

Joint Low-Rank and Sparse Bayesian Channel Estimation for Ultra-Massive MIMO Communications

Jiangnan Ji, Cheng-Xiang Wang[✉], *Fellow, IEEE*, Shuaifei Chen[✉], *Member, IEEE*, Chen Huang[✉], *Member, IEEE*, Xiping Wu[✉], *Senior Member, IEEE*, and Emil Björnson[✉], *Fellow, IEEE*

Abstract—This letter investigates channel estimation for ultra-massive multiple-input multiple-output (MIMO) communications. We propose a joint low-rank and sparse Bayesian estimation (LRSBE) algorithm for spatial non-stationary ultra-massive channels by exploiting the low-rankness and sparsity in the beam domain. Specifically, the channel estimation integrates sparse Bayesian learning and soft-threshold gradient descent within the expectation-maximization framework. Simulation results show that the proposed algorithm significantly outperforms the state-of-the-art alternatives under different signal-to-noise ratio conditions in terms of estimation accuracy and overall complexity.

Index Terms—Ultra-massive MIMO, channel estimation, beam domain, Bayesian learning, low-rank, sparse optimization.

I. INTRODUCTION

EMPLOYING a large number of antennas significantly improves the spectral and energy efficiencies of sixth-generation (6G) wireless communications with enhanced beamforming gain and spatial interference suppression capability [1], [2]. However, having such large channel dimensions poses a critical channel estimation challenge. This issue is further aggravated by the spatial non-stationary phenomenon of ultra-massive multiple-input multiple-output (MIMO) channels and becomes more prominent in the near-field region [3]. Conventional least squares and minimum mean-square-error (MSE) algorithms are ill-suited for such scenarios, as they fail to exploit channel's structural properties, i.e., low-rankness and sparsity [4]. Consequently, these methods cannot achieve

a satisfactory trade-off between estimation accuracy and computational complexity.

Compressive sensing (CS)-based channel estimation is well-attended by leveraging the channel sparsity and recovering the channel from limited measurements through a small number of samples. CS algorithms can be classified into three categories: convex optimization [5], greedy [6], and Bayesian algorithms [7], [8], emphasizing estimation accuracy, efficiency, and robustness, respectively. The low-rank matrix recovery-based channel estimation algorithms utilize the channel low-rankness, which means the channel matrix has a small rank. Among these, nuclear norm minimization-based approaches lead to a satisfactory estimation solution by transforming the rank minimization problem into the convex nuclear norm minimization [9]. Tensor decomposition approaches capture structural information about the higher dimensions of the channel through tensor representations [10], [11]. However, estimation based on sole channel low-rankness or sparsity is suboptimal, and thus, a joint design is needed to further improve channel estimation performance.

Beam domain channel estimation is promising since the ultra-massive MIMO channel's low-rankness and sparsity are both observable in the beam domain, where signal energies are concentrated in specific beam directions due to the typically limited signal propagation paths [12]. In [13], a two-stage channel estimation algorithm was proposed by exploiting both low-rankness and sparsity but in a separate way. Though the method can improve the estimation accuracy to some extent, it failed to address the joint optimization, causing a possible error propagation across the two stages. To this end, [4] performed directions-of-arrival (DoA) tracking by Bayesian inference, assuming that stationary paths are low-rank and time-varying paths are sparse. Nevertheless, the modeling assumptions are relatively simple, and the adopted DoA tracking approach in the space domain cannot be directly applied to channel estimation due to the inhibited complexity.

To fill the research gaps, this letter proposes a high-accuracy joint low-rank and sparse Bayesian estimation (LRSBE) algorithm for ultra-massive MIMO channels. The main contributions are summarized as follows:

- We propose to decompose the beam domain channel into independent low-rank and sparse components based on the visibility across the antenna array.
- We formulate a joint low-rank and sparse recovery problem for channel estimation optimization and solve it through Bayesian learning within the expectation-maximization (EM) framework.
- The proposed algorithm is compared with state-of-the-art alternatives for validation of effectiveness and efficiency.

Received 12 October 2025; accepted 18 November 2025. Date of publication 26 November 2025; date of current version 17 December 2025. This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 62401643, the China Postdoctoral Science Foundation under Grant 2024M752439, the Jiangsu Major Project on Fundamental Researches under Grant BK20243059, and the Research Fund of National Mobile Communications Research Laboratory, Southeast University, under Grant 2025A05. E. Björnson was supported by the Grant 2022-04222 from the Swedish Research Council. The associate editor coordinating the review of this letter and approving it for publication was W. D. C. Freitas. (*Corresponding authors: Cheng-Xiang Wang; Shuaifei Chen.*)

Jiangnan Ji is with the National Mobile Communications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing 211189, China (e-mail: ji_jh@seu.edu.cn).

