Group Bayesian Sparse Channel Estimation for Massive MIMO Systems

Huajian Chen and Chenhao Qi
School of Information Science and Engineering
Southeast University, Nanjing 210096, China
Email: qch@seu.edu.cn

***Abstract*—In massive MIMO systems, different wireless channels starting from the same user antenna to different base station (BS) antennas usually share a common channel support. In this paper, Bayesian channel estimation is studied to jointly recover uplink channels sharing common support in massive MIMO systems. Since wireless channel is usually modeled as a mixed Bernoulli-Gaussian random process, it is employed as a prior knowledge for Bayesian estimation. A group Bayesian sparse channel estimation algorithm is proposed by enhancing the posterior probability of common support after jointly processing the received signal at BS. To reduce the computational complexity of Bayesian estimation, the genuine support set of uplink channels is approximated by a main support set. Moreover, the estimation of channel sparsity is also derived. Simulation results show that the proposed group Bayesian algorithm performs much better than the existing sparse recovery algorithms.**

***Index Terms*—Massive MIMO, channel estimation, sparse recovery, Bayesian estimation**

I. Introduction

Massive multiple-input multiple-output (MIMO) system, which has a large number of antennas equipped as antenna arrays at base station (BS), can sufficiently explore the spatial degree of freedom (DoF) to simultaneously serving multiple user terminals (UTs) with the same temporal and frequency resource. Therefore, massive MIMO has been regarded as an enabling technology for next generation wireless communications [1]. To benefit from spatial division multiplex access (SDMA) that can best exploit spatial DoF for wireless transmission, efficient beamforming formed by antenna arrays at BS is essentially important for massive MIMO. However, it requires accurate channel state information (CSI) at BS. Since the number of wireless links grows linearly with the number of antennas, the number of wireless channel to be estimated is large in massive MIMO systems, leading to huge pilot overhead and substantially increased computational complexity for pilot-assisted channel estimation. Regarding such issues, earlier works on massive MIMO study the time-division duplex (TDD) system, which can obtain downlink CSI from uplink channel estimation based on channel reciprocity. In this way, the heavy-burden uplink channel estimation can be performed at BS instead of at UTs.

Recent research on channel modelling shows that wireless channels are sparse. Limited number of scatters during wireless transmission result in a small number of channel multipath components sparsely distributed in relatively long delay spread. Furthermore, different channels linked to B-S shares a common support because the time of arrival (ToA) or time of departure (ToD) at different BS antennas is similar while the amplitude and phase of channel taps are distinct [2]. Based on these findings, several research works exploiting the common sparsity or partially common sparsity for massive MIMO channels are presented. In [3], a distributed compressive CSI estimation scheme is proposed so that the compressed measurements are observed at the users and the CSI recovery is jointly performed at BS. In [4], a spatially common sparsity based adaptive channel estimation and feedback scheme which adapts training overhead and pilot design to reliably estimate and feed back the downlink CSI is proposed. In [5], the block coherence of MIMO channels is analyzed for joint sparse channel estimation, which shows that as the number of BS antennas grows, the probability of joint recovery of the positions of nonzero channel entries will increase. An algorithm named block optimized orthogonal matching pursuit (BOOMP) is also proposed in [5] for joint sparse channel estimation. In [6], an algorithm named support agnostic Bayesian matching pursuit (SABMP) that is specially beneficial for massive MIMO system where nonzero entries of sparse channel obey non-Gaussian distribution is proposed.

In this paper, we consider Bayesian estimation and joint sparse recovery for uplink channels sharing a common support in massive MIMO systems. Since wireless channel is usually modeled as a mixed Bernoulli-Gaussian random process, where the position and the coefficients of channel taps obey Bernoulli distribution and Gaussian distribution respectively, we take it as prior knowledge for Bayesian estimation. A group Bayesian sparse channel estimation algorithm is proposed by enhancing the posterior probability of common support after jointly processing the received signal at BS. Moreover, considering that the dimension of channels in massive MIMO systems is large, it is computationally expensive to compute all posterior probabilities and expectations. To reduce the computational complexity of Bayesian estimation, we use a main support set to approximate the genuine support set with approximated minimum mean square error (MMSE) estimation. Since the sparsity of channels is unknown in practice, we also derive the estimation of the channel sparsity for the proposed group Bayesian sparse channel estimation algorithm.

The notations used in this paper are defined as follows. Symbols for matrices (upper case) and vectors (lower case) are

978-1-5090-5019-2/17/$31.00

in boldface. $(\cdot)^T$, $(\cdot)^H$, $\|\cdot\|_2$, diag$\{\cdot\}$, $\boldsymbol{I}_L$, $\mathbb{C}^M$, $\mathbf{0}^M$, $\mathbf{0}^{M\times N}$, E$\{\cdot\}$, det$\{\cdot\}$, $\mathcal{CN}$ and $\varnothing$, denote the matrix transpose, the matrix conjugate transpose (Hermitian), ℓ_2-norm, the diagonal matrix, the identity matrix of size L, the set of complex vectors with dimension M, the zero vector with dimension M, the $M\times N$ matrix, expectation, determinant of a matrix $\boldsymbol{a}$, the complex Gaussian distribution and the empty set, respectively.

II. Problem formulation

We consider a massive MIMO system including a BS equipped with N_B antennas and several user terminals (UTs) each equipped with a single antenna. We use orthogonal frequency division multiplexing (OFDM) for uplink transmission from the UT to BS. Suppose the total number of OFDM subcarriers is N_p, which is usually in an integer order of 2, e.g., $N_p = 256$. The UT employs $M(0 < M \le N_p)$ subcarriers with the corresponding indices $\boldsymbol{p} = [p_1, p_2, ..., p_M]^T (1 \le p_1 < p_2 < \cdots < p_M \le N_p)$ to transmit pilot symbols for pilot-assisted uplink channel estimation. The transmit pilot vector by UT is denoted as $\boldsymbol{x} = [x(p_1), x(p_2), \ldots, x(p_M)]^T$. Then the BS will receive N_B different pilot vectors, denoted as $\boldsymbol{y}^{(i)} = [y^{(i)}(p_1), y^{(i)}(p_2), \ldots, y^{(i)}(p_M)]^T, i = 1, 2, \ldots, N_B$, each experiencing different multipath fading. We denote the channel impulse response (CIR) of each multipath channel as $\boldsymbol{h}^{(i)} = [h^{(i)}(1), h^{(i)}(2), \ldots, h^{(i)}(N)]^T, i = 1, 2, \ldots, N_B$. We have

