete DFT dictionary is used to achieve sparse channel recovery. However, the pre-designed fixed dictionary cannot adapt to the channel structure information according to the cell propagation environment. In order to reduce the system complexity in the dictionary training stage, a better method is to design an incomplete dictionary that is specific to the transmission environment and suitable for channel characteristics. Ding Yacong et al. proposed a channel estimation method based on learning dictionary for channel thinning recovery in Liu et al. [17]. Different from the pre-designed dictionary method, the proposed method learns a dictionary from channel measurements. In the learning process, the sparse representation is optimized, and the sparse vector of the channel is extracted through sparse constraints. Channel characteristics at different locations in the cell are measured indefinitely, and channel measurement vectors are trained to obtain an adaptive over-complete dictionary. However, this scheme carries out channel measurements randomly and indefinitely, thus lacking selection of typical positions of channels, which may lose channel state information indicating important structural features of the channels, decrease the accuracy of dictionary training, and increase the redundancy of channel measurements.

Therefore, we adopt the DNN-based channel measurement method to classify and measure the key features of downlink mmWave channels. On the one hand, the measurement load of the system is reduced. On the other hand, this method can fully extract most channel structure characteristics, thus constructing an incomplete dictionary suitable for link characteristics. This part corresponds to the third stage of the overall framework of the algorithm. The specific implementation process is shown in Fig. 2.

**Table 1** Downlink received measured signal extraction algorithm based on non-convex low-rank approximation theory and alternated direction (ALM) method

---

Algorithm 1:Estimation of downlink received pilot signals based on non-convex low-rank approximation

Input:the measurements $Y$, the matrices $A_{BS}$, and $A_{MS}$, pilot sequences $X$.
output: $Y^* = Y^{k+1}$

Initialization :iterative times $k = 0$, $\lambda$, $w_i^k = 0$ $i = 1, 2, \cdots, n$ 。

while not converge
Solve Eq. (18) to obtain the solution of $Y^{k+1}$;

$\omega_i^{k+1} = \partial g(\sigma_i(Y^{k+1}))$, calculate the sub-gradient of $g(t)$, where
$t = \sigma_i(Y^{k+1})$, $\sigma_1(Y^{k+1}) \geq \cdots \geq \sigma_n(Y^{k+1})$,
Convergence condition: $\|Y - XD\beta\|_F^2 \leq \varepsilon$, $\varepsilon = 10^{-5}$
end while

---

We classify path loss and other features utilizing DNN. Firstly, we set non–line of sight (NLOS) and line of sight (LOS) path propagation scenes in the actual transmission environment. Secondly, we use the DNN to classify channel measurement vectors from the view of path loss, multipath component energy, delay, phase and angle domains, and other characteristics. During training, links with similar parameters are grouped into the same group and group-based channel measurements are obtained.

Sample the channel measurements to obtain the following expression:

$$h(mT_S) = \sum_{l=0}^{L} \sum_{k=0}^{K} \alpha_{k,l}\delta(mT_S - T_l - \tau_{k,l}), m = 1, 2, \cdots, N \quad (19)$$

where $T_S$ is the sampling interval. We make Modulus Operation on the sampled signals to obtain

$$g_m = |h(mT_S)|, m = 1, 2, \cdots, N \quad (20)$$

With the evolution of time, the signal $g_m$ tends to approach zero. Therefore, only $M^2(M^2 < N)$ sample signals are required to construct the input set $X$ of the DNN:

$$X = \begin{bmatrix} g_1 & g_2 & \cdots & g_M \\ g_{M+1} & g_{M+2} & \cdots & g_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ g_{M(M+1)+1} & g_{M(M+1)+2} & \cdots & g_{M^2} \end{bmatrix} \quad (21)$$