Cheng-Xiang Wang and Xiping Wu are with the National Mobile Communications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing 211189, China, and also with the Purple Mountain Laboratories, Nanjing 211111, China (e-mail: chxwang@seu.edu.cn; xiping.wu@seu.edu.cn).

Shuaifei Chen and Chen Huang are with the Purple Mountain Laboratories, Nanjing 211111, China, and also with the National Mobile Communications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing 211189, China (e-mail: shuaifeichen@seu.edu.cn; huangchen@pmlabs.com.cn).

Emil Björnson is with the Department of Computer Science, KTH Royal Institute of Technology, Kista, 164 40 Stockholm, Sweden (e-mail: emilbj@kth.se).

Digital Object Identifier 10.1109/LCOMM.2025.3637318

The rest of this letter is organized as follows. Section II introduces the beam domain channel model and analyzes its characteristics. In Section III, the proposed LRSBE algorithm is described in detail. Section IV provides simulation results and analysis. Finally, conclusions are drawn in Section V.

II. SYSTEM MODEL

Let us consider K single-antenna mobile terminals (MTs) and a base station (BS) equipped with a large uniform planar array comprising $M = M_h \times M_v$ half-wavelength-spaced antennas, where M_h and M_v are the numbers of antennas in the horizontal and vertical directions, respectively. We consider the block fading model and let $\mathbf{h}_k = \text{vec}(\mathbf{H}_k) \in \mathbb{C}^M$ denote beam domain channel transfer function between MT k and the BS in a specific coherence block, where $\text{vec}(\cdot)$ denotes vectorization. The beam domain channel matrix $\mathbf{H}_k \in \mathbb{C}^{M_h \times M_v}$ is given by

$$\mathbf{H}_k = \mathbf{U}_h^H \bar{\mathbf{H}}_k \mathbf{U}_v \quad (1)$$

where $\bar{\mathbf{H}}_k \in \mathbb{C}^{M_h \times M_v}$ is the corresponding space domain channel matrix resulting from the 6G pervasive channel model (6GPCM) [14], by which the spatial non-stationarity of ultra-massive MIMO can be characterized by dividing the clusters into partially visible and wholly visible parts [15]. More details will be elaborated later in this section. Matrices $\mathbf{U}_h \in \mathbb{C}^{M_h \times M_h}$ and $\mathbf{U}_v \in \mathbb{C}^{M_v \times M_v}$ are horizontal and vertical discrete Fourier transform-based (DFT) beamforming matrices, respectively.

The space domain ultra-massive MIMO channel matrix $\bar{\mathbf{H}}_k$ and its beam domain representative \mathbf{H}_k are both low-rank by having limited scattering clusters and strong spatial correlation, while only the latter is sparse due to the focusing effect of DFT in (1). Fig. 1 quantitatively exhibits the channel low-rankness and sparsity in both the beam and space domains by showing the distributions of the singular values of the channel matrices and the amplitude of channel elements. In Fig. 1(a), most of the information is carried by a few singular values, and the low-rankness is evident in both the beam and space domains by having 5 dominant singular values constituting 90% of the energy. As for the channel sparsity illustrated in Fig. 1(b), channel elements are non-negligible in the space domain, and the beam domain channels consist of a few dominant elements with the remaining elements very close to zero. According to the above observations, we can rewrite the beam domain channel \mathbf{H}_k as

$$\mathbf{H}_k = \mathbf{H}_k^L + \mathbf{H}_k^S \quad (2)$$

where $\mathbf{H}_k^L \in \mathbb{C}^{M_h \times M_v}$ and $\mathbf{H}_k^S \in \mathbb{C}^{M_h \times M_v}$ are low-rank and sparse components, respectively. One physical realization of (2) is the channel decomposition according to the visibility across a huge BS antenna array [15, Eq. (29)]. In this case, \mathbf{H}_k^L and \mathbf{H}_k^S coincide with the partially visible component and wholly visible component representing the clusters of \mathbf{H}_k seen by a subset of antennas and all antennas, respectively. This comes from the fact that the channel constructed by the wholly visible component consists of paths in the major few beam directions and thus exhibits a strong sparsity. The channel composed of the partially visible component, on the other hand, has a relatively slow amplitude change due to the presence of scatterers, and the main fluctuations are concentrated in a few columns of the channel matrix, which exhibits a strong low-rankness.