$$\boldsymbol{y}^{(i)} = \boldsymbol{DFh}^{(i)} + \boldsymbol{\eta}^{(i)}, i = 1, 2, \ldots, N_B \tag{1}$$

where $\boldsymbol{D} \triangleq \text{diag}\{\boldsymbol{x}\}$ denotes a diagonal matrix with the kth diagonal entries being $x(p_m)$, $m = 1, 2, \ldots, M$, $\boldsymbol{F}$ is an M by N DFT submatrix with row indices $\boldsymbol{p}$ and column indices $[1, 2, \ldots, N]$ selected from from a standard N_p by N_p discrete Fourier transform (DFT) matrix, and $\boldsymbol{\eta}^{(i)} \sim \mathcal{CN}(\mathbf{0}, \sigma_\eta^2 \boldsymbol{I}_M)$ is an additive white Gaussian noise (AWGN) term of ith uplink channel. We define a measurement matrix $\boldsymbol{\phi} \triangleq \boldsymbol{DF}$, then (1) is rewritten as

$$\boldsymbol{y}^{(i)} = \boldsymbol{\phi h}^{(i)} + \boldsymbol{\eta}^{(i)}, i = 1, 2, \ldots, N_B \tag{2}$$

Many literatures show that the wireless channel is typically sparse, where the number of nonzero taps of the channel, denoted as K, is much smaller than the channel length $N(0 < K \ll N)$. Then the sparse recovery algorithms can be applied for sparse channel estimation. Furthermore, it is shown in [2] that the CIR of different uplink channels shares a common support because the time of arrival (ToA) at different BS antennas is similar while the path amplitude and phase are distinct. In other words, the nonzero positions of $\boldsymbol{h}^{(i)}$ are the same for $i = 1, 2, \ldots, N_B$, while their nonzero coefficients are different. We can jointly consider the N_B equations in (2) and explore their joint sparsity.

III. Group Bayesian estimation

Now we will propose a group Bayesian estimation algorithm that exploits the knowledge of distribution of the sparse channel as well as taking advantage of the joint sparsity of massive MIMO uplink channels.

Wireless channel can usually be modeled as a mixed Bernoulli-Gaussian random process, which means the positions and the coefficients of channel taps obey Bernoulli distribution and Gaussian distribution, respectively [7] [8]. Since the CIR of different uplink channels in massive MIMO system shares a common support, we denote $\boldsymbol{s}$ as the common support of channel $\boldsymbol{h}^{(i)}, i = 1, 2, \ldots, N_B$, indicating a set of positions of nonzero channel entries, where the entries of $\boldsymbol{s}$ can be considered to be randomly selected from $\{1, 2, ..., N\}$ according to the Bernoulli distribution [9]. Define $\boldsymbol{\Theta}$ as a support set containing all possible occurrence of $\boldsymbol{s}$. Denote P_1 as the probability that each position is successfully extracted. The probability to generate a given $\boldsymbol{s}$ is

$$\mathcal{P}(\boldsymbol{s}) = P_1^K (1 - P_1)^{N-K} \tag{3}$$

where $K \triangleq |\boldsymbol{s}|$ denotes the number of entries in $\boldsymbol{s}$. It can be observed that the sparsity of $\boldsymbol{h}^{(i)}$ is essentially controlled by P_1, since $\text{E}\{K\} = NP_1$.

Define $\boldsymbol{h}_{\boldsymbol{s}}^{(i)}$ as a subset of $\boldsymbol{h}^{(i)}$ by selecting entries of $\boldsymbol{h}^{(i)}$ on positions of $\boldsymbol{s}$. Define $\boldsymbol{\phi}_{\boldsymbol{s}}$ as a submatrix by selecting columns of $\boldsymbol{\phi}$ on indices of $\boldsymbol{s}$. Assume that each entry of $\boldsymbol{h}_{\boldsymbol{s}}^{(i)}$ independently obeys complex Gaussian distribution with zero mean and variance of σ_1^2, i.e., $\boldsymbol{h}_s^{(i)} \sim \mathcal{CN}(0, \sigma_1^2 \boldsymbol{I}_K)$.

After the BS receives the signal $\boldsymbol{y}^{(i)},\ i = 1, 2, \ldots, N_B$ from all antennas, the MMSE estimation of $\boldsymbol{h}^{(i)}$ can be denoted as

$$\begin{aligned}\hat{\boldsymbol{h}}_{\text{MMSE}}^{(i)} &= \text{E}\left\{\boldsymbol{h}^{(i)} | \boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N_B)}\right\} \\ &= \sum_{\boldsymbol{s}} \text{E}\left\{\boldsymbol{h}^{(i)} | \boldsymbol{y}^{(i)}, \boldsymbol{s}\right\} \mathcal{P}(\boldsymbol{s} | \boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N_B)})\end{aligned} \tag{4}$$

where $\text{E}\{\boldsymbol{h}^{(i)} | \boldsymbol{y}^{(i)}, \boldsymbol{s}\}$ is the expectation of $\boldsymbol{h}^{(i)}$ given $\boldsymbol{y}^{(i)}$ and $\boldsymbol{s}$, and $\mathcal{P}(\boldsymbol{s} | \boldsymbol{y}^{(i)})$ is a posteriori probability. Note that given $\boldsymbol{s}$, $\boldsymbol{h}^{(i)}$ is only determined by $\boldsymbol{y}^{(i)}$. Here we can improve the MMSE estimation of $\boldsymbol{h}^{(i)}$ by jointly considering $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N_B)}$ exploring the common support of massive MIMO uplink channels.

Now we derive $\text{E}\{\boldsymbol{h}^{(i)} | \boldsymbol{y}^{(i)}, \boldsymbol{s}\}$. Since

$$\boldsymbol{y}^{(i)} = \boldsymbol{\phi}_{\boldsymbol{s}} \boldsymbol{h}_{\boldsymbol{s}}^{(i)} + \boldsymbol{\eta}^{(i)}, \tag{5}$$

the MMSE estimation of $\boldsymbol{h}_{\boldsymbol{s}}^{(i)}$ is

$$\text{E}\{\boldsymbol{h}^{(i)} | \boldsymbol{y}^{(i)}, \boldsymbol{s}\} = \boldsymbol{R}_{\boldsymbol{h}_{\boldsymbol{s}}^{(i)} \boldsymbol{h}_{\boldsymbol{s}}^{(i)}} \boldsymbol{\phi}_{\boldsymbol{s}}^H (\boldsymbol{\phi}_{\boldsymbol{s}} \boldsymbol{R}_{\boldsymbol{h}_{\boldsymbol{s}}^{(i)} \boldsymbol{h}_{\boldsymbol{s}}^{(i)}} \boldsymbol{\phi}_{\boldsymbol{s}}^H + \sigma_\eta^2 \boldsymbol{I}_M)^{-1} \boldsymbol{y}^{(i)}. \tag{6}$$