DNN includes three hidden layers, each with 1024 hidden cells and 64 output cells. Channel measurement matrix x under different transmission environments is obtained from the input of training samples. The dimension of the input unit is equal to the dimension of the sparse feature vector. In this paper, the depth neural network adopts feedforward structure, the input layer is 1845-dimensional feature vector, the output layer has classification function, and the channel feature vector forms an incomplete dictionary. In order to improve the convergence speed of network algorithm, cross entropy loss function and Sigmoid are used as activation functions of hidden layer elements. Hiding layer initialization adopts He initialization method. In addition, the depth neural network covers about 4M adjustable parameters. This paper uses statistical gradient descent and NAG method to optimize the parameters. It is assumed that the output layer result is quite different from the preset result; i.e., the hidden layer feeds back to the input layer to modify the weight of each unit. The gradient is calculated by backward propagation based on the minimum mean square error function; that is, the minimum mean square error between the channel measurement samples and the estimated channel estimation matrix is taken as the loss function. The basic objective function is defined as

$$J_{MSE}(W, b) = \frac{1}{M} \sum_{m=1}^{M} \|\hat{X}_m(U, W, b) - X_m\|_2^2 \quad (22)$$

**Fig. 2** Incomplete dictionary based on DNN



input layer      hidden layer      output layer
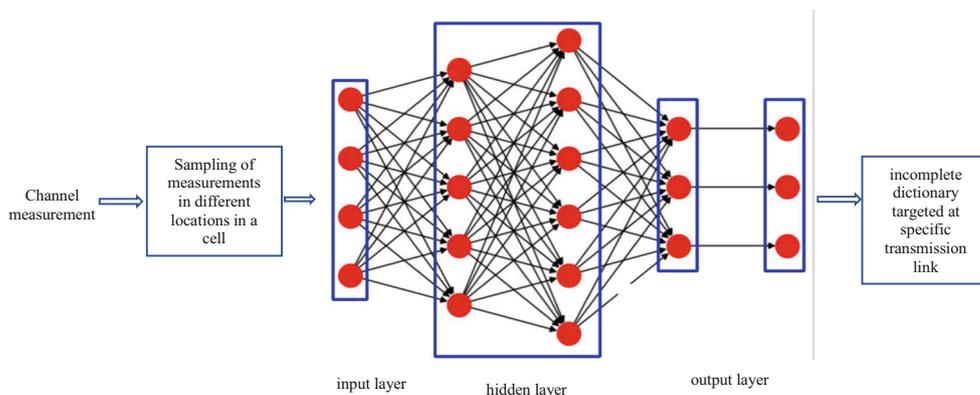
where $W$ and $b$ represent the weight and the basis matrix to be trained for the estimated channel state information. $M$ is the dimension of the training matrix, and $U$ is the input feature vector. The objective function $J_{MSE}(W, b)$ aims to minimize the mean square error between the estimated channel estimate and the channel measurement. The three hidden layers use 20% dropout to suppress over-fitting images in the training process. The wireless channel measurement matrix is taken as the input signal of the deep learning network. The output of the $l$th hidden layer is

$$G_j^l = f\left(\sum_{i=1}^{J} \omega_{ij}^{l-1} H_i^{l-1}\right), j = 1, 2, \cdots, J, l = 1, 2, \cdots L \quad (23)$$

in which $\omega_{ij}^{l-1}$ is the weights of the $i$th node in the $(l-1)^{\text{th}}$ layer and the $j$th node and $H_i^{l-1}$ is the output of the $i$th node in $(l-1)$th layer. $J$ is the number of nodes per layer. $f(\cdot)$ is the activation function of the hidden layer. We obtain the following parameters using gradient descent method to obtain

$$\Delta\left(W_{m+1}^k, b_{m+1}^k\right) = -\lambda \frac{\partial J}{\partial\left(W_{m+1}^k, b_{m+1}^k\right)}$$
$$+ \omega\Delta\left(W_m^k, b_m^k\right), 1 \leq k \leq K + 1 \quad (24)$$

in which $\lambda$ is the learning rate, momentum $\omega$ is used to accelerate the convergence rate of the algorithm, and $K$ is the number of hidden layers. The process of DNN-based incomplete dictionary learning is summarized as follows:

Step 1. Construct the measurement data set to be classified. Equation (19) represents a massive MIMO channel model.
Step 2. DNN training algorithm is mainly divided into two stages: forward feedback and reverse propagation.
Step 3. At the end of training, the trained DNN performs feature extraction on a massive MIMO channel to identify channel features. The dictionary learning

algorithm proposed in this paper is shown in Table 2.

