CrossMark

EM-based parameter iterative approach for sparse Bayesian channel estimation of massive MIMO system

Sulin Mei and Yong Fang[*]

Abstract

One of the main challenges for a massive multi-input multi-output (MIMO) system is to obtain accurate channel state information despite the increasing number of antennas at the base station. The Bayesian learning channel estimation methods have been developed to reconstruct the sparse channel. However, these existing methods depend heavily on the channel distribution. In this paper, based on sparse Bayesian method, an expectation maximization-based parameter iterative approach is proposed to estimate the massive MIMO channel with unknown channel distribution. Using the approximate sparse feature, the massive MIMO channel is modeled as a non-zero Gaussian mixture and the sparse Bayesian channel estimation is introduced. The channel marginal probability density function is expressed by using the general approximate message-passing algorithm. All of the required channel parameters are iteratively estimated by the EM method. Simulation results show that the proposed scheme enables evident performance in channel estimation accuracy with a lower complexity when channel distribution is unknown.

Keywords: Massive MIMO, Expectation maximization, Sparse Bayesian learning, General approximate message passing

1 Introduction

Massive multi-input multi-output (MIMO) [1], which equipping a large number antennas at the base station (BS) to simultaneously serve tens of users in the same time-frequency channel, is widely considered as one of the key techniques for future communication network. Such systems can greatly improve the system capacity and energy efficiency by exploiting the increased degree of spatial freedom. Accurate downlink channel state information (CSI) is essential for massive MIMO systems to realize high-speed communication, and it is also vital for signal detection, beam forming, and other operations. However, it is difficult for the BS to obtain accurate downlink CSI due to the prohibitively high overhead used for the downlink channel estimation as the number of antennas increases in massive MIMO systems. Most of the researches exploit the time division

duplex(TDD) protocol to overcome this challenge. In TDD, the downlink CSI could be achieved through the uplink channel estimation by using the channel reciprocity property. The CSI in the uplink is more easily acquired at the BS due to the limited number of users. However, the CSI acquired in the uplink is not always accurate for the downlink. In addition, current wireless networks are mainly dominated by the frequency division duplex (FDD) protocol. Thus, it is of great importance to explore the downlink CSI. Compressed sensing (CS), which can reconstruct the sparse channel through few pilots, is viewed as a promising method to solve this problem.

Experimental results in [2, 3] show that the massive MIMO channel is approximate sparse due to the limited scatters at the BS. By exploiting the temporal correlation and the sparsity of massive MIMO channels in time-domain, [4] applied a sparse channel estimation scheme and sharply reduced the pilot overhead. But the sparsity in time domain maybe disappear with the increasing number of scatters at the user side [5]. Gao et al. [6] exploited the spatial correlation of massive MIMO channels and

*Correspondence: yfang@staff.shu.edu.cn
Key laboratory of Specialty Fiber Optics and Optical Access Networks, Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai, 200444 China

estimated massive MIMO channels through CS. In [7], both spatial and temporal correlation were considered, and a structured compressive sensing (SCS)-based spatio-temporal joint channel estimation scheme is proposed to further reduce the pilot overhead and improve the estimation accuracy. These schemes estimate channels mostly based on greedy algorithms, such as the orthogonal matching pursuit (OMP) algorithm, the subspace pursuit (SP) algorithm, etc. The sparse Bayesian learning [8] has been developed to estimate the sparse channel recently. It takes channel distribution into consideration, resulting in a high accuracy of estimator. However, the parameters of channel distribution are hard to obtain in practice.

In this paper, an expectation maximization (EM)-based parameter iterative approach is proposed to estimate the massive MIMO channel based on sparse Bayesian method. With a large antenna array, the massive MIMO channel matrix is approximate sparse in the beam domain. We model the channel non-zero distribution as Gaussian mixture (GM) and introduce expectation maximization (EM) algorithm to learn all the required channel parameters. And all of the parameters needed for EM update are computed by the general approximate message passing (GAMP) [9, 10] algorithm, which provides a huge gain in reducing the computational complexity. Hence, our approach does not require any channel parameters, which can adaptively adjust channel parameters by using EM algorithm. Numerical results show that the proposed scheme can accurately estimate the massive MIMO channel with a much reduced complexity while channel distribution is unknown. Compared with the OMP algorithm, the proposed EM-GAMP algorithm has a much better performance. Moreover, the accuracy of the EM-GAMP algorithm is not affected by the number of antennas at the BS.

The rest of this paper is organized as follows. In Section 2, we describe the system model and introduce the sparse Bayesian channel estimation. In Section 3, we propose an EM-based parameter iterative approach to estimate the massive MIMO channel based on sparse Bayesian method. The performance evaluation is described in Section 4. Finally, we conclude the paper in Section 5.