The covariance matrix of $\boldsymbol{h}_{\boldsymbol{s}}^{(i)}$ can be derived as

$$\boldsymbol{R}_{\boldsymbol{h}_{\boldsymbol{s}}^{(i)} \boldsymbol{h}_{\boldsymbol{s}}^{(i)}} = \text{E}\left\{\boldsymbol{h}_{\boldsymbol{s}}^{(i)} \{\boldsymbol{h}_{\boldsymbol{s}}^{(i)}\}^H\right\} = \sigma_1^2 \boldsymbol{I}_K. \tag{7}$$

Therefore,

$$\text{E}\{\boldsymbol{h}^{(i)} | \boldsymbol{y}^{(i)}, \boldsymbol{s}\} = \sigma_1^2 \boldsymbol{\phi}_{\boldsymbol{s}}^H (\sigma_1^2 \boldsymbol{\phi}_{\boldsymbol{s}} \boldsymbol{\phi}_{\boldsymbol{s}}^H + \sigma_\eta^2 \boldsymbol{I}_M)^{-1} \boldsymbol{y}^{(i)}. \tag{8}$$

Before deriving $\mathcal{P}(\boldsymbol{s} | \boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N_B)})$, we first derive $\mathcal{P}(\boldsymbol{s} | \boldsymbol{y}^{(i)})$. According to the Bayesian rule,

$$\mathcal{P}(\boldsymbol{s} | \boldsymbol{y}^{(i)}) = \frac{\mathcal{P}(\boldsymbol{y}^{(i)} | \boldsymbol{s}) \mathcal{P}(\boldsymbol{s})}{\mathcal{P}(\boldsymbol{y}^{(i)})}. \tag{9}$$

where $\mathcal{P}(\boldsymbol{s}|\boldsymbol{y}^{(i)})$ denotes a priori probability. Since $\mathcal{P}(\boldsymbol{y}^{(i)})$ can be treated as a constant and $\mathcal{P}(\boldsymbol{s})$ has been already given by (3), we have to compute $\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{s})$. Note that $\boldsymbol{h}_{\boldsymbol{s}}^{(i)} \sim \mathcal{CN}(0, \sigma_1^2 \boldsymbol{I}_K)$, we have $\boldsymbol{y}^{(i)}|\boldsymbol{s} \sim \mathcal{CN}(0, \boldsymbol{\Sigma}_{\boldsymbol{s}})$, where

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{s}} &= \mathrm{E}\left\{(\boldsymbol{\phi}_{\boldsymbol{s}}\boldsymbol{h}_{\boldsymbol{s}}^{(i)} + \boldsymbol{\eta}^{(i)})(\boldsymbol{\phi}_{\boldsymbol{s}}\boldsymbol{h}_{\boldsymbol{s}}^{(i)} + \boldsymbol{\eta}^{(i)})^H\right\} \\ &= \sigma_1^2 \boldsymbol{\phi}_{\boldsymbol{s}}\boldsymbol{\phi}_{\boldsymbol{s}}^H + \sigma_\eta^2 \boldsymbol{I}_M. \end{aligned} \tag{10}$$

According to the multidimensional Gaussian distribution, $\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{s})$ can be derived as

$$\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{s}) = \frac{1}{(\sqrt{2\pi})^N |\boldsymbol{\Sigma}_{\boldsymbol{s}}|^{\frac{1}{2}}} \exp\left(\frac{1}{2}\{\boldsymbol{y}^{(i)}\}^H \boldsymbol{\Sigma}_{\boldsymbol{s}}^{-1} \boldsymbol{y}^{(i)}\right). \tag{11}$$

Then we have

$$\mathcal{P}(\boldsymbol{s}|\boldsymbol{y}^{(i)}) = \frac{P_1^K (1-P_1)^{N-K} \exp\left(\frac{1}{2}\{\boldsymbol{y}^{(i)}\}^H \boldsymbol{\Sigma}_{\boldsymbol{s}}^{-1} \boldsymbol{y}^{(i)}\right)}{(\sqrt{2\pi})^N |\boldsymbol{\Sigma}_{\boldsymbol{s}}|^{\frac{1}{2}} \mathcal{P}(\boldsymbol{y}^{(i)})}. \tag{12}$$

In massive MIMO systems, the dimension of $\boldsymbol{h}^{(i)}$, $i = 1, 2, \ldots, N_B$ is large, which makes it impractical to compute all posterior probabilities and expectations. In order to reduce the computational complexity of Bayesian estimation, it is necessary to limit the number of possible support. We define $\boldsymbol{\Theta}^*$ as the main support set to approximate the genuine support set $\boldsymbol{\Theta}$. Then (4) can be approximated by

$$\hat{\boldsymbol{h}}_{\text{AMMSE}}^{(i)} = \sum_{\boldsymbol{s} \in \boldsymbol{\Theta}^*} \mathrm{E}\left\{\boldsymbol{h}^{(i)}|\boldsymbol{y}^{(i)}, \boldsymbol{s}\right\} \mathcal{P}(\boldsymbol{s}|\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N_B)}) \tag{13}$$

which is named as approximate MMSE (AMMSE). Therefore, how to find the main support is an important step to reduce the computational complexity of Bayesian estimation.

Since $\boldsymbol{h}^{(i)}$ can be modeled as a mixed Gaussian-Bernoulli random process and $\boldsymbol{h}^{(i)}, i = 1, 2, \ldots, N_B$ is sparse with a common support, we introduce a binary vector $\boldsymbol{z} \triangleq [z_1, z_2, \ldots z_N]^T$ with the same dimension as $\boldsymbol{h}^{(i)}$ to indicate the sparse pattern, i.e., the positions of nonzero entries of $\boldsymbol{h}^{(i)}$. We define $h^{(i)}[n]$ and $z[n]$ as the nth entry of $\boldsymbol{h}^{(i)}$ and $\boldsymbol{z}$, respectively. We use $z[n] \in \{0, 1\}$ to indicate whether $h^{(i)}[n]$ is nonzero, i.e., $z[n] = 1$ showing $h^{(i)}[n]$ is nonzero and $z[n] = 0$ showing $h^{(i)}[n]$ is zero. Since $h^{(i)}[n]$ obeys complex Gaussian distribution with mean zero and variance of σ_q^2, so