After obtaining the dictionary, the channel matrix is sparsely recovered in the angular domain. The sparse recovery process is implemented employing the block-sparse compressed sensing method.

$$:\hat{h}$$

$$\text{method: } \blacksquare Y^{k+1} = \arg\min_\beta \lambda\|\beta\|_1 + \frac{\mu}{2}\|Y - \left(Y^k + \frac{1}{\mu}(Y - XD\beta)\right)\|_F^2$$

$$H_\nu = \hat{D}\hat{\beta}$$

## 6 Simulation results

We now carry out simulation results to illustrate the performance of our proposed non-convex dictionary learning method. We compare our method with the following algorithms: Dictionary Learning based Channel Model (DLCM) [2], Joint Sparse and Low-rank Bayesian Learning(SLAB) [15], and compressed sensing–based sparse channel estimation (CSSCE).

We consider a scenario where the BS and the MS employ a uniform linear array with $N = 128$, $M = 1$ antenna. The distance between neighboring antenna elements is assumed to be

**Table 2** Incomplete dictionary learning algorithm

Algorithm 2:Incomplete dictionary learning

Input:sparsity of the mmWave channel
Output: incomplete dictionary $D$

Step 1: Measure the links from BS to MS to extract measurements of downlink vectors $h_i$, $i = 1, 2, \cdots, L$;

Step 2: Take the measurements into the DNN to perform Eqs. (22), 23, and (24);

Step 3: Construct the incomplete dictionary base on the output of DNN.

half the wavelength of the signal. The mmWave channel is assumed to follow the geometric channel model with $L = 6$ clusters. The mean AoAs and AoDs for these six clusters are set to $\theta_1 = \phi_1 = \frac{\pi}{6}, \theta_2 = \phi_2 = \frac{\pi}{3}, \theta_3 = \phi_3 = \frac{\pi}{2}$ , $\theta_4 = \phi_4 = -\frac{\pi}{2}, \theta_5 = \phi_5 = -\frac{\pi}{3}, \theta_6 = \phi_6 = -\frac{\pi}{6}$. The relative AoA and AoD shifts are uniformly generated within the angular spreads $(\theta_l - \delta_\theta/2, \theta_l + \delta_\theta/2)$, $(\phi_l - \delta_\phi/2, \phi_l + \delta_\phi/2)$, Suppose the base station is equipped with a uniform rectangular array of antennas and the terminal is equipped with a single antenna. The wireless channel model uses NLOS (non–line of sight) UMI-Street Canyon (Urban Microcellular Channel) scenario with a carrier frequency of 28 GHz.

The performance is evaluated via one metric, namely, the normalized mean square error (NMSE). The NMSE is calculated as

$$NMSE = E\left[\frac{\left\|\hat{\boldsymbol{H}}-\boldsymbol{H}\right\|_F^2}{\|\boldsymbol{H}\|_F^2}\right] \qquad (25)$$

where $\hat{\boldsymbol{H}}$ denotes the estimate of the true channel $\boldsymbol{H}$.

Figure 3 a and b illustrates the NMSE performance versus channel matrix rank for all algorithms with SNR = 5 dB and 15 dB. As can be seen from the figure, the DLCM and CSSCE comparison algorithms are not sensitive to the change of matrix rank, and compared with SLAB, the performance of our proposed algorithm has been significantly improved in the case of low rank. This is because the algorithms DLCM and CSSCE do not consider the low-rank performance of the channel, and for non-convex methods, their algorithms are superior to SLAB algorithms using kernel norm approximation. Although the channel has a low rank in MIMO scenarios, traditional algorithms do not make full use of these features. The algorithm in this paper makes full use of the low-rank characteristics of the channel. In Fig. 3a, when the channel

