Downlink Channel Estimation for Massive MIMO Systems Relying on Vector Approximate Message Passing

Sheng Wu, Haipeng Yao, Chunxiao Jiang, Xi Chen, Linling Kuang, and Lajos Hanzo

Abstract—To reduce the pilot overhead of downlink channel estimation in massive multiple-input–multiple-output (MIMO) systems, a sparse recovery algorithm relying on the vector approximate message passing (VAMP) technique is proposed. More specifically, an a-priori channel model characterized by a multivariate Bernoulli-Gaussian distribution is invoked for exploiting the common sparsity of massive MIMO channels, and the VAMP technique is used for jointly estimating the spatially correlated channels. Moreover, the hyperparameters of the a-priori model are learned by invoking the expectation maximization (EM) algorithm. Our numerical results demonstrate that the proposed algorithm is capable of reducing the pilot overhead by 50% in massive MIMO systems.

Index Terms—Block sparsity, Channel estimation, Massive MIMO, OFDM, Vector Approximate Message Passing.

I. INTRODUCTION

To maximize the gain of massive multiple-input–multiple-output (MIMO) systems, having accurate channel state information (CSI) at the base station (BS) is essential [1]. In frequency division duplex (FDD) systems, the CSI is typically obtained by downlink (DL) channel estimation and then fed back to the BS [2], [3]. However, the conventional least squares channel estimation method requires a high pilot overhead that increases linearly with the number of BS antennas and the length of channel impulse response (CIR), and becomes inefficient in massive MIMO systems.

In recent years, compressed sensing (CS)-based channel estimation has attracted much attention due to its ability to reduce required pilots. In [4], the sparsity of the massive MIMO CIR taps was exploited by Rao and Lau. As a further advance, Rao and Lau [5] studied CS-based channel estimation in the presence of temporal correlation. Gao et al. [1] proposed an adaptive structured subspace pursuit based algorithm for exploiting both the common spatial and temporal sparsity in massive MIMO channels. Then, a distributed CS-based algorithm was studied by Gong et al. [6] for leveraging common sparsity in doubly-selective channels. Furthermore, a turbo-CS algorithm relying on the Markov chain prior was proposed by Chen et al. [7] for estimating MIMO channels that are structured sparse in the angular domain. Moreover, many message passing algorithms [8], [9], [10], [11], [12] have been applied to MIMO channel estimation. Mo et al. used the generalized AMP (GAMP) algorithm for exploiting the joint sparsity of the mmWave MIMO channel both in the angle and in the delay domain [13]. Huang et al. [14] proposed an iterative channel estimation algorithm based on the least square estimation and sparse message passing algorithm invoked for mmWave MIMO systems.

Against this background, we exploit the common sparsity of massive MIMO channels in the delay domain and conceive a channel estimator that intrinsically amalgamates the vector approximate message passing (VAMP) technique of Rangan et al. [12] and the EM algorithm [15] of Neal and Hinton. Specifically, an a-priori channel model characterized by a multivariate Bernoulli-Gaussian distribution is employed for exploiting the common sparsity, and the VAMP relying in a non-separable denoiser is used for jointly estimating the spatially correlated channels. At the same time, the hyperparameters of the model are learned by invoking the expectation maximization (EM) algorithm. Our numerical results demonstrate that the proposed algorithm can reduce the pilot overhead of massive MIMO systems by up to 50%, while maintaining superior performance approaching the idealized bound relying on perfectly known channel supports and hyperparameters.

Throughout the paper, the superscript T denotes the transpose operation, and H represents the conjugate transpose operation. \( \| \cdot \|_2 \) and \( \| \cdot \|_F \) return the Euclidean norm and Frobenius norm of a vector or a matrix, respectively. \( N_{\mathbb{C}}(x, \hat{x}, \Sigma) \) represents the Gaussian distribution function of a complex random vector \( x \) with mean \( \hat{x} \) and covariance matrix \( \Sigma \). Furthermore, \( \langle \cdot \rangle \) is the empirical averaging operation \( \langle x \rangle = \frac{1}{N} \sum_{n=1}^{N} x_n \), and \( I \) is an identity matrix. Finally, \( \text{Tr} (\cdot) \) returns the trace of a matrix, and \( \text{Diag} (\cdot) \) creates a diagonal matrix with a vector or get the diagonal elements of a matrix.