$$h^{(i)}[n]\big|\{z[n] = q\} \sim \mathcal{CN}(0, \sigma_q^2). \tag{14}$$

If $z[n] = 0$ ($q = 0$), we define $\sigma_0^2 \triangleq 0$; otherwise if $z[n] = 1$ ($q = 1$), we define $h^{(i)}[n] \sim \mathcal{CN}(0, \sigma_1^2)$. We have $h^{(i)}[n]|\{z[n] = 1\} \sim \mathcal{CN}(0, \sigma_1^2)$ and $h^{(i)}[n]|\{z[n] = 0\} \sim \mathcal{CN}(0, \sigma_0^2)$. Note that $\mathcal{P}(z[n] = 1) = P_1$, $\mathcal{P}(z[n] = 0) = 1 - P_1$. (14) can be rewritten in vector form as

$$\boldsymbol{h}^{(i)}|\boldsymbol{z} \sim \mathcal{CN}(0, \boldsymbol{R}_{zz}) \tag{15}$$

where $\boldsymbol{R}_{\boldsymbol{zz}} = \mathrm{E}\left\{\boldsymbol{z}\boldsymbol{z}^H\right\}$ is defined as the variance matrix of $\boldsymbol{z}$. It is seen that $\boldsymbol{R}_{\boldsymbol{zz}}$ is a diagonal matrix with the nth diagonal entry denoted as $R_{zz}[n, n] = \sigma_{z_n}^2$.

According to the Bayesian rule, the relationship between $\boldsymbol{z}$, $\boldsymbol{h}^{(i)}$ and $\boldsymbol{y}^{(i)}$ can be expressed as

$$\mathcal{P}(\boldsymbol{y}^{(i)}, \boldsymbol{h}^{(i)}|\boldsymbol{z}) = \frac{\mathcal{P}(\boldsymbol{y}^{(i)}, \boldsymbol{h}^{(i)}, \boldsymbol{z})}{\mathcal{P}(\boldsymbol{z})} = \mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{h}^{(i)}, \boldsymbol{z})\mathcal{P}(\boldsymbol{h}^{(i)}|\boldsymbol{z}). \tag{16}$$

Since $\boldsymbol{z}$ is determined by $\boldsymbol{h}^{(i)}$, we have

$$\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{z}, \boldsymbol{h}^{(i)}) = \mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{h}^{(i)}, \boldsymbol{z}) = \mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{h}^{(i)}). \tag{17}$$

Then (16) is

$$\mathcal{P}(\boldsymbol{y}^{(i)}, \boldsymbol{h}^{(i)}|\boldsymbol{z}) = \mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{h}^{(i)})\mathcal{P}(\boldsymbol{h}^{(i)}|\boldsymbol{z}). \tag{18}$$

Therefore the model can be formulated as

$$\begin{bmatrix}\boldsymbol{y}^{(i)} \\ \boldsymbol{h}^{(i)}\end{bmatrix}\Bigg|\boldsymbol{z} = \begin{bmatrix}\boldsymbol{y}^{(i)}|\boldsymbol{z} \\ \boldsymbol{h}^{(i)}|\boldsymbol{z}\end{bmatrix} \sim \mathcal{CN}\left(0, \begin{bmatrix}\boldsymbol{\varphi}(\boldsymbol{z}) & \boldsymbol{\phi}\boldsymbol{R}_{zz} \\ \boldsymbol{R}_{zz}\boldsymbol{\phi}^H & \boldsymbol{R}_{zz}\end{bmatrix}\right) \tag{19}$$

where

$$\boldsymbol{\varphi}(\boldsymbol{z}) \triangleq \boldsymbol{\phi}\boldsymbol{R}_{zz}\boldsymbol{\phi}^H + \boldsymbol{\sigma}_\eta^2 \boldsymbol{I}_M. \tag{20}$$

It is seen that the determination of nonzero entry's position of $\boldsymbol{h}^{(i)}$ is essentially the estimation of $\boldsymbol{z}$. The posterior probability of $\boldsymbol{z}$ is

$$\mathcal{P}(\boldsymbol{z}|\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N_B)}) = \frac{\mathcal{P}(\boldsymbol{z})\prod_{i=1}^{N_B}\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{z})}{\sum_{\boldsymbol{z}' \in \boldsymbol{\Lambda}}\mathcal{P}(\boldsymbol{z}')\prod_{i=1}^{N_B}\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{z}')} \tag{21}$$

where $\boldsymbol{\Lambda} \subset \{0, 1\}^N$. It's seen that $\boldsymbol{z}$ is a subset of $\boldsymbol{\Lambda}$. The equality is due to the fact that $\boldsymbol{y}^{(i)}$, $i = 1, 2, \ldots, N_B$ is independent to each other since $\boldsymbol{h}_{\boldsymbol{s}}^{(i)}$, $i = 1, 2, \ldots, N_B$ is independent to each other. Note that we can enhance the estimation of $\boldsymbol{z}$ via (21) by jointly processing the received signal at BS in massive MIMO systems.

According to (11), we have

$$\begin{aligned}&\ln\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{z}) \\ &= \ln\left(\frac{1}{(\sqrt{2\pi})^N|\varphi(\boldsymbol{z})|^{\frac{1}{2}}}\exp\left(\frac{1}{2}\{\boldsymbol{y}^{(i)}\}^H\varphi(\boldsymbol{z})^{-1}\boldsymbol{y}^{(i)}\right)\right) \\ &= -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln\det\big(\varphi(\boldsymbol{z})\big) - \frac{1}{2}(\boldsymbol{y}^{(i)})^H\varphi(\boldsymbol{z})^{-1}\boldsymbol{y}^{(i)}.\end{aligned} \tag{22}$$

Now we can define the selection metric for $\boldsymbol{z}$ as

$$\begin{aligned}v(\boldsymbol{z}) &\triangleq \ln\mathcal{P}(\boldsymbol{z})\prod_{i=1}^{N_B}\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{z}) \\ &= \ln\prod_{n=1}^{N}\mathcal{P}(z_n) + \ln\prod_{i=1}^{N_B}\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{z}) \\ &= \|z\|_0\ln P_1 + (N - \|\boldsymbol{z}\|_0)\ln(1-P_1) + \ln\prod_{i=1}^{N_B}\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{z}) \\ &= \|\boldsymbol{z}\|_0\ln\frac{P_1}{1-P_1} + N\ln(1-P_1) + \sum_{i=1}^{N_B}\ln\mathcal{P}(\boldsymbol{y}^{(i)}|\boldsymbol{z}) \\ &= \|\boldsymbol{z}\|_0\ln\frac{P_1}{1-P_1} + N\ln(1-P_1) - \frac{N_B N}{2}\ln 2\pi \\ &\quad - \frac{N_B}{2}\ln\det\big(\varphi(\boldsymbol{z})\big) - \frac{1}{2}\sum_{i=1}^{N_B}(\boldsymbol{y}^{(i)})^H\varphi(\boldsymbol{z})^{-1}\boldsymbol{y}^{(i)}.\end{aligned} \tag{23}$$