matrix rank is less than 15, the performance of CSSCE and SLAB algorithms is not as good as that of this algorithm. This is because from the perspective of least squares, CSSCE has the best performance, but the algorithm is sensitive to noise and has poor performance under low signal-to-noise ratio, while SLAB algorithm is based on Bayesian posterior mean, so its anti-noise robustness is better than the least squares algorithm. Therefore, SLAB algorithm is superior to CSSCE when SNR is low and the rank of matrix is large. When the bit energy signal-to-noise ratio EbNo is increased to 15 dB, the result is similar to Fig. 3b.

Figure 4 is a graph of normalized mean square error performance versus EbNo for four algorithms. The algorithm proposed in this paper outperforms DLCM, SLBF, and CSSCE for the following reasons:

(1) DLCM algorithm carries out random large-scale measurements on the channels in the cell. Thus, DLCM cannot accurately capture all link characteristics and has large data storage load. The method proposed in this paper classifies the channel scenes and extracts the key features of the measured values according to the channel categories to form an incomplete dictionary. Therefore, compared with DLCM method, our proposed algorithm can obtain higher channel estimation accuracy; that is, the normalized mean square error of this algorithm is lower than DLCM algorithm under the same bit energy signal-to-noise ratio.

(2) When SLBF performs channel low-rank approximation, nuclear norm method was adopted. However, this method is not optimal. The nonconvex approximation algorithm used in this paper is an optimal solution for low-rank approximation. Therefore, compared with SLBF, the main features of the channel can be extracted more accurately.

Fig. 3 a Performance comparison of channel rank and NMSE (EbNo = 5 dB). b Performance comparison of channel rank and NMSE (EbNo = 15 dB)
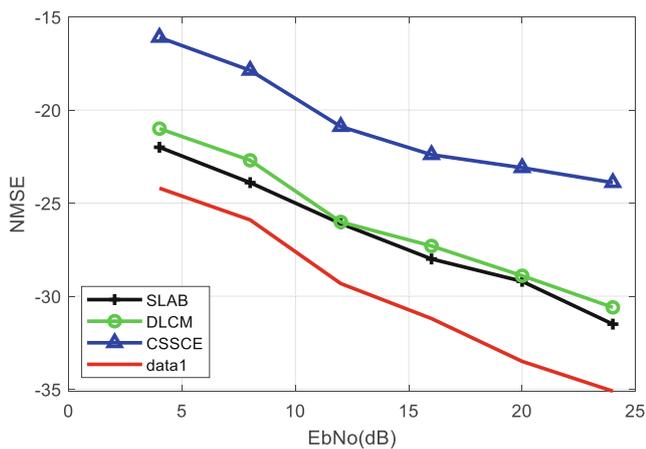


(a) performance comparison of channel rank and NMSE(EbNo=5dB)

(b) performance comparison of channel rank and NMSE(EbNo=15dB)
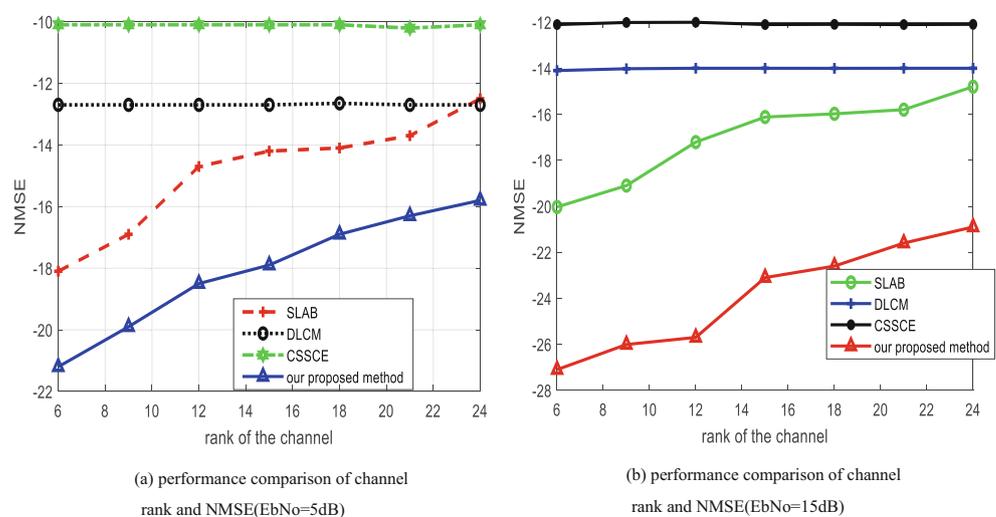
**Fig. 4** Performance comparison of EbNo and NMSE