Notations. We use boldface capital letters like $\mathbf{A}$ to denote matrices. For any matrix $\mathbf{A} \in \mathbb{C}^{N \times M}$, $\mathbf{A}_{nm}$ denotes the (n, m)th element of $\mathbf{A}$. $\mathbf{A}^{\mathsf{T}}$ and $\mathbf{A}^{\mathsf{H}}$ are the transpose and the conjugate transpose of $\mathbf{A}$, respectively. $\mathbf{I}_n$ denotes the identity matrix of size $n \times n$. For a Gaussian random vector $\mathbf{x}$ with mean μ and variance Ω, we denote the pdf by $\mathcal{N}(x; \mu, \Omega)$. In addition, we use $\mathcal{N}_{\mathbb{C}}(x; \mu, \Omega)$ to denote its circular complex Gaussian counterpart. $E\{\cdot\}$ and $\delta(\cdot)$ denote the expectation operator and Dirac delta, respectively.

2 Sparse Bayesian channel estimation
2.1 Sparse channel model
Consider a multi-user massive MIMO system with one cell. There is one BS and K users in the network, where the BS is equipped with M antennas and each user has a single antenna. During the downlink training phase, the BS simultaneously transmits a sequence of N training pilot symbols to all users. Denote the pilot sequences as a matrix $\mathbf{S} \in \mathbb{C}^{N \times M}$. The channel matrix between the users and the BS is denoted by $\mathbf{G} \triangleq [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_K] \in \mathbb{C}^{M \times K}$, where $\mathbf{g}_k$ is the channel vector from the k-th users to the BS. The received signal $\mathbf{Y}$ at the user side can be expressed as

$$\mathbf{Y} \triangleq \mathbf{SG} + \mathbf{W}, \tag{1}$$

where $\mathbf{W} \in \mathbb{C}^{N \times K}$ is the white Gaussian noise.

Accurate CSI is the basic of resource allocation, precoding, and other operations in massive MIMO systems. The common channel estimation algorithms, such as least square (LS) and minimum mean square error (MMSE) methods, usually exploit the pilot sequence $\mathbf{S}$ and the received signal $\mathbf{Y}$ to reconstruct the channel. However, due to massive antennas at the BS, traditional methods cannot be applied to massive MIMO channel estimation directly. Because the pilot overhead, which is proportional to the number of antennas at the BS, will be prohibitively high. Current researches show that the massive MIMO channels exhibit the sparsity in the beam domain. Wen et al. [9] points out that the channel vector $\mathbf{g}_k$ can be written as

$$\mathbf{g}_k^{\mathsf{T}} \triangleq \mathbf{v}_k^{\mathsf{T}} \mathbf{R}_k^{\frac{1}{2}}, \tag{2}$$

where $\mathbf{R}_k$ is the channel covariance matrix and $\mathbf{v}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{v}_k; \mathbf{0}, \mathbf{I}_M)$. After taking the discrete Fourier transform (DFT) [11] of $\mathbf{g}_k$, we have

$$\mathbf{h}_k^{\mathsf{T}} \triangleq \mathbf{g}_k^{\mathsf{T}} \mathbf{F} = \mathbf{v}_k^{\mathsf{T}} \mathbf{R}_k^{\frac{1}{2}} \mathbf{F}, \tag{3}$$

where $\mathbf{h}_k^{\mathsf{T}}$ is the channel representation of $\mathbf{g}_k^{\mathsf{T}}$ in beam domain and $\mathbf{F}$ is the DFT matrix. The channel matrix in beam domain is proved to be approximate sparse [12]. With the number of antennas at the BS increasing, the channel covariance matrix $\mathbf{R}_k$ can be represented by

$$\mathbf{R}_k \triangleq \mathbf{F} \mathbf{\Lambda}_k \mathbf{F}^{\mathsf{H}}, \tag{4}$$

where $\mathbf{\Lambda}_k$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{R}_k$.

Plugging (4) into (3) and simplifying, we can obtain the sparse channel matrix

$$\mathbf{h}_k^{\mathsf{T}} = \mathbf{v}_k^{\mathsf{T}} \mathbf{R}_k^{\frac{1}{2}} \mathbf{F} = \mathbf{v}_k^{\mathsf{T}} \mathbf{F} \mathbf{\Lambda}_k^{\frac{1}{2}}. \tag{5}$$

We can get the received signal in the beam domain by taking the DFT of (1) both side: $\mathbf{Y}^B \triangleq \mathbf{SH} + \mathbf{W}^B$, where $\mathbf{Y}^B = \mathbf{YF}$, $\mathbf{H} = \mathbf{GF}$, $\mathbf{W}^B = \mathbf{WF}$. Simulation results

in [12] point out that over 99% of the channel power is located only within 16% of the beam indices, so we can conclude that $\mathbf{H}$ is approximately sparse. The CS techniques can reconstruct the sparse signal from a small set of linear measurements; thus, it is feasible for massive MIMO channel estimation.