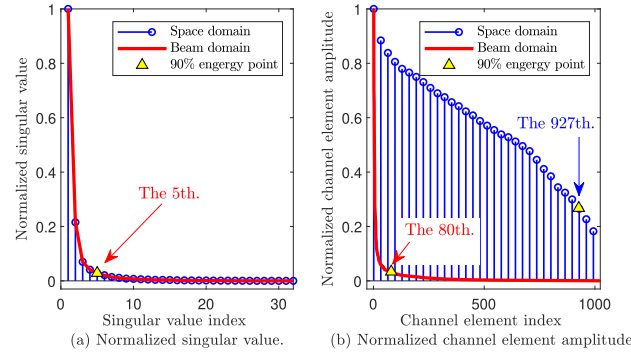


Fig. 1. Illustration of (a) low-rankness and (b) sparsity of the space domain and beam domain channels ($M = 1024$).

III. JOINT LOW-RANK AND SPARSE CHANNEL ESTIMATION

A. Joint Optimization Problem Formulation

Let $\mathbf{h} = [\mathbf{h}_1^T, \dots, \mathbf{h}_K^T]^T \in \mathbb{C}^{MK}$ denote the collective channel and $\mathbf{x}_k \in \mathbb{C}^N$ represent the pilot sequence transmitted the k th MT in the uplink. Here, $N \ll K$ is the number of orthogonal pilot sequences. It is worth noting that the so-called pilot contamination deteriorating the coherent transmission could occur in this case, due to the pilot reuse among the MTs. By leveraging the sparsity of the channel, a reduced number of pilots can be designed to alleviate the impact of pilot contamination, which will be further investigated in the future work. The received signal at BS is given by [7]

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\mathbf{h} + \mathbf{n} \\ &= \mathbf{A}(\mathbf{h}^S + \mathbf{h}^L) + \mathbf{n} \in \mathbb{C}^{MN} \end{aligned} \quad (3)$$

where $\mathbf{A} = [\mathbf{x}_1, \dots, \mathbf{x}_K] \otimes \mathbf{I}_M \in \mathbb{C}^{MN \times MK}$ is the prior measurement matrix, $\mathbf{h}^S = [(\mathbf{h}_1^S)^T, \dots, (\mathbf{h}_K^S)^T]^T$ with $\mathbf{h}_k^S = \text{vec}(\mathbf{H}_k^S)$, $\mathbf{h}^L = [(\mathbf{h}_1^L)^T, \dots, (\mathbf{h}_K^L)^T]^T$ with $\mathbf{h}_k^L = \text{vec}(\mathbf{H}_k^L)$, and $\mathbf{n} \sim \mathcal{N}_C(\mathbf{0}, \sigma^2 \mathbf{I}_{MN})$ is the receiver noise with power σ^2 .

The relationship between the prior distribution $P(\mathbf{h})$, the posterior distribution $P(\mathbf{h} | \mathbf{y})$, and the likelihood function $P(\mathbf{y} | \mathbf{h})$ of channel \mathbf{h} can be expressed as

$$P(\mathbf{h} | \mathbf{y})P(\mathbf{y}) = P(\mathbf{y} | \mathbf{h})P(\mathbf{h}) \quad (4)$$

according to Bayes' theorem, where the likelihood $P(\mathbf{y} | \mathbf{h}) = P(\mathbf{y} | \mathbf{h}^S + \mathbf{h}^L)$ describes the distribution of \mathbf{y} for a given $\mathbf{h} = \mathbf{h}^S + \mathbf{h}^L$. Since the noise in the system is complex Gaussian, its distribution is

$$P(\mathbf{n}) = \frac{1}{(\pi\sigma^2)^{MN}} \exp\left(-\frac{\|\mathbf{n}\|_2^2}{\sigma^2}\right) \quad (5)$$

where $\|\cdot\|_2$ stands for the ℓ_2 -norm of the vector. Therefore, the likelihood function also follows the complex Gaussian distribution, i.e.,

$$\begin{aligned} P(\mathbf{y} | \mathbf{h}) &= P(\mathbf{y} | \mathbf{h}^S + \mathbf{h}^L) \\ &= \frac{1}{(\pi\sigma^2)^{MN}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{A}(\mathbf{h}^S + \mathbf{h}^L)\|_2^2}{\sigma^2}\right). \end{aligned} \quad (6)$$

In order to reflect the sparsity and low-rankness while maintaining mathematical solvability, we model the prior distributions using complex Gaussian distributions with adjustable variances [4], [7]. The prior distributions of \mathbf{h}^S and \mathbf{h}^L are given by

$$\begin{aligned} P(\mathbf{h}^S) &= \mathcal{N}_C(\mathbf{0}, \alpha^{-1} \mathbf{I}_{MK}) \\ P(\mathbf{h}^L) &= \mathcal{N}_C(\mathbf{0}, \beta^{-1} \mathbf{I}_{MK}) \end{aligned} \quad (7)$$

where α and β are hyperparameters controlling the levels of sparsity and low-rankness, respectively. As α and β increase, the variances decrease and more elements are concentrated around zero, resulting in stronger sparsity and low-rankness.