(3) CSSCE algorithm only mines the sparse characteristics of the channel and ignores the low-rank characteristics of the channel angle domain. Therefore, compared with the algorithm proposed in this paper, it cannot accurately mine the potential internal structural characteristics of the channel.

Figure 5 shows the NMSE performance comparison of the four algorithms on the base station for different numbers of pilots. In our experiment, EbNo = 10 dB is set. The normalized mean square error decreases as the number of training pilots increases. Compared with the other two methods, SLAB and the algorithm proposed in this paper reduce the number of downlink training pilots. This is because SLAB and this paper make use of the low-rank characteristics of the channel, which further restricts the sparse performance of the channel and reduces the effective dimension and training load of the downlink channel compared with DLCM and CSSCE. Since the optimal solution can be obtained through nonconvex low-rank approximation, the algorithm in this
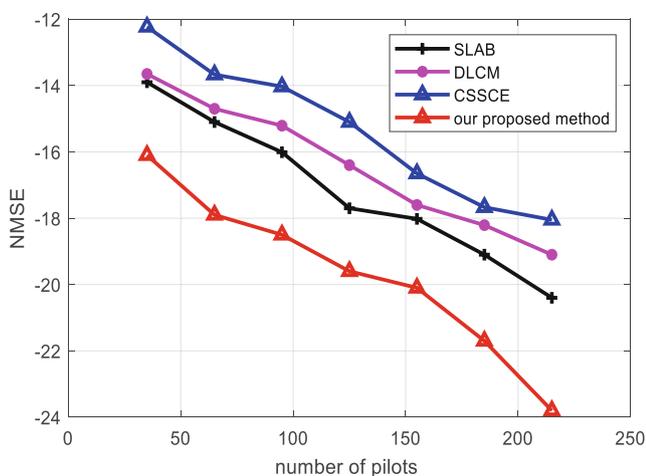
paper can perform low-rank approximation better. Therefore, under the same NMSE performance, the number of training sequences required in this paper is smaller than SLAB algorithm.

As can be seen from Fig. 4, to achieve the same NMSE performance, the algorithm proposed in this paper requires the lowest downlink pilot sequence length. Theoretically, the longer the pilot sequence is, the longer the channel estimation takes, and the more accurate the channel state information is. However, at the same time, it needs to occupy more frequency band resources to transmit pilot signals, resulting in lower frequency band utilization rate. Therefore, the frequency band utilization rate of the algorithm in this paper is higher than that of the other three methods, and accurate channel estimation results can be obtained by using relatively few pilot sequences.

# 7 Conclusion

For massive MIMO FDD systems, this paper proposes an estimation channel estimation framework that exploits the joint low rank and sparse characteristics of channels. In the proposed scheme, the base station sends a downlink pilot sequence to the mobile station. The mobile station quantizes the received downlink pilot signal, then feeds back the uplink to the base station and performs a channel estimation algorithm at the base station. The innovation lies in using non-convex algorithm instead of traditional nuclear norm to approximate the rank of wireless channel, and learning sparse coefficient through non-preset dictionary. In this method, the construction of dictionary is independent of the transmission signal. The channel acquisition matrix is obtained by measuring multiple channel impulse response experiments. Then the key characteristics of the channel are extracted by DNN to form an incomplete dictionary, and the channel state is obtained by sparse representation. Simulation experiments verify the superiority of the proposed method from NMSE and pilot number.