II. SYSTEM MODEL

Consider a FDD massive MIMO-OFDM system employing \( K \) OFDM subcarriers, where the BS has \( M \) antennas, while each user has only one antenna. Let \( h_m = [h_{m1}, \ldots, h_{mL}]^\top \) denote the DL CIR taps from the \( m \)th transmit antenna (TA)
to a user, where $h_{ml}$ denotes the $l$th CIR tap and $L$ denotes the length of CIR taps. According to the analysis in [16], $h_m$ and $h_{m'}$ with $m \neq m'$ tend to share an identical support set in the delay domain if $d_{\text{max}} \leq \frac{C}{\text{BW}}$, where $d_{\text{max}}$ is the farthest distance between the TAs, $C$ is the speed of light, and $\text{BW}$ is the signal bandwidth. All the $l$th DL CIR taps from the $M$ TAs to the user are denoted by $h_{l} = [h_{l1}, \ldots, h_{lM}]^T$. Due to having a limited number of CIR-tap clusters, $h_{l}$ is spatially correlated. Therefore, the Kronecker product correlation model [17][18] is applied for characterizing $h_{l}$, which is formulated as $h_{l} = (R_{lz} \otimes R_{lz})^{\frac{1}{2}}h_{l}^{id}$, where $R_{lz} \otimes R_{lz}$ denote the Kronecker product of the azimuth correlation matrix $R_{lz}$ and the elevation correlation matrix $R_{lz}$, while $h_{l}^{id}$ obeys the Gaussian distribution $\mathcal{N}([h_{l}^{id}; 0, \alpha_{l}I])$, where $\alpha_{l}$ is the variance of the $l$th CIR tap.

The pilot sequences transmitted over different TAs are randomly and independently generated, denoted by $x_{m} \in \mathbb{C}^{N \times L}$, for all $m$, but occupy the same subcarriers indexed by the set $\mathcal{P}$. Let $y \in \mathbb{C}^{N \times L}$ denote the samples after discarding the cyclic prefix and the fast Fourier transformation (FFT) at the receiver, then

$$y = \sum_{m=1}^{M} \Psi_{m} x_{m} + \sigma = \sum_{l=1}^{L} \Phi_{l} h_{l} + \sigma = \Phi h + \sigma,$$  
where $\Psi_{m} = \text{diag}(x_{m}) F_{P}$, $F_{P} \in \mathbb{C}^{N \times L}$ is comprised of the first $L$ columns of the $K$-point discrete Fourier transformation (DFT) matrix and the $N$ rows of the $K$-point DFT matrix specified by the indices in set $\mathcal{P}$, $\sigma = [\sigma_{1} \cdots \sigma_{ML}]^T \in \mathbb{C}^{L \times 1}$ is the complex Gaussian noise obeying the distribution $p(\sigma) \sim \mathcal{N}(0, \sigma I)$. $\Phi_{l} = [\varphi_{m} \cdots \varphi_{ML}] \in \mathbb{C}^{N \times M}$ stacks $\varphi_{ml}$ being the $l$th column of matrix $\Psi_{m}$, $\Phi = [\Phi_{1} \cdots \Phi_{L}] \in \mathbb{C}^{N \times ML}$, and $h = [h_{1}^{T} \cdots h_{L}^{T}]^T \in \mathbb{C}^{ML \times 1}$.

### III. SPARSE CHANNEL ESTIMATION RELYING ON VAMP

To characterize the common sparsity of $h$, i.e., all the CIR taps in $h$, $\forall l$ are either zero simultaneously or non-zero simultaneously, a multivariate Bernoulli-Gaussian $a$-priori model [19] is employed:

$$p(h; \theta_{p}) = \prod_{l=1}^{L} \left[ \lambda_l \mathcal{N}(h_{l}; 0, \tau_l I) + (1 - \lambda_l) \delta(h_{l}) \right],$$

where $\theta_{p} = [\lambda_1, \lambda_L, \tau_1, \cdots, \tau_L]^T$ contains the hyperparameters of the $a$-priori channel model, $\lambda_{l} \in (0, 1)$ denotes the $a$-priori sparsity ratio of $h_{l}$, $\tau_{l}$ denotes the $a$-priori variance of every CIR tap in $h_{l}$ when $h_{l}$ are non-zero, and $\delta(\cdot)$ denotes the Dirac delta function. Note that as the spatial correlation matrix $R_{lz} \otimes R_{lz}$ is different across tap index $l$, modeling each $h_{l}$ with a different $a$-priori covariance matrix in (2) would lead to overfitting, and modeling all the $h_{l}$ by a common $a$-priori covariance matrix would not yield any performance gain. In view of this, the proposed prior model (2) does not capture the amplitude correlation among the taps in $h_{l}$. As the random noise $\sigma$ in (1) is Gaussian, the likelihood function of $h$ is written as