The posterior distribution $P(\mathbf{h}|\mathbf{y})$ is derived through Bayesian learning based on the prior distribution and likelihood function [7]. Assuming that channel multi-paths are independent regarding phase, delay, and amplitude without coupling or interference, \mathbf{h}^S and \mathbf{h}^L can be considered mutually independent. Substituting (6) and (7) into (4), the posterior distribution of \mathbf{h} can be expressed as

$$\begin{aligned} P(\mathbf{h}|\mathbf{y}) &= P(\mathbf{h}^S, \mathbf{h}^L|\mathbf{y}) \\ &\propto P(\mathbf{y}|\mathbf{h})P(\mathbf{h}^S)P(\mathbf{h}^L) \\ &\propto \exp\left(-\frac{\|\mathbf{y}-\mathbf{A}(\mathbf{h}^S+\mathbf{h}^L)\|_2^2}{\sigma^2}-\alpha\|\mathbf{h}^S\|_2^2-\beta\|\mathbf{h}^L\|_2^2\right). \end{aligned} \quad (8)$$

The optimal estimate of \mathbf{h} can be obtained by maximizing the posterior distribution presented in (8). Since the posterior distribution is a monotonically increasing exponential function, by removing the negative sign and neglecting the constant terms, we can estimate \mathbf{h} by solving

$$P_1: \min_{\mathbf{h}^S, \mathbf{h}^L} \|\mathbf{y}-\mathbf{A}(\mathbf{h}^S+\mathbf{h}^L)\|_2^2 + \alpha\|\mathbf{h}^S\|_0 + \beta \text{Rank}(\mathbf{H}^L) \quad (9)$$

where $\mathbf{H}^L = [\mathbf{H}_1^L, \dots, \mathbf{H}_K^L]$ is the low-rank component of the collective channel matrix with $\mathbf{h}^L = \text{vec}(\mathbf{H}^L)$. Moreover, $\|\cdot\|_0$ and $\text{Rank}(\cdot)$ represent the ℓ_0 -norm of the vector and rank of the matrix, respectively. The first term in (9) is the data fitting term in (8), representing the estimation error. The second and third terms are the sparse penalty term and the low-rank penalty term, respectively. We minimize the estimation error while simultaneously promoting sparsity and low-rankness through the corresponding penalty terms. This unconstrained formulation is equivalent to the Lagrangian form of a constrained problem that imposes explicit bounds on the sparsity and rank. Furthermore, in order to ensure the convexity of the problem, ℓ_1 -norm is introduced to constrain the sparsity and the nuclear norm is used as a convex approximation of the rank function [16]. As a result, we formulate the process of estimating the beam domain channel \mathbf{h} as a joint low-rank and sparse recovery problem P2, which is given by

$$P_2: \min_{\mathbf{h}^S, \mathbf{h}^L} \|\mathbf{y}-\mathbf{A}(\mathbf{h}^S+\mathbf{h}^L)\|_2^2 + \alpha\|\mathbf{h}^S\|_1 + \beta\|\mathbf{H}^L\|_* \quad (10)$$

where $\|\cdot\|_1$ is the ℓ_1 -norm of the vector and $\|\cdot\|_*$ is the nuclear norm of the matrix.

B. EM Framework for Iterative Updating

Since (10) contains non-smooth ℓ_1 -norm and the hyperparameters α and β need to be dynamically updated according to the current estimation results, we employ the EM framework [7], which is particularly suitable for probabilistic models containing hidden or unobserved variables, to solve the formulated problem P2. Within the EM framework, \mathbf{h}^S and \mathbf{h}^L , which is equivalent to \mathbf{H}^L , are treated as hidden variables and α and β as model parameters.

1) *E-Step*: This step calculates posterior expectations of \mathbf{h}^S and \mathbf{h}^L with fixed hyperparameters. The nuclear norm and ℓ_1 -norm terms make (10) difficult to be directly optimized.

Algorithm 1 The LRSBE Algorithm

Input: $\mathbf{y} \in \mathbb{C}^{MN}$, $\mathbf{A} \in \mathbb{C}^{MN \times MK}$, maximum allowed iterations Q .