**Fig. 5** Performance of number of pilots and NMSE

# References

1. Feng C, Jing Y, Jin S (2016) Interference and outage probability analysis for massive MIMO downlink with MF precoding. IEEE Signal Proc Lett (S1070-9908) 23(3):366–370

2. Ding Y, Rao BD (2018) Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems. IEEE Trans Wireless Commun (S1536-1276) 17(8):5437–5451

3. Hirantha BA, Abeysekera S, Matsuda T (2008) Dynamic contention window control mechanism to achieve fairness between uplink and downlink flows in IEEE 802.11 Wireless LANs. IEEE Trans Wireless Commun (S1536-1276) 7(9):3517–3525

4. Li Q, Lin XE, Zhang J (2009) Advancement of MIMO technology in WiMAX: from IEEE 802.16d/e/j to 802.16m. IEEE Commun Mag (S0163-6804) 47(6):100–107

5. Ding Z, Liu Y, Choi J (2017) Application of non-orthogonal multiple access in LTE and 5G networks. IEEE Communications Mag(S2169-3536) 55(2):185–191

6. El-Khamy SE, Moussa KH (2016) Millimeter multi-user massive MIMO simplified using antenna selection and one-bit quantized channel estimation:// URSI Asia-Pacific Radio Science Conference (URSI AP-RASC). IEEE, Seoul, pp 769–772

7. Rao X, Lau VKN (2014) Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. IEEE Transactions on Signal Proc(S1053-587X) 62(12):3261–3271

8. Vlachos E, Alexandropoulos GC, Thompson J (2018) Masive MIMO channel estimation for millimeter wave systems via matrix completion. IEEE Signal Proc Lett 25(11):1675–1679

9. Yang Y, Gao F, Ma X, Zhang S (2019) Deep learning-based channel estimation for doubly selective fading channels. IEEE Acess 7: 36579–36589

10. Gao Z, Dai L, Dai W, Shim B, Wang Z (2016) Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO. IEEE Trans Commun(S0090-6778) 64(2): 601–617

11. Fang J, Li X, Li H (2017) Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems. IEEE Trans Wireless Commun (S1536-1276) 16(3):1935–1947

12. Sun H, Zhao Z, Xiao F (2018) Limited feedback double directional massive MIMO channel estimation: from low-rank modeling to deep learning// IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, Kalamata, pp 1948–3252

13. Wei J (2019) Channel estimation based on matrix completion for massive MIMO systems. J Nanjing Univ Posts Telecommun (Natural Science Edition)(S1673-5439) 39(5):20–25

14. Menglu S (2018) Channel Estimation based on low-rank matrix completion for Massive MIMO systems. Appl Res Comp(S1001-3695) 35(6):1841–1844

15. Sun H. (2018) Limited Feedback Double Directional Massive MIMO Channel estimation: from low-rank modeling to deep learning// IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece: IEEE. 1948-3252.

16. Lu C, Tang J (2016) Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm. IEEE Trans Image Proc(S1057-7149) 25(2):829–839

17. Liu K, Li X, Fang J, Li H (2019) Bayesian mmWave channel estimation via exploiting joint sparse and low-rank structures. IEEE Access(S2169-3536) 7:48961–48970

18. He X, Song R, Zhu W-P (2016) Pilot Allocation for Distributed-compressed sensing Based Sparse channel estimation in MIMO-OFDM Systems. IEEE Transactions Vehicular Technol(S0018-9545) 65(10):2990–3004

19. Li X, Fang J (2018) Millimeter wave channel estimation via exploiting joint sparse and low-rank structures. IEEE Trans Wireless Commun(S1536-1276) 17(2):1123–1133

20. Li L, Fu T, Wang X, Huang M, Wan L, Yang Y (2020) DOA estimation of strictly noncircular sources in wireless sensor array network via block sparse representation, in IEEE Access 8:47500-47508. https://doi.org/10.1109/ACCESS.2020.2979055