2.2 Sparse Bayesian channel estimation

The received signal vector at the kth user can be expressed as

$$\mathbf{y}_k^B = \mathbf{S}\mathbf{h}_k + \mathbf{w}_k^B. \tag{6}$$

Each element of sparse vector $\mathbf{h}_k$ is a Gaussian random variable with different variances, so we assume the elements of $\mathbf{h}_k$ follow Gaussian-mixture distributions:

$$
\begin{aligned}
P_{\mathbf{h}_k}(h_{km}; \lambda_k, \boldsymbol{\rho}_k, \boldsymbol{\theta}_k, \boldsymbol{\sigma}_k) &\triangleq (1 - \lambda_k)\delta(h_{km}) \\
&+ \lambda_k \sum_{l=1}^{L} \rho_{kl} \mathcal{N}(h_{km}; \theta_{kl}, \sigma_{kl}),
\end{aligned}
\tag{7}
$$

where λ_k is the sparsity rate and L is the number of mixture components; $\boldsymbol{\rho}_k \triangleq [\rho_{k1}, \rho_{k2}, \ldots, \rho_{kL}]^\mathsf{T}$, $\boldsymbol{\theta}_k \triangleq [\theta_{k1}, \theta_{k2}, \ldots, \theta_{kL}]^\mathsf{T}$ and $\boldsymbol{\sigma}_k \triangleq [\sigma_{k1}, \sigma_{k2}, \ldots, \sigma_{kL}]^\mathsf{T}$ are the weight, mean and variance matrix of the k-th GM component, respectively. In addition, $\sum_{l=1}^{L} \rho_{kl} = 1$. We assume that the noise $\mathbf{w}_k^B$ is independent to the channel $\mathbf{h}_k$ and is independent and identically distributed (i.i.d.) Gaussian with mean zero and variance φ_k, so we have

$$P_{\mathbf{w}_k}(w_k; \varphi_k) \triangleq \mathcal{N}(w_k; 0, \varphi_k). \tag{8}$$

And, the channel $\mathbf{h}_k$ also contains i.i.d. components, so we have

$$P_{\mathbf{h}_k}(\mathbf{h}_k; \mathbf{q}_k) = \prod_{m=1}^{M} P_{\mathbf{h}_k}(h_{km}; \mathbf{q}_k), \tag{9}$$

where $\mathbf{q}_k \triangleq [\lambda_k, \boldsymbol{\rho}_k, \boldsymbol{\theta}_k, \boldsymbol{\sigma}_k, \varphi_k]$. By exploiting the sparse Bayesian learning method to estimate the channel $\mathbf{h}_k$, we can get

$$\hat{h}_{km} = \int h_{km} Q(h_{km}) dh_{km}, \tag{10}$$

where $Q(h_{km}) = \int P_{\mathbf{h}_k | \mathbf{y}_k^B}(\mathbf{h}_k | \mathbf{y}_k^B; \mathbf{q}_k) \prod_{i \neq m} dh_{ki}$ is the marginal probability density function (PDF) of h_{km}.

As we can see in (10), the Bayesian algorithm is computationally intractable and it requires to know the channel statistical properties in advance. However, it is difficult to obtain in massive MIMO systems. So, new methods should be developed to solve these problems. In this paper, we propose an EM-based parameter iterative approach to estimate sparse channel by employing the GAMP to reduce the computational complexity and the EM algorithm to learn all the required channel knowledge.

3 EM-based parameter iterative approach for sparse Bayesian channel estimation

In massive MIMO systems, the channel statistical property is getting harder to obtain with the increasing number of antennas at the BS. Unlike traditional Bayesian method, this paper proposes an EM-based parameter iterative approach to learn all the required channel knowledge.

3.1 EM-GAMP algorithm

We introduce the EM algorithm to learn the channel parameters $\mathbf{q}_k$. The EM algorithm constantly updates $\mathbf{q}_k$ by increasing the lower bound on the likelihood $\ln P_{\mathbf{h}_k | \mathbf{y}_k^B}(\mathbf{h}_k | \mathbf{y}_k^B; \mathbf{q}_k)$ at each iteration [13]. The updating rules are given as:

$$\hat{\mathbf{q}}_k^{t+1} = \arg \max_{\mathbf{q}_k} E\left\{\ln P_{\mathbf{h}_k^t | \mathbf{y}_k^B}\left(\mathbf{h}_k^t | \mathbf{y}_k^B; \hat{\mathbf{q}}_k^t\right)\right\}, \tag{11}$$

where t represents the iteration index and $\ln P_{\mathbf{h}_k^t | \mathbf{y}_k^B}\left(\mathbf{h}_k^t | \mathbf{y}_k^B; \hat{\mathbf{q}}_k^t\right)$ can be calculated by the GAMP algorithm.