Output: $\hat{\mathbf{h}} \in \mathbb{C}^{MK}$.

```

1: Initialization:  $i = 1$ ,  $\hat{\mathbf{h}}_i^S = \mathbf{0}$ ,  $\hat{\mathbf{h}}_i^L = \mathbf{0}$ ,  $\alpha = 1$ ,  $\beta = 1$ ;
2: while  $i \leq Q$  do
3:   Update  $\mathbf{r}^S = \mathbf{y} - \mathbf{A}\hat{\mathbf{h}}_i^L$ ;
4:   Calculate  $\mu$  and  $\Sigma$  through (12) and (13);
5:   Calculate  $\Gamma$  and  $\mathbf{C}$  through (14) and (15);
6:   Update  $\hat{\mathbf{h}}_{i+1}^S = \mu$ ;
7:   Update  $\mathbf{r}^L = \mathbf{y} - \mathbf{A}\hat{\mathbf{h}}_{i+1}^S$ ;
8:   Update  $\hat{\mathbf{h}}_{i+1}^L$  through (18) and (19);
9:   Limit the number of singular values through (20);
10:  Combine estimation result  $\hat{\mathbf{h}}_{i+1} = \hat{\mathbf{h}}_{i+1}^S + \hat{\mathbf{h}}_{i+1}^L$ ;
11:  Update  $\alpha$  and  $\beta$  through (22);
12:  if  $\|\hat{\mathbf{h}}_{i+1} - \hat{\mathbf{h}}_i\|_2 / \|\hat{\mathbf{h}}_i\|_2 \leq 10^{-4}$  then
13:    break;
14:  end if
15:   $i \leftarrow i + 1$ ;
16: end while
return  $\hat{\mathbf{h}} = \hat{\mathbf{h}}_i^S + \hat{\mathbf{h}}_i^L$ .

```

To this end, we propose a stepwise optimization approach that constrains the low-rankness and sparsity separately.

Considering that the channel sparse component exhibits block sparsity structure [17], \mathbf{h}^S can be processed in blocks. Let $\mathbf{h}^S = [(\mathbf{h}^S_1)^T, \dots, (\mathbf{h}^S_G)^T]^T$, where G is the number of blocks, L is the length of the single block, and \mathbf{h}^S_k is the k th block. The block length is designed based on the antenna dimension. Fixing \mathbf{h}^L , the optimal \mathbf{h}^S is achieved by solving

$$P_3: \min_{\{\mathbf{h}_k^S\}} \|\mathbf{r}^S - \mathbf{A}_k \mathbf{h}_k^S\|_2^2 + \alpha \sum_{k=1}^G \|\mathbf{h}_k^S\|_1 \quad (11)$$

where $\mathbf{r}^S = \mathbf{y} - \mathbf{A}\mathbf{h}^L$ denotes the sparse residual of the channel and $\mathbf{A}_k \in \mathbb{C}^{MN \times L}$ is an extracted measurement matrix in which the columns are associated with the k th block. Utilizing the block sparse Bayesian learning [8], the posterior mean and covariance of \mathbf{h}^S are

$$\mu = \Gamma \mathbf{A}^H (\mathbf{A} \Gamma \mathbf{A}^H + \sigma^2 \mathbf{I})^{-1} \mathbf{r}^S \quad (12)$$

$$\Sigma = \Gamma - \Gamma \mathbf{A}^H (\mathbf{A} \Gamma \mathbf{A}^H + \sigma^2 \mathbf{I})^{-1} \mathbf{A} \Gamma \quad (13)$$

where $\Gamma = \text{diag}(\gamma_1 \mathbf{C}_1, \dots, \gamma_G \mathbf{C}_G)$ represents the block covariance matrix, with \mathbf{C}_k and γ_k being the correlation matrix and weight of \mathbf{h}_k^S , respectively. Since assigning each block with an exclusive \mathbf{C}_k easily leads to overfitting [7], the correlation matrices are set to be identical. We have

$$\mathbf{C}_k = \mathbf{C} = \frac{1}{G} \sum_{k=1}^G \frac{\Sigma_k + \mu_k \mu_k^H}{\gamma_k}, \quad \forall k \quad (14)$$

$$\gamma_k = \frac{1}{L} \text{tr}(\mathbf{C}^{-1}(\Sigma_k + \mu_k \mu_k^H)) \cdot \frac{1}{1 + \alpha}. \quad (15)$$

Since the block sparsity is susceptible to noise, L is set to be a variable to achieve adaptive adjustment for different scenarios.

The low-rank component \mathbf{h}^L can be obtained by solving

$$P_4: \min_{\mathbf{h}^L} \|\mathbf{r}^L - \mathbf{A}\mathbf{h}^L\|_2^2 + \beta \|\mathbf{H}^L\|_* \quad (16)$$

where $\mathbf{r}^L = \mathbf{y} - \mathbf{A}\mathbf{h}^S$ represents the low-rank residual of the channel. To minimize the first term of (16), its gradient is