We define $\boldsymbol{z}_n$ as a vector where all entries are equal to $\boldsymbol{z}$ except the nth entry. The difference between $\boldsymbol{z}_n$ and $\boldsymbol{z}$ is denoted as

$$\Delta_n(\boldsymbol{z}_n) = v(\boldsymbol{z}_n) - v(\boldsymbol{z}). \tag{24}$$

Note that $\boldsymbol{\varphi}(\boldsymbol{0}) = \sigma_\eta^2 \boldsymbol{I}_M$ according to (20). At the initial state that $\boldsymbol{z} = \boldsymbol{0}$, we have

$$\begin{aligned} v(\boldsymbol{0}) =& N \ln(1-P_1) - \frac{N_B N}{2} \ln 2\pi - N_B M \ln \sigma_\eta \\ & - \frac{1}{2\sigma_\eta^2} \sum_{i=1}^{N_B} \left\| \boldsymbol{y}^{(i)} \right\|_2^2 \end{aligned} \tag{25}$$

In order to derive $v(\boldsymbol{z}_n)$, we first compute $\boldsymbol{\varphi}(\boldsymbol{z}_n)$ based on (20) as

$$\begin{aligned} \boldsymbol{\varphi}(\boldsymbol{z}_n) &= \boldsymbol{\phi} \boldsymbol{R}_{\boldsymbol{z}_n \boldsymbol{z}_n} \boldsymbol{\phi}^H + \sigma_n^2 \boldsymbol{I}_M = \boldsymbol{\phi}(\boldsymbol{R}_{\boldsymbol{z}\boldsymbol{z}} + \boldsymbol{A})\boldsymbol{\phi}^H \\ &= \boldsymbol{\varphi}(\boldsymbol{z}) + \sigma_1^2 \boldsymbol{\phi}_n \boldsymbol{\phi}_n^H, \end{aligned} \tag{26}$$

where $\boldsymbol{A}$ is an $N \times N$ square matrix, with all the entries equal zero except $A[n,n] = \sigma_1^2$, and $\boldsymbol{\phi}_n$ denotes the nth column of $\boldsymbol{\phi}$. According to the matrix inversion lemma, we have

$$\begin{aligned} \boldsymbol{\varphi}(\boldsymbol{z}_n)^{-1} =& \boldsymbol{\varphi}(\boldsymbol{z})^{-1} - \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{\phi}_n \big(\boldsymbol{\phi}_n^H \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{\phi}_n \\ & + \sigma_1^{-2} \big)^{-1} \boldsymbol{\phi}_n^H \boldsymbol{\varphi}(\boldsymbol{z})^{-1}. \end{aligned} \tag{27}$$

We define

$$\boldsymbol{b}_n \triangleq \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{\phi}_n \tag{28}$$

and

$$\beta_n \triangleq \sigma_1^2 \big(1 + \sigma_1^2 \boldsymbol{\phi}_n^H \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{\phi}_n \big)^{-1} = \sigma_1^2 (1 + \sigma_1^2 \boldsymbol{\phi}_n^H \boldsymbol{b}_n). \tag{29}$$

Then (27) can be further denoted as

$$\boldsymbol{\varphi}(\boldsymbol{z}_n)^{-1} = \boldsymbol{\varphi}(\boldsymbol{z})^{-1} - \beta_n \boldsymbol{b}_n \boldsymbol{b}_n^H. \tag{30}$$

Suppose $\boldsymbol{z}$ is obtained by changing the entry of $\boldsymbol{z}^{\text{pre}}$ on the position n^{pre} while $\boldsymbol{z}_n$ is obtained by changing the entry of $\boldsymbol{z}$ on the position n. We define

$$\begin{aligned} \boldsymbol{b}_{n^{\text{pre}}}^{\text{pre}} &\triangleq \boldsymbol{\varphi}(\boldsymbol{z}^{\text{pre}})^{-1} \boldsymbol{\phi}_{n^{\text{pre}}} \\ \boldsymbol{b}_{n}^{\text{pre}} &\triangleq \boldsymbol{\varphi}(\boldsymbol{z}^{\text{pre}})^{-1} \boldsymbol{\phi}_{n} \\ \beta_{n^{\text{pre}}}^{\text{pre}} &\triangleq \sigma_1^2 \left(1 + \sigma_1^2 \boldsymbol{\phi}_{n^{\text{pre}}}^H \boldsymbol{b}_{n^{\text{pre}}}^{\text{pre}} \right)^{-1} \end{aligned}$$

Then $\boldsymbol{b}_n$ can be further denoted as

$$\begin{aligned} \boldsymbol{b}_n &= \left(\boldsymbol{\varphi}\left(\boldsymbol{z}^{\text{pre}}\right)^{-1} - \beta_{n^{\text{pre}}}^{\text{pre}} \boldsymbol{b}_{n^{\text{pre}}}^{\text{pre}} \left(\boldsymbol{b}_{n^{\text{pre}}}^{\text{pre}}\right)^H \right) \boldsymbol{\phi}_n \\ &= \boldsymbol{b}_n^{\text{pre}} - \beta_{n^{\text{pre}}}^{\text{pre}} \boldsymbol{b}_{n^{\text{pre}}}^{\text{pre}} \left(\boldsymbol{b}_{n^{\text{pre}}}^{\text{pre}}\right)^H \boldsymbol{\phi}_n \end{aligned} \tag{31}$$

According to (26) and (30), we can compute the fourth term of (23) as

$$\begin{aligned} \ln\det\big(\boldsymbol{\varphi}(\boldsymbol{z}_n)\big) &= \ln\det\big(\boldsymbol{\varphi}(\boldsymbol{z}) + \sigma_1^2 \boldsymbol{\phi}_n \boldsymbol{\phi}_n^H\big) \\ &= \ln\left(\det\big(\boldsymbol{\varphi}(\boldsymbol{z})\big) \det\big(\boldsymbol{I}_M + \sigma_1^2 \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{\phi}_n \boldsymbol{\phi}_n^H\big) \right) \\ &= \ln\big(\det\boldsymbol{\varphi}(\boldsymbol{z})\big)\big(1 + \sigma_1^2 \boldsymbol{\phi}_n^H \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{\phi}_n\big) \\ &= \ln\det\big(\boldsymbol{\varphi}(\boldsymbol{z})\big) + \ln\big(1 + \sigma_1^2 \boldsymbol{\phi}_n^H \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{\phi}_n\big) \\ &= \ln\det\big(\boldsymbol{\varphi}(\boldsymbol{z})\big) - \ln \frac{\beta_n}{\sigma_1^2}. \end{aligned} \tag{32}$$