Denote $\mathbf{z}_k^t = \mathbf{S}\mathbf{h}_k^t$, $z_{kn}^t = \mathbf{s}_n^\mathsf{T} \mathbf{h}_k^t$, where $\mathbf{s}_n^\mathsf{T}$ is the nth row of $\mathbf{S}$. The GAMP algorithm models the relationship between $\mathbf{y}_{kn}^B$ and z_{kn}^t by the conditional PDF. The true marginal posterior can be estimated by the whole probability formula

$$
P_{\mathbf{z}_k^t | \mathbf{y}_k^B}\left(z_{kn}^t | \mathbf{y}_k^B; \mathbf{q}_k^t\right) = \frac{P_{\mathbf{y}_k^B | \mathbf{z}_k^t}\left(y_{kn}^B | z_{kn}^t; \mathbf{q}_k^t\right) \mathcal{N}\left(z_{kn}^t; \hat{p}_{kn}^t, \mu_{kn}^{p,t}\right)}{\int_{\mathbf{z}_k^t} P_{\mathbf{y}_k^B | \mathbf{z}_k^t}\left(y_{kn}^B | z_k^t; \mathbf{q}_k^t\right) \mathcal{N}\left(z_k^t; \hat{p}_{kn}^t, \mu_{kn}^{p,t}\right)},
\tag{12}
$$

where $\hat{p}_{kn}^t$ and $\mu_{kn}^{p,t}$ are the mean and variance of $\mathbf{z}_k^t$, respectively. The noise vector $\mathbf{w}_k^t = \mathbf{y}_k^B - \mathbf{z}_k^t$ is assumed to be additive white Gaussian, so we can get $P_{\mathbf{y}_k^B | \mathbf{z}_k^t}\left(y_{kn}^B | z_{kn}^t; \mathbf{q}_k^t\right) = \mathcal{N}\left(y_{kn}^B; z_{kn}^t, \varphi_k^t\right)$. Plugging it into (12), we can obtain

$$
\begin{aligned}
P_{\mathbf{z}_k^t | \mathbf{y}_k^B}\left(z_{kn}^t | \mathbf{y}_k^B; \mathbf{q}_k^t\right) &= \frac{\mathcal{N}\left(y_{kn}^B; z_{kn}^t, \varphi_k^t\right) \mathcal{N}\left(z_{kn}^t; \hat{p}_{kn}^t, \mu_{kn}^{p,t}\right)}{\int_{\mathbf{z}_k^t} \mathcal{N}\left(y_{kn}^B; \mathbf{z}_k^t, \varphi_k^t\right) \mathcal{N}\left(\mathbf{z}_k^t; \hat{p}_{kn}^t, \mu_{kn}^{p,t}\right)} \\
&= \mathcal{N}\left(z_{kn}^t; \frac{\mu_{kn}^{p,t} y_{kn}^B + \hat{p}_{kn}^t \varphi_k^t}{\mu_{kn}^{p,t} + \varphi_k^t}, \frac{\mu_{kn}^{p,t} \varphi_k^t}{\mu_{kn}^{p,t} + \varphi_k^t}\right).
\end{aligned}
\tag{13}
$$

The posterior PDF of h_{km}^t could be approximated by the GAMP algorithm

$$
P_{\mathbf{h}_k^t | \mathbf{y}_k^B}\left(h_{km}^t | \mathbf{y}_k^B; \hat{r}_{km}^t, u_{km}^{r,t}, \mathbf{q}_k^t\right) = \frac{P_{\mathbf{h}_k^t}\left(h_{km}^t; \mathbf{q}_k^t\right) \mathcal{N}\left(h_{km}^t; \hat{r}_{km}^t, u_{km}^{r,t}\right)}{\int_{\mathbf{h}_k^t} P_{\mathbf{h}_k^t}\left(\mathbf{h}_k^t; \mathbf{q}_k^t\right) \mathcal{N}\left(\mathbf{h}_k^t; \hat{r}_{km}^t, u_{km}^{r,t}\right)},
\tag{14}
$$

where $\hat{r}_{km}^t$ and $u_{km}^{r,t}$ are the mean and variance of $\mathbf{h}_k^t$, respectively.

Taking the prior information (7) into (14), we have

$$P_{\mathbf{h}_k^t|\mathbf{y}_k^B}\left(h_{km}^t|\mathbf{y}_k^B;\hat{r}_{km}^t, u_{km}^{r,t}, \mathbf{q}_k^t\right) = \left(1 - \pi_{km}^t\right)\delta\left(h_{km}^t\right)$$
$$+ \pi_{km}^t \Sigma_{l=1}^L \bar{\beta}_{km,l}^t \mathcal{N}\left(h_{km}^t; \gamma_{km,l}^t, \upsilon_{km,l}^t\right), \quad (15)$$

where

$$\pi_{km}^t = \int_{\mathbf{h}_k^t} P_{\mathbf{h}_k^t}\left(\mathbf{h}_k^t; \mathbf{q}_k^t\right)\mathcal{N}\left(\mathbf{h}_k^t; \hat{r}_{km}^t, u_{km}^{r,t}\right), \quad (16a)$$

$$\beta_{km,l}^t = \lambda_k^t \rho_{kl}^t \mathcal{N}\left(\hat{r}_{km}^t; \theta_{kl}^t, u_{km}^{r,t} + \sigma_{kl}^t\right), \quad (16b)$$

$$\bar{\beta}_{km,l}^t = \frac{\beta_{km,l}^t}{\Sigma_{l=i}^L \beta_{km,i}^t}, \quad (16c)$$