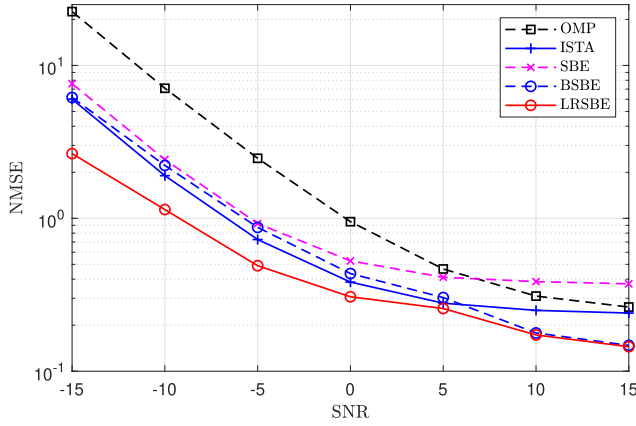


Fig. 2. Comparison of average NMSEs versus SNR for different algorithms ($M = 256, N = 5$).

calculated as

$$\nabla_{\mathbf{h}^L} \left(\|\mathbf{r}^L - \mathbf{A}\mathbf{h}^L\|_2^2 \right) = -\mathbf{A}^H (\mathbf{r}^L - \mathbf{A}\mathbf{h}^L). \quad (17)$$

Derived by the gradient descent formula, the initial estimates for the i th iteration is

$$\mathbf{h}_{i+1}^L = \mathbf{h}_i^L + \frac{1}{T} \mathbf{A}^H (\mathbf{r}^L - \mathbf{A}\mathbf{h}_i^L) \quad (18)$$

where $T = \lambda_{\max}(\mathbf{A}^H \mathbf{A})$ is the step length, set to be the maximum eigenvalue of $\mathbf{A}^H \mathbf{A}$ to ensure convergence.

To minimize the regularization term $\beta \|\mathbf{H}^L\|_*$ at the same time, we utilize a soft-thresholding operation [5] to impose a regularization constraint, i.e.,

$$h_{k,i+1}^L = \text{sign}(h_{k,i+1}^L) \cdot \max(0, |h_{k,i+1}^L| - \tau) \quad (19)$$

where $\text{sign}(\cdot)$ is the sign function, $h_{k,i+1}^L$ and $h_{k,i+1}^L$ represent the k th elements of \mathbf{h}^L and \mathbf{h}'^L obtained in the i th iteration, respectively. The threshold $\tau = \frac{\beta}{2T}$ controls the level of low-rankness. As the low-rank parameter β increases, more entries of \mathbf{h}^L are “compressed” to zero.

After performing singular value decomposition on \mathbf{H}^L , we apply soft thresholding to the singular values as

$$\sigma_k = \max \left(\sigma_k - \frac{\sqrt{\beta}}{2}, 0 \right). \quad (20)$$

As a result, the small singular values are removed.

2) *M-Step*: Based on the current estimations of \mathbf{h}^S and \mathbf{h}^L , the M-step adjusts the hyperparameters α and β by maximizing the likelihood functions, thus dynamically adapting to different channel conditions. The log-likelihood functions of \mathbf{h}^S and \mathbf{h}^L with respect to α and β are

$$\begin{aligned} \log P(\mathbf{h}^S) &= -\alpha \|\mathbf{h}^S\|_2^2 + MK \log \alpha \\ \log P(\mathbf{h}^L) &= -\beta \|\mathbf{h}^L\|_2^2 + MK \log \beta. \end{aligned} \quad (21)$$

Considering the posterior expectations of \mathbf{h}^S and \mathbf{h}^L , we have

$$\begin{aligned} \alpha &= \frac{MK}{\|\mathbf{y} - \mathbf{A}(\mathbf{h}^S + \mathbf{h}^L)\|_2^2 + \frac{1}{\alpha} \sum_k (1 - \theta_k)} \\ \beta &= \frac{MK}{\|\mathbf{h}^L\|_2^2 + \text{tr}(\mathbf{\Sigma}^L)} \end{aligned} \quad (22)$$

where $\theta_k = (\mu_k^S)^2 / \Sigma_{kk}^S$ represents the posterior weight for the k th entry of \mathbf{h}^S , in which μ_k^S and Σ_{kk}^S are posterior mean and covariance of the k th component of \mathbf{h}^S , respectively. Besides, $\mathbf{\Sigma}^L$ denotes the posterior covariance of \mathbf{h}^L .

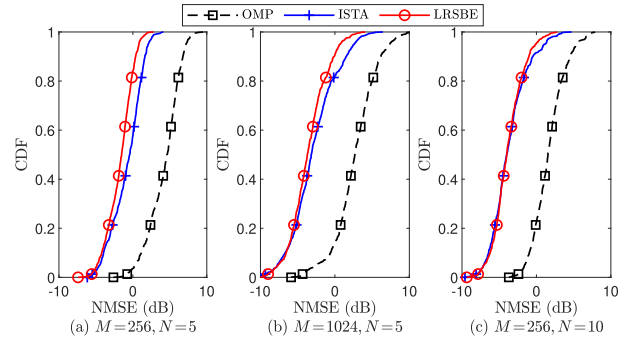


Fig. 3. CDFs of NMSEs with (a) $M = 256, N = 5$, (b) $M = 1024, N = 5$ and (c) $M = 256, N = 10$ (SNR = -10 dB).