The third term of (23) is

$$\begin{aligned} (\boldsymbol{y}^{(i)})^H \boldsymbol{\varphi}(\boldsymbol{z}_n)^{-1} \boldsymbol{y}^{(i)} &= (\boldsymbol{y}^{(i)})^H \big(\boldsymbol{\varphi}(\boldsymbol{z})^{-1} - \beta_n \boldsymbol{b}_n \boldsymbol{b}_n^H \big) \boldsymbol{y}^{(i)} \\ &= (\boldsymbol{y}^{(i)})^H \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{y}^{(i)} - \beta_n |(\boldsymbol{y}^{(i)})^H \boldsymbol{b}_n|^2. \end{aligned} \tag{33}$$

The first term of (23) is

$$\begin{aligned} \|\boldsymbol{z}_n\|_0 \ln \frac{P_1}{1-P_1} &= (\|\boldsymbol{z}\|_0 + 1) \ln \frac{P_1}{1-P_1} \\ &= \|\boldsymbol{z}\|_0 \ln \frac{P_1}{1-P_1} + \ln \frac{P_1}{1-P_1}. \end{aligned} \tag{34}$$

Finally we substitute (32), (33) and (34) into (23), we have

$$\begin{aligned} v(\boldsymbol{z}_n) =& -\frac{N_B N}{2} \ln 2\pi - \frac{N_B}{2} \left(\ln\det\big(\boldsymbol{\varphi}(\boldsymbol{z})\big) - \ln \frac{\beta_n}{\sigma_1^2} \right) \\ & - \frac{1}{2} \sum_{i=1}^{N_B} \Big((\boldsymbol{y}^{(i)})^H \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{y}^{(i)} - \beta_n |(\boldsymbol{y}^{(i)})^H \boldsymbol{b}_n|^2 \Big) \\ & + \|\boldsymbol{z}\|_0 \ln \frac{P_1}{1-P_1} + \ln \frac{P_1}{1-P_1} + N \ln(1-P_1) \\ =& v(\boldsymbol{z}) + \Delta_n(\boldsymbol{z}_n). \end{aligned} \tag{35}$$

Now we can obtain $\Delta_n(\boldsymbol{z}_n)$ based on (24) as

$$\Delta_n(\boldsymbol{z}_n) = \frac{N_B}{2} \ln \frac{\beta_n}{\sigma_1^2} + \frac{1}{2} \beta_n \sum_{i=1}^{N_B} |(\boldsymbol{y}^{(i)})^H \boldsymbol{b}_n|^2 + \ln \frac{P_1}{1-P_1}. \tag{36}$$

After finishing the estimation of $\boldsymbol{z}$, we now further estimate $\boldsymbol{h}^{(i)},\ i = 1, 2, \ldots, N_B$. We define $\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, ...\boldsymbol{b}_N]$ which is a matrix formed by the columns from $\boldsymbol{b}_1$ to $\boldsymbol{b}_N$. According to (28), we have

$$\boldsymbol{B} = \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{\phi}. \tag{37}$$

Based on (8), we define

$$\begin{aligned} \boldsymbol{h}_e^{(i)} &\triangleq \mathrm{E}\left\{ \boldsymbol{h}^{(i)} | \boldsymbol{y}^{(i)}, \boldsymbol{z} \right\} = \sigma_1^2 \boldsymbol{\phi}_{\boldsymbol{z}}^H (\sigma_1^2 \boldsymbol{\phi}_{\boldsymbol{z}} \boldsymbol{\phi}_{\boldsymbol{z}}^H + \sigma_\eta^2 \boldsymbol{I}_M)^{-1} \boldsymbol{y}^{(i)} \\ &= \sigma_1^2 \boldsymbol{\phi}_{\boldsymbol{z}}^H \boldsymbol{\varphi}(\boldsymbol{z})^{-1} \boldsymbol{y}^{(i)} = \sigma_1^2 \boldsymbol{B}_z^H \boldsymbol{y}^{(i)} \end{aligned} \tag{38}$$

where $\boldsymbol{B}_z$ is a submatrix formed by the columns indexed by $\boldsymbol{z}$ from $\boldsymbol{B}$. According to (23), we have

$$\mathcal{P}(\boldsymbol{z}) \prod_{i=1}^{N_B} \mathcal{P}(\boldsymbol{y}^{(i)} | \boldsymbol{z}) = e^{v(\boldsymbol{z})}. \tag{39}$$

Since $\boldsymbol{\Lambda}$ is large, the computational complexity of (21) is high. To reduce the computational complexity, we introduce $\boldsymbol{\Lambda}^*$ to approximate $\boldsymbol{\Lambda}$. Substituting (39) into (21), we have

$$\mathcal{P}(\boldsymbol{z} | \boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N_B)}) = \frac{e^{v(\boldsymbol{z})}}{\sum_{\boldsymbol{z}' \in \Lambda} e^{v(\boldsymbol{z}')}} \approx \frac{e^{v(\boldsymbol{z})}}{\sum_{\boldsymbol{z}' \in \Lambda^*} e^{v(\boldsymbol{z}')}}. \tag{40}$$

We define

$$p_e \triangleq \frac{e^{v(\boldsymbol{z})}}{\sum_{\boldsymbol{z}' \in \Lambda^*} e^{v(\boldsymbol{z}')}}. \tag{41}$$

Finally, the AMMSE estimation of $\boldsymbol{h}$ in (13) can be rewritten as

$$\hat{\boldsymbol{h}}_{\text{AMMSE}}^{(i)} = \sum_{\Lambda^*} \boldsymbol{h}_e^{(i)} p_e. \tag{42}$$