$$\pi_{km}^t = \left(1 + \left(\frac{\Sigma_{l=i}^L \beta_{km,i}^t}{\left(1 - \lambda_k^t\right)\mathcal{N}\left(0; \hat{r}_{km}^t, u_{km}^{r,t}\right)}\right)^{-1}\right)^{-1}, (16d)$$

$$\gamma_{km,l}^t = \frac{\hat{r}_{km}^t/u_{km}^{r,t} + \theta_{kl}^t/\sigma_{kl}^t}{1/u_{km}^{r,t} + 1/\sigma_{kl}^t}, \quad (16e)$$

$$\upsilon_{km,l}^t = \frac{1}{1/u_{km}^{r,t} + 1/\sigma_{kl}^t}. \quad (16f)$$

$P_{\mathbf{h}_k^t|\mathbf{y}_k^B}\left(h_{km}^t|\mathbf{y}_k^B; \mathbf{q}_k^t\right)$ obeys a GM distribution according to (15). Plugging (15) into $Q\left(h_{km}^t\right)$ and estimating by the GAMP algorithm, we find that $Q\left(h_{km}^t\right)$ could be viewed as a Gaussian distribution. So it can be written as $Q\left(h_{km}^t\right) = \mathcal{N}\left(h_{km}^t; a_{km}^t, var_{km}^t\right)$, where a_{km}^t and var_{km}^t are the mean and variance, respectively. And then plugging it into (10), the posterior estimate of h_{km}^t can be obtained as a_{km}^t.

Plugging (15) into (11) and following similar steps of manipulation in [13], we have the update for the parameters

$$\lambda_k^{t+1} = \frac{1}{M}\Sigma_{m=1}^M \pi_{km}^t, \quad (17a)$$

$$\rho_{kl}^{t+1} = \frac{\Sigma_{m=1}^M \pi_{km}^t \bar{\beta}_{km,l}^t}{\Sigma_{m=1}^M \pi_{km}^t}, \quad (17b)$$

$$\theta_{kl}^{t+1} = \frac{\Sigma_{m=1}^M \pi_{km}^t \bar{\beta}_{km,l}^t \gamma_{km,l}^t}{\Sigma_{m=1}^M \pi_{km}^t \bar{\beta}_{km,l}^t}, \quad (17c)$$

$$\sigma_{kl}^{t+1} = \frac{\Sigma_{m=1}^M \pi_{km}^t \bar{\beta}_{km,l}^t \left(|\theta_{kl}^t - \gamma_{km,l}^t|^2 + \upsilon_{km,l}^t\right)}{\Sigma_{m=1}^M \pi_{km}^t \bar{\beta}_{km,l}^t}, \quad (17d)$$

$$\varphi_k^{t+1} = \frac{1}{N}\Sigma_{n=1}^N \left(|y_{kn}^B - \hat{z}_{kn}^t|^2 + \mu_{kn}^{z,t}\right), \quad (17e)$$

where π_{km}^t, $\bar{\beta}_{km,l}^t$, $\gamma_{km,l}^t$, $\upsilon_{km,l}^t$ are calculated by (16) and $\mu_{kn}^{z,t}$ is the variance of $P_{\mathbf{z}_k^t|\mathbf{y}_k^B}\left(z_{kn}^t|\mathbf{y}_k^B; \mathbf{q}_k^t\right)$, namely, $\mu_{kn}^{z,t} = \mu_{kn}^{p,t}\varphi_k^t/\left(\mu_{kn}^{p,t} + \varphi_k^t\right)$. As we can see from (17), accurate channel parameters could be achieved after several iterations.

Statistical properties of the adjacent sub-channels change slowly in beam domain because of the limited

scatters. So, we can use the learned parameters to initial the next sub-channel parameters, namely, $\hat{\mathbf{q}}_{k+1}^0 = \hat{\mathbf{q}}_k^t$. The proposed EM-GAMP algorithm is summarized in Algorithm 1.

Algorithm 1 Proposed EM-GAMP Algorithm.

Input: $P_{\mathbf{q}_k}(\cdot)$, $\mathbf{S} = \{S_{mn}\}$, $\mathbf{y}_k^B$, ε_{gamp}, L, ε_{em}.
Initialization: Set $\hat{\mathbf{q}}_1^0 = [\lambda_1^0, \boldsymbol{\rho}_1^0, \boldsymbol{\theta}_1^0, \boldsymbol{\sigma}_1^0, \varphi_1^0]$.
for $k = 1 : K$ **do**
    $t = 1$;
    repeat
      E-Step: Compute $\hat{\mathbf{h}}_k^t$ from
$P_{\mathbf{h}_k^t|\mathbf{y}_k^B}(h_{km}^t|\mathbf{y}_k^B; \hat{r}_{km}^t, u_{km}^{r,t}, \mathbf{q}_k^t)$ as described in (15)
      by exploiting the GAMP algorithm
      M-Step: Update $\hat{\mathbf{q}}_k^t$ using (17)
      $t = t + 1$
    until $\Sigma_{m=1}^M |\hat{h}_{km}^t - \hat{h}_{km}^{t-1}|^2 < \varepsilon_{em}$
    $\hat{\mathbf{q}}_{k+1}^0 = \hat{\mathbf{q}}_k^t$
end for
Output: $\hat{\mathbf{H}}$