This algorithm handles low-rankness and sparsity with nuclear norm and ℓ_1 -norm regularization, respectively. The procedure is summarized in Algorithm 1.

IV. SIMULATION RESULTS

An ultra-massive MIMO system consisting of $K = 10$ MTs is considered in an urban microcell scenario where the BS is equipped with M antennas, $M \in \{16 \times 16 = 256, 32 \times 32 = 1024\}$. The system operates at a carrier frequency of 11 GHz and the channel is generated according to [15]. The simulations are performed in MATLAB R2023b on a computer with Intel Core i7-12700 CPU @ 2.1 GHz and 16 GB of RAM. We conduct a comparative analysis to evaluate the performance of our proposed LRSBE algorithm. The analysis involves four representative comparative algorithms, which cover three categories of the CS algorithm. Specifically, the comparative algorithms selected include the greedy algorithm orthogonal matching pursuit (OMP) [6], the convex optimization-based iterative shrinkage-threshold algorithm (ISTA) [5], the sparse Bayesian learning estimation (SBE) algorithm [7], and our previously proposed block SBE (BSBE) algorithm [8]. The normalized MSE (NMSE) $\frac{1}{K} \sum_{k=1}^K \mathbb{E}\{\|\mathbf{h}_k - \hat{\mathbf{h}}_k\|^2 / \|\mathbf{h}_k\|^2\}$ is employed to evaluate the estimation accuracy.

Fig. 2 demonstrates that the proposed LRSBE outperforms other algorithms regarding NMSE under different signal-to-noise ratio (SNR) conditions. Since greedy algorithms are vulnerable to noise, OMP performs relatively unsatisfactorily under low SNR. In contrast, SBE has lower NMSE due to the self-inflicted prior assumption, with BSBE improving the estimation accuracy by further utilizing the block structure of channels. By using ℓ_1 -norm regularization, ISTA can suppress the effect of noise so that its accuracy is even better than BSBE but encounters a bottleneck due to a fixed soft threshold. However, these algorithms only utilize the channel sparsity while ignoring the channel low-rankness in the low SNR region that can be exploited to overcome the influence of noise. By integrating the low-rankness and sparsity simultaneously, the proposed LRSBE further exploits the block sparsity of the channel after 5 dB, thus significantly improves the estimation accuracy with a 3 dB ($\approx 50\%$) NMSE reduction compared to the best benchmark ISTA when SNR = -15 dB and a 1 dB ($\approx 20\%$) reduction compared to the best benchmark BSBE when SNR = 15 dB.

Moreover, Fig. 3 and Fig. 4 evaluate the NMSE of the proposed LRSBE with the greedy algorithm OMP and

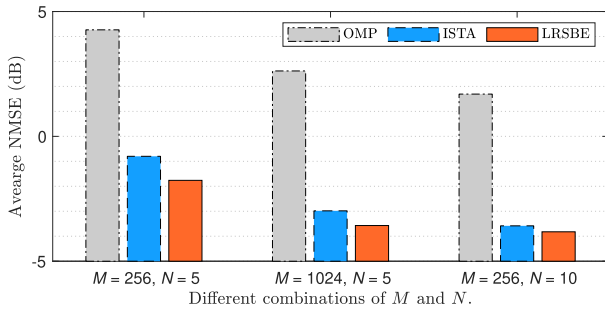


Fig. 4. Average NMSEs with different combinations of M and N (SNR = -10 dB).

TABLE I

AVERAGE RUNTIME OF THE CONSIDERED ALGORITHMS TO MEET NMSE = $5 \cdot 10^{-1}$ ($M = 256$, $N = 5$, AND SNR = 10 dB)

Scheme	OMP	ISTA	SBE	BSBE	LRSBE
Iteration numbers	601	43	14	36	5
Runtime (s)	108.482	13.823	54.392	229.173	23.965

convex-based algorithm ISTA, varying the number of pilots and antenna dimensions. Fig. 3(a), (b), and (c) display the cumulative distribution functions (CDFs) of NMSEs with three combinations of numbers of pilots and antenna dimensions, revealing that the estimation accuracy is improved by equipping with more antennas and utilizing more pilots. Quantitatively, Fig. 4 demonstrates the average NMSEs of the aforementioned three cases. Compared to the LRSBE with $M = 256$ and $N = 5$, the LRSBE with $M = 1024$ and $N = 5$ enjoys an almost 2 dB reduction of average NMSE through increasing the size of the antenna array. Moreover, increasing the number of pilots from 5 to 10 leads to a further 2 dB reduction in average NMSE, effectively mitigating pilot contamination when $M = 256$.