Algorithm 1 Group Bayesian Sparse Channel Estimation

1: *Input:* $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N_B)}, M, N, D, P_0, P_1, \sigma_1, \sigma_\eta, \boldsymbol{\phi}$.
2: Obtain an estimate of the sparsity $\hat{K}$ via (44).
3: Set $\boldsymbol{p}_e \leftarrow \mathbf{0}^{\hat{K}D}$ and $\boldsymbol{w} \leftarrow \mathbf{0}^{\hat{K}D}$.
4: Set $\boldsymbol{H}_e^{(i)} \leftarrow \mathbf{0}^{N\times\hat{K}D},\ i = 1, 2, \ldots, N_B$.
5: Obtain $v(\mathbf{0})$ via (25).
6: **for** $n = 1, ..., N$ **do**
7: Compute $\boldsymbol{b}_n$ and β_n via (28) and (29), respectively.
8: Compute $\Delta_n(\boldsymbol{z}_n)$ via (36).
9: $v(\boldsymbol{z}_n) \leftarrow v(\mathbf{0}) + \Delta_n(\boldsymbol{z}_n)$
10: **end for**
11: **for** $d = 1 : D$ **do**
12: $\Omega \leftarrow \varnothing$, $\Psi \leftarrow \{1, 2, ...N\}$.
13: **for** $k = 1 : \hat{K}$ **do**
14: **if** $k = 1$ **then**
15: $n^* \leftarrow$ the index corresponding to the dth largest value in $\{v(\boldsymbol{z}_n), n \in \Psi\}$.
16: **else**
17: $n^* \leftarrow \arg\max_{l\in\Psi} v(\boldsymbol{z}_i)$.
18: **end if**
19: $\Omega \leftarrow \Omega \bigcup \{n^*\}$, $\Psi \leftarrow \{1, 2, ...N\} \setminus \Omega$.
20: $\boldsymbol{B}_\Omega \leftarrow \{\boldsymbol{b}_l, l \in \Omega\}$.
21: $\boldsymbol{H}_e^{(i)}[\Omega, (d-1)\hat{K} + k] = \sigma_1^2 \boldsymbol{B}_\Omega^H \boldsymbol{y}^{(i)}, i = 1, \ldots, N_B$.
22: $w[(d-1)\hat{K} + k] \leftarrow v(\boldsymbol{z}_{n^*})$.
23: Update $\boldsymbol{b}_l$ by $\boldsymbol{b}_l \leftarrow \boldsymbol{b}_l - \beta_{n^*} \boldsymbol{b}_{n^*} \boldsymbol{b}_{n^*}^H \boldsymbol{\phi}_l, l \in \Psi$.
24: Update $\beta_l, l \in \Psi$ via (29).
25: Update $\Delta_n(\boldsymbol{z}_n)$ via (36).
26: Update $v(\boldsymbol{z}_n)$ by $v(\boldsymbol{z}_n) \leftarrow v(\boldsymbol{z}_{n^*}) + \Delta_n(\boldsymbol{z}_n)$.
27: **end for**
28: **end for**
29: Compute each entry of $\boldsymbol{p}_e$ via (45).
30: Compute $\mathrm{E}\{\boldsymbol{h}^{(i)}|\boldsymbol{y}^{(i)}\} = \boldsymbol{H}_e^{(i)} \boldsymbol{p}_e,\ i = 1, 2, \ldots, N_B$.
31: *Output:* $\mathrm{E}\{\boldsymbol{h}^{(i)}|\boldsymbol{y}^{(i)}\}$, $i = 1, 2, \ldots, N_B$.

Since the sparsity of the channel, K, is usually unknown, we have to obtain an estimate of K, denoted as $\hat{K}$.

Note that $\|\boldsymbol{z}\|_0$ obeys the Binomial distribution with the mean of NP_1 and the variance of $NP_1(1-P_1)$. If $NP_1 > 5$, the Binomial distribution can be well approximated by the Gaussian distribution with the same mean and variance. Given $\hat{K}$, we define

$$\begin{aligned} P_0 &\triangleq \mathcal{P}(\|\boldsymbol{z}\|_0 > \hat{K}) = 1 - \mathcal{P}(\|\boldsymbol{z}\|_0 \le \hat{K}) \\ &= 1 - \int_{-\infty}^{\hat{K}} \frac{1}{\sqrt{2\pi NP_1(1-P_1)}} \exp\left(\frac{-(x-NP_1)^2}{2NP_1(1-P_1)}\right) \mathrm{d}x \\ &= 1 - \boldsymbol{\Phi}\left(\frac{\hat{K} - NP_1}{\sqrt{2NP_1(1-P_1)}}\right) \\ &= \frac{1}{2}\mathrm{erfc}\left(\frac{\hat{K} - NP_1}{\sqrt{2NP_1(1-P_1)}}\right) \end{aligned} \tag{43}$$

where

$$\boldsymbol{\Phi}(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dx = \frac{1}{2}\left[1 + \mathrm{erf}(\frac{x}{\sqrt{2}})\right]$$

denotes the cumulative distribution function of standard Gaussian distribution. Further considering that the number of nonzero entries of the sparse vector is no larger than the dimension of observation, $\hat{K}$ can be derived from (43) as follows

$$\hat{K} = \min\left\{M, \left\lceil \mathrm{erfc}^{-1}(2P_0)\sqrt{2NP_1(1-P_1)} + NP_1 \right\rceil\right\}. \tag{44}$$

We usually set P_0 small, e.g., $P_0 = 0.01$ to obtain $\hat{K}$, which may be slightly larger than K. In our proposed group Bayesian estimation algorithm, we do not require $\hat{K}$ exactly the same as K.

We summarize the above procedures in Algorithm 1. At first, we obtain an estimate of the sparsity $\hat{K}$ via (44). We set $\boldsymbol{w}$ and $\boldsymbol{p}_e$ to be zero vectors. We set $\boldsymbol{H}_e$ to be a zero matrix to store the results of $\boldsymbol{h}_e$ for all iterations. Then we obtain $v(\mathbf{0})$ via (25). From step 6 to step 10, we obtain $\{v(\boldsymbol{z}_l), l = 1, 2\ldots, N\}$. From step 11 to step 28, we use D outer-loop iterations, with each iteration using the $d(d = 1, 2, \ldots, D)$th largest value of $\{v(\boldsymbol{z}_l), l = 1, 2\ldots, N\}$ as the starting value for the inner-loop iterations from step 19 to step 26. Different starting value may lead to different final results, where the weighted average based on the probability will be computed in the end. All steps follow our precedent derivations, e.g., step 21 is based on (38) and step 21 is based on (31). After the outer-loop iterations are finished, based on (41) we can compute each entry of $\boldsymbol{p}_e$, denoted as $p_e(l),\ l = 1, 2, \ldots, \hat{K}D$ by the following equation

$$p_e(l) = \frac{e^{w[l]}}{\sum_{k=1}^{\hat{k}D} e^{w[k]}},\ l = 1, 2, \ldots, \hat{K}D. \tag{45}$$

Finally we use $\boldsymbol{H}_e^{(i)}$ and $\boldsymbol{p}_e$ to compute $\mathrm{E}\{\boldsymbol{h}^{(i)}|\boldsymbol{y}^{(i)}\}$ via (42), i.e., $\mathrm{E}\{\boldsymbol{h}^{(i)}|\boldsymbol{y}^{(i)}\} = \boldsymbol{H}_e^{(i)}\boldsymbol{p}_e,\ i = 1, 2, \ldots, N_B$.