3.2 Performance analysis

The GAMP algorithm transforms a multiple integral non-linear problem into a linear problem, so the computational complexity in each iteration mainly depends on the simple matrix multiplication and addition operations in formula (10)–(16). In E-step, the computational complexity is dominated by formula (15), namely, $\vartheta(LNM)$. Similarly, the computation of the M-step is $\vartheta(4M + N)$. Therefore, the proposed EM-GAMP algorithm has the computational complexity on the order of $\vartheta(T(LNM + 4M + N))$, where T and L represent the iteration times and the number of Gaussian-mixture components, respectively.

The complexity of classic LS and OMP algorithm is $\vartheta(M^3)$ and $\vartheta(NM^2)$, respectively. In massive MIMO systems, the number of antennas at the BS is much larger than the length of pilot sequences, namely, $N \ll M$. Simulation results show that the MSE tends to be stable after five to six iterations, namely, $T \ll N$. In this paper, L is fixed to 3, which contributes to better performance. Thus, the complexity of the EM-GAMP algorithm can be simplified as $\vartheta(TNM)$. It is obvious that the EM-GAMP algorithm proposed in this paper can greatly reduce the computational complexity.

Classic CS algorithms, such as OMP and SP, can also accurately reconstruct the channel. However, these schemes need to know channel statistical properties in advance, which is unachievable in practice. In contrast to the OMP and SP, our approach exploits the EM algorithm to update all the required channel parameters as a part of the estimation procedure.

4   Simulation results

We consider a single cell massive MIMO system including a BS equipped with $M = 200$ antennas and $K = 40$ users. The BS uses $N = \lambda M$ pilot symbols for channel estimation. We denote $\lambda = 0.2$ in this paper because over 99% of the channel power is located only within 16% of the beam indices, which is mentioned in Section 2.1. It is essential for the EM algorithm to initial the unknown parameters $\hat{\mathbf{q}}_k$ properly. We initialize the parameters as $\rho_{1l}^0 = \frac{1}{3}, \theta_{1l}^0 = 0, \boldsymbol{\sigma}_1^0 = [\,0.01, 0.1, 1\,], \varphi_1^0 = \frac{\|\mathbf{y}_k^B\|^2}{N(1+SNR)}$, where SNR is the signal to noise ratio. We define the mean square error (MSE) as

$$MSE_k = \frac{1}{KM} \sum_{k=1}^{K} \sum_{m=1}^{M} |\hat{h}_{km} - h_{km}|^2 \qquad (18)$$

Figure 1 compares the MSE performance of the proposed algorithm and the traditional algorithms as functions of *SNR*. As can be seen from the picture, the MSE of all algorithms decrease with *SNR* increasing. Clearly, the proposed EM-GAMP has much better performance than the LS and OMP and it approaches the perfect CSI bound when $SNR > 15dB$. In addition, we also compare the MSE performance with different GM components. The MSE would reduce when the number of GM components increases. This is because the channel model could be more accurate with more GM components.

Figure 2 compares the MSE performance of the EM-GAMP algorithm as functions of the number of iterations. From the picture, we can see that the MSE tends to be stable after five to six iterations. And the convergence rate is inversely proportional to *L*, the number of GM components. However, the MSE performance degrades with small *L*. Therefore, it is necessary to choose appropriate *L*
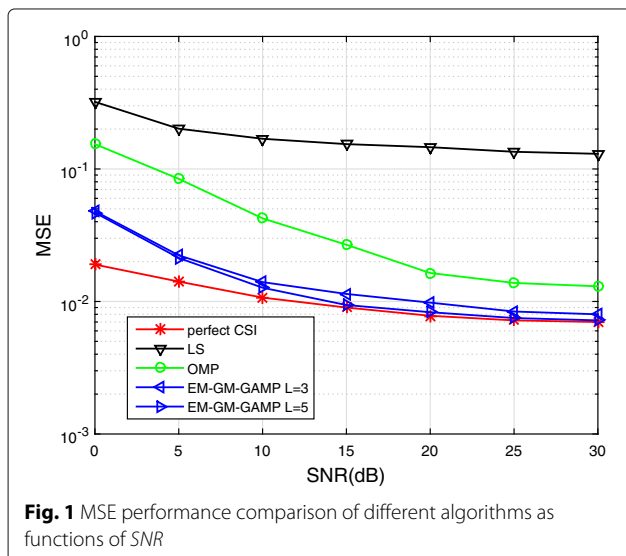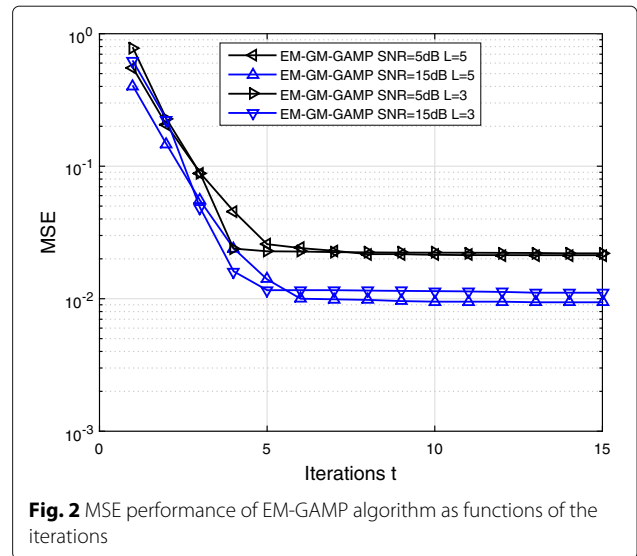