In terms of the complexity per iteration of the proposed LRSBE, it is concentrated on the posterior mean and covariance operations, where the inversion makes the single iteration complexity $\mathcal{O}(M^3)$, which is the same as the SBE and BSBE [7], [8]. The complexity of the OMP and ISTA is $\mathcal{O}(MN)$ [5], [6]. Fortunately, the fast convergence of the proposed LRSBE ensures low overall complexity with only a few iterations, whereas the high estimation accuracy of the BSBE is achieved at the expense of high overall complexity. This is clearly reflected in Table I, which presents the average runtime and number of iterations required to meet the specified performance criteria. Here, the value of SNR is set to be 10 dB, which can ensure a stable connectivity of wireless communications. Meanwhile, to comprehensively measure the complexity of all algorithms, we adopt a precision requirement of $5 \cdot 10^{-1}$. As shown in Table I, the average iteration times of the proposed LRSBE is reduced by more than 80% compared to the BSBE, which leads to a 90% reduction of complexity. Despite the slight decrease in complexity of ISTA compared to the LRSBE, the estimation accuracy is much lower than that of LRSBE.

V. CONCLUSION

In this letter, we have proposed a LRSBE algorithm for channel estimation in ultra-massive MIMO systems. We have

formulated a joint low-rank and sparse recovery problem and solved it through Bayesian learning within the EM framework, where the beam domain channel is decomposed into independent low-rank and sparse components. Simulation results have exhibited significant accuracy advantages of the proposed LRSBE algorithm compared to state-of-the-art alternatives, thanks to the exploitations of channel low-rankness and sparsity at low and high SNR regions, respectively. In addition, the complexity of the LRSBE is dramatically reduced compared to other high-precision channel estimation algorithms.

REFERENCES

- [1] C.-X. Wang et al., "On the road to 6G: Visions, requirements, key technologies and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, 2nd Quart., 2023.
- [2] S. Chen, J. Zhang, E. Björnson, J. Zhang, and B. Ai, "Structured massive access for scalable cell-free massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1086–1100, Apr. 2021.
- [3] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—what is next: Five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, pp. 3–20, Nov. 2019.
- [4] A. Das, "A Bayesian sparse-plus-low-rank matrix decomposition method for direction-of-arrival tracking," *IEEE Sensors J.*, vol. 17, no. 15, pp. 4894–4902, Aug. 2017.
- [5] A. Nasser, M. Elsabrouty, and O. Muta, "Weighted fast iterative shrinkage thresholding for 3D massive MIMO channel estimation," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–5.
- [6] J. Dong and L. Wu, "Comparison and simulation study of the sparse representation matching pursuit algorithm and the orthogonal matching pursuit algorithm," in *Proc. Int. Conf. Wireless Commun. Smart Grid (ICWCSG)*, Hangzhou, China, Aug. 2021, pp. 317–320.
- [7] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Apr. 2013.
- [8] J. Ji, H. Chang, C. Wang, J. Huang, and C. Huang, "A beam domain channel estimation algorithm based on sparse Bayesian learning," in *Proc. IEEE ICCT*, Oct. 2023, pp. 384–389.
- [9] M. Hawej and Y. R. Shayan, "Compressive sensing based nuclear norm minimization method for massive MU-MIMO channel estimation," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, May 2018, pp. 1–5.
- [10] L. Cheng, G. Yue, X. Xiong, Y. Liang, and S. Li, "Tensor decomposition-aided time-varying channel estimation for millimeter wave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1216–1219, Aug. 2019.
- [11] R. Zhang et al., "Tensor-based channel estimation for extremely large-scale MIMO-OFDM with dynamic metasurface antennas," *IEEE Trans. Wireless Commun.*, vol. 24, no. 7, pp. 6052–6068, Jul. 2025.
- [12] S. Chen et al., "Channel map-based angle domain multiple access for cell-free massive MIMO communications," *IEEE J. Sel. Topics Signal Process.*, vol. 19, no. 2, pp. 366–380, Mar. 2025.
- [13] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, Feb. 2018.
- [14] C.-X. Wang, Z. Lv, X. Gao, X. You, Y. Hao, and H. Haas, "Pervasive wireless channel modeling theory and applications to 6G GBsMs for all frequency bands and all scenarios," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9159–9173, Sep. 2022.
- [15] J. Bian et al., "A novel 3D beam domain channel model for massive MIMO communication systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1618–1632, Mar. 2023.
- [16] F. Y. Wu and F. Tong, "Non-uniform norm constraint LMS algorithm for sparse system identification," *IEEE Commun. Lett.*, vol. 17, no. 2, pp. 385–388, Feb. 2013.
- [17] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3705–3716, Dec. 2012.