IV. Simulation Results

We consider a massive MIMO system including a BS equipped with $N_B = 64$ antennas and several single-antenna UTs. We employ OFDM for uplink transmission, where $M = 16$ subcarriers among totally $N_p = 256$ OFDM subcarriers are used to transmit pilot symbols for uplink channel estimation. QPSK modulation is used for simulations. The length of CIR is set to $N = 60$ where the number of nonzero channel taps is set to be $K = 10$. For each channel implementation, we randomly select $K = 10$ taps from totally $N = 60$ taps as nonzero taps where the coefficients of these nonzero channel taps independently obey zero mean and unit variance complex Gaussian distribution.

We compare the proposed group Bayesian sparse channel estimation with the existing sparse channel estimation methods, including OMP, BOOMP and Bayesian algorithms in terms of mean squared error (MSE) and bit error rate (BER) in Fig. 1 and Fig. 2, respectively. The comparisons are averaged over 1000 random channel implementations. We define the MSE as

$$\mathrm{MSE} = \frac{1}{N_B} \sum_{i=1}^{N_B} \frac{\|\hat{\boldsymbol{h}}^{(i)} - \boldsymbol{h}^{(i)}\|_2^2}{\|\boldsymbol{h}^{(i)}\|_2^2}. \tag{46}$$

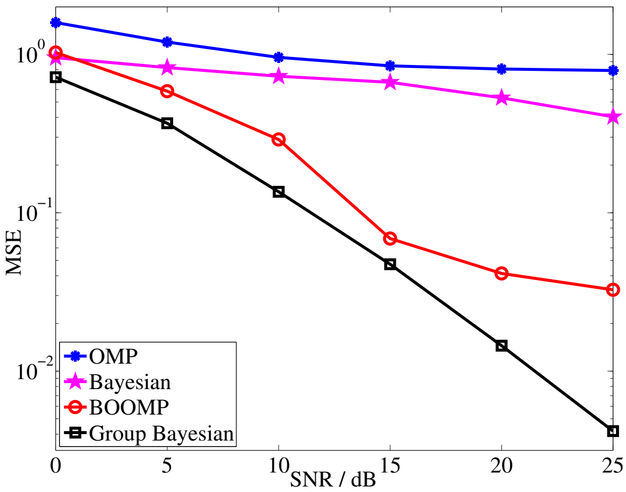

Fig. 1. Comparisons of MSE performance for different channel estimation algorithms.

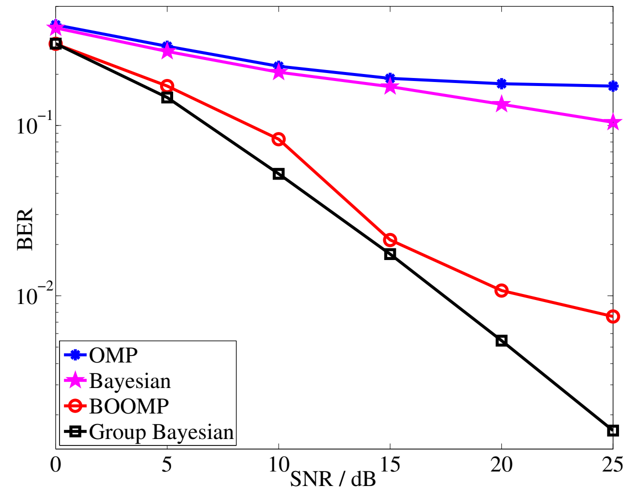

Fig. 2. Comparisons of BER performance for different channel estimation algorithms.

We set $P_0 = 0.005$, $D = 5$ and $P_1 = 0.167$ for Bayesian and group Bayesian algorithm. The estimate of the sparsity $\hat{K} = 16$ can be obtained via (44). It is seen that $\hat{K}$ is a bit larger than K.

As shown in Fig. 1 and Fig. 2, the proposed group Bayesian performs much better than original Bayesian and OMP algorithm, since the beneficial group sparsity and common support of massive MIMO channels are not utilized by Bayesian and OMP algorithm. Note that in our simulations we set the sparsity as $K = 10$, which is larger than the half of the length of observation, i.e., $K > M/2$, theoretically it is difficult to make sparse recovery since it violates the rule that the observation length should be at least twice of the sparsity. It is the essential reason for the poor performance of original Bayesian and OMP.

Considering that BER around 0.01 can be completely removed after the channel encoding and decoding are additionally included, e.g., turbo channel coding, we fixed BER to be 0.008. It is seen from Fig. 2 that to achieve the same BER of 0.008, group Bayesian can save around 5dB SNR compared to BOOMP.

V. Conclusions

In this paper we have studied the Bayesian estimation and joint sparse recovery for uplink channels sharing common support in massive MIMO systems. Since wireless channel is usually modeled as a mixed Bernoulli-Gaussian random process, we have employed it as prior knowledge for Bayesian estimation. A group Bayesian sparse channel estimation algorithm has been proposed by enhancing the posterior probability of common support after jointly processing the received signal at BS. To reduce the computational complexity of Bayesian estimation, we have used a main support set to approximate the genuine support set with AMMSE estimation. We have also derived the estimation of channel sparsity for the proposed group Bayesian sparse channel estimation algorithm.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61302097, by the Natural Science Foundation of Jiangsu Province under Grant BK20161428, and by the National Science and Technology Major Project of China under Grant 2016ZX03001016-003.

References

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

[2] 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.

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

[4] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[5] C. Qi, Y. Huang, S. Jin, and L. Wu, "Sparse channel estimation based on compressed sensing for massive MIMO systems," in *Proc. ICC*, London, UK, 2015, pp. 4458–4563.

[6] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 104–118, Jan. 2014.

[7] O. Rabaste and T. Chonavel, "Estimation of multipath channels with long impulse response at low SNR via an MCMC method," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1312–1325, Apr. 2007.

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

[9] P. Schniter, L. C. Potter, and J. Ziniel, "Fast Bayesian matching pursuit," in *Proc. ITA*, San Diego, California, Jan. 2008, pp. 326–333.