Fig. 2 MSE performance of EM-GAMP algorithm as functions of the iterations

to improve the operation rate as much as possible within the margin of error.

Figure 3 compares the MSE performance of the EM-GAMP, the LS, and the OMP as functions of M with $SNR = 15dB$, $L = 3$, $\frac{N}{M} = 0.2$. From Fig. 3, it is clearly that performance of the LS and OMP degrades with M increasing. By contrast, the proposed EM-GAMP algorithm is not affected by the number of antennas at the BS. The estimation error keeps stable no matter how large M is. This is because that we exploit the EM algorithm to learn and update the channel parameters adaptively, which makes our algorithm adapt to different environments.

Figure 4 compares the MSE performance of different channels and M as functions of $\frac{N}{M}$ with $SNR = 15dB$, $\lambda = 0.2$ and $L = 3$. From the results in Fig. 4, when
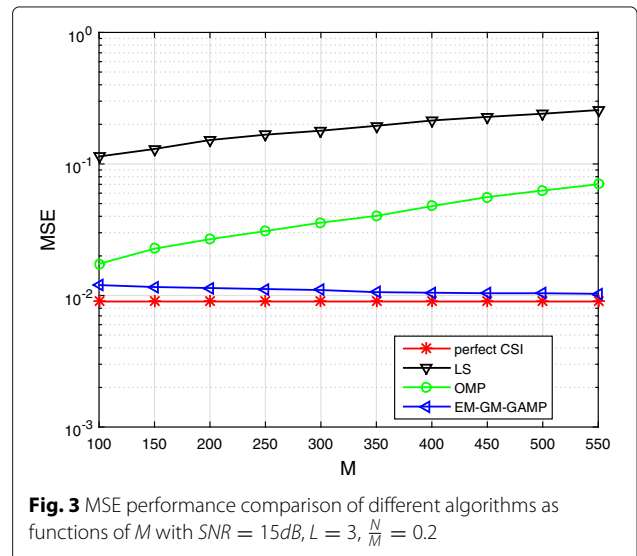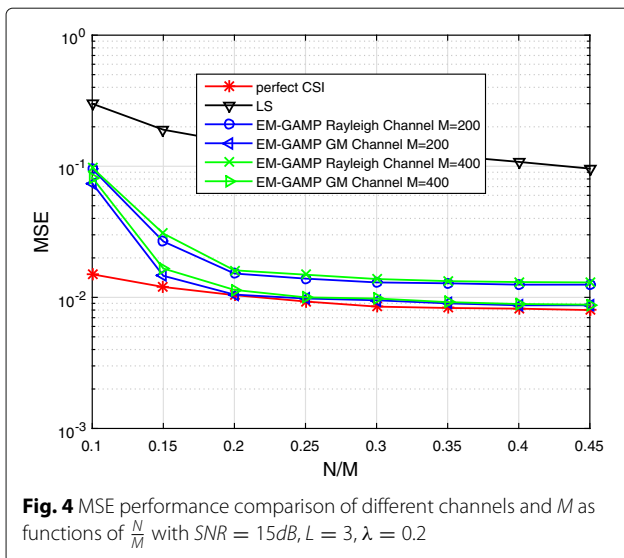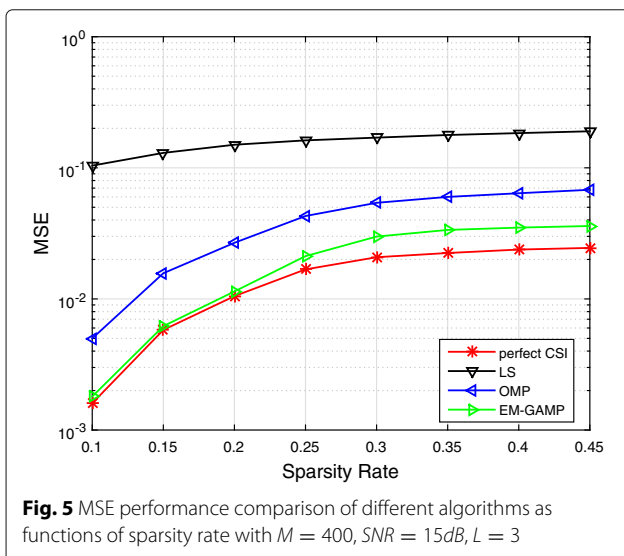

Fig. 1 MSE performance comparison of different algorithms as functions of *SNR*

Fig. 3 MSE performance comparison of different algorithms as functions of M with $SNR = 15dB$, $L = 3$, $\frac{N}{M} = 0.2$

Fig. 4 MSE performance comparison of different channels and M as functions of $\frac{N}{M}$ with $SNR = 15dB, L = 3, \lambda = 0.2$

5 Conclusions

To obtain an accurate CSI of massive MIMO system, we propose an EM-based parameter iterative approach based on sparse Bayesian method. Channel statistical properties are essential for sparse Bayesian methods to obtain accurate CSI; however, it is barely accessible in massive MIMO systems due to the larger number of antennas at the BS. In this paper, we model the channel non-zero distribution as a Gaussian mixture and employ the EM algorithm to learn all the required channel parameters. Besides, all of the parameters needed for EM update are computed by the GAMP algorithm. Simulation results show that our approach provides a huge gain in inducing complexity and has a much better performance compared to the LS and OMP algorithms with channel parameters unknown. In addition, the proposed EM-GAMP algorithm is robust to the number of antennas at the BS.

Pilot contamination exists when the same orthogonal pilot sequences used in all cells for multi-cell massive MIMO systems. We have designed a kind of non-orthogonal pilot sequences [14], which can efficiently reduce the impact of pilot contamination. Thus, the proposed method can be extended to multi-cell massive MIMO systems and still have a good performance and complexity reducing.

increasing $\frac{N}{M}$, it is observed that the MSE decreases and tends to be stable when $N \geq \lambda M$ as excepted. Different channel models, including Rayleigh channel model and Gaussian mixture channel model, are considered. The proposed approach achieves much better performance in all cases illustrating a substantial improvement over the LS estimator.

Figure 5 compares the MSE performance of different algorithms as functions of sparsity rate with $M = 400$, $SNR = 15dB, L = 3$. The performance of all of the three algorithms increases with the sparsity rate getting larger. However, the EM-GAMP algorithm approaches the perfect CSI when the sparsity rate $\lambda \leq 0.2$, and it always achieves a much better performance than others.

Authors' contributions
MS conceived and designed the study. MS performed the experiments and wrote the paper. FY reviewed and revised the manuscript. Both authors read and approved the manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. EG Larsson, F Tufvesson, O Edfors, TL Marzetta, Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)
2. F Kaltenberger, D Gesbert, R Knopp, M Kountouris, Correlation and capacity of measured multi-user MIMO channels. Proc. IEEE Int. Symp. Personal Indoor Mobile Radio Commun. (PIMRC), 1–5 (2008)
3. J Hoydis, C Hoek, T Wild, S ten Brink, Channel measurements for large antenna arrays. Proc. IEEE Int. Symp. Wireless Commun. Systems (ISWCS). **91**(12), 811–815 (2012)
4. L Dai, Z Wang, Z Yang, Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems. IEEE J. Sel. Areas Commun. **31**(2), 251–263 (2013)
5. Z Gao, L Dai, Z Wang, S Chen, Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO. IEEE Trans. Signal Process. **63**(23), 6169–6183 (2015)
6. Z Gao, L Dai, Z Wang, Structured compressive sensing based superimposed pilot design for large-scale MIMO systems. Electron. Lett. **50**(12), 896–898 (2014)

Fig. 5 MSE performance comparison of different algorithms as functions of sparsity rate with $M = 400, SNR = 15dB, L = 3$

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

8. C Zhu, Z Zheng, B Jiang, W Zhong, X Gao, Bayesian channel estimation for massive MIMO communications. IEEE 83rd Veh. Technol. Conf, 1–5 (2016)

9. C-K Wen, K-K Wong, J-C Chen, P Ting, Channel estimation for massive MIMO using gaussian-mixture bayesian learning. IEEE Trans. Commun. **14**(3), 1356–1368 (2015)

10. S Rangan, Generalized approximate message passing for estimation with random linear mixing. Proc. IEEE Int. Symp. Inf. Theory (ISIT), 2168–2172 (2011)

11. J Zhang, B Zhang, S Chen, X Mu, M El-Hajjar, L Hanzo, Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems. IEEE J. Sel. Topics Signal Process. **8**(5), 759–772 (2014)

12. E Bjonson, J Hoydis, M Kountouris, M Debbah, Massive MIMO systems with non-ideal hardware: Energy efficiency estimation and capacity limits. IEEE Trans. Inf. Theory. **60**(11), 7112–7139 (2014)

13. JP Vila, P Schiner, Expectation-maximization Gaussian-mixture approximate message passing. IEEE Trans. Signal Process. **61**(19), 4658–4672 (2013)

14. F Yong, M Sulin, W Du, Pilot assignment of massive mimo systems for pilot decontamination based on inter-cell cross gain. Processings 2016 Int. Conf. Sci. Comput, 125–129 (2016)