$$p(y \mid h; \sigma) = \mathcal{N}(y; \Phi h, \sigma I).$$

For the channel estimation, the $a$-posteriori distribution of $h$ can be approximately calculated by the VAMP [12], and then we can get the $a$-posteriori mean of $h$ as its estimate. However, the VAMP desires the specification of the hyperparameters in the $a$-priori model (2) and the likelihood function (3), i.e., $\theta = [\theta_{p} \sigma]^T$. To this end, the EM algorithm is used to get the specification of hyperparameters

$$Q(\theta) = \mathbb{E} \left[ \ln p(y \mid h; \sigma) + \ln p(h; \theta_{p}) \right],$$

$$(\theta(i)) = \arg \max_{\theta} Q(\theta),$$

where the expectation in (4) is w.r.t the $a$-posteriori distribution of $h$ (as shown in the following (12)) that is approximately calculated by the VAMP, and the specification of hyperparameters $\theta(i)$ is updated in the maximization step as shown by (5).

Fig. 1 shows the factor graph (for more details on factor graph, please refer to [20]) corresponding to the factorization $p(y, h; \theta_{p}, \sigma) = p(h; \theta_{p}) p(y \mid h; \sigma)$, which is comprised of the only variable node $h$, and two function nodes, namely $p(h; \theta_{p})$ and $p(y \mid h; \sigma)$. Upon invoking the VAMP and the specification of the hyperparameters, the $a$-posteriori probability of $h$ at the variable node $h$ at the $i$th iteration is given by

$$\beta_{h}(i, h) = \frac{\mathcal{N}(h; r(i), \gamma_{r}^{-1}(i) I) \prod_{l=1}^{L} p(h_{l}; \theta_{p}(i-1))}{\int_{h} \mathcal{N}(h; r(i), \gamma_{r}^{-1}(i) I) \prod_{l=1}^{L} p(h_{l}; \theta_{p}(i-1))},$$

where we have

$$\mu_{l}(i) = \frac{\tau_{l}(i) \gamma_{r}(i)}{1 + \tau_{l}(i) \gamma_{r}(i)} r_{l}(i),$$

$$\Sigma_{l}(i) = \frac{\tau_{l}(i)}{1 + \tau_{l}(i) \gamma_{r}(i)} I,$$

$$\pi_{l}(i) = \frac{\lambda_{l}(i) + (1 - \lambda_{l}(i)) \mathcal{N}(r_{l}(i); \Phi_{l}(i) \gamma_{r}^{-1}(i) I)}{\mathcal{N}(r_{l}(i); \Phi_{l}(i) \gamma_{r}^{-1}(i) I)},$$

in terms of $r_{l}(i) = [r_{l1}(i), \cdots, r_{lM}(i)]' \in \mathbb{C}^{ML \times 1}$, $r_{l}(i) \in [r_{l-1}(i+1), \cdots, r_{LM}(i)]' \in \mathbb{C}^{ML \times 1}$, and $\gamma_{r}(i)$ defined in Tab. 1. Then, the $a$-posteriori mean and covariance matrix of $h_{l}$ w.r.t (6) are given by

$$\hat{h}_{l}(i) = \pi_{l}(i) \mu_{l}(i),$$

$$C_{h_{l}}(i) = \pi_{l}(i) \Sigma_{l}(i) + \pi_{l}(i)(1 - \pi_{l}(i)) \mu_{l}(i) \mu_{l}(i)'$$

The operations of (6)-(11) show that our proposed algorithm uses an inseparable denoiser, which is different from the separable denoiser studied by most VAMP.

Then, the message emanating from the variable node $h$ to the function node $p(y \mid h; \sigma)$ is written as

$$p(h; \theta_{p})$$

$$p(y \mid h; \sigma)$$

Figure 1: Factor graph for DL channel estimation.
Finally, the derivation of $Q(\theta)$ w.r.t the $a$-priori variance is updated by

$$\eta_{\gamma}(i) = (\pi_{\gamma}(i) - \lambda_{\gamma}(i))/(1 - \lambda_{\gamma}(i)).$$

Finally, the derivation of $Q(\theta)$ w.r.t the $a$-priori variance is updated by

$$\eta_{\lambda}(i) = (\pi_{\lambda}(i) - \lambda_{\lambda}(i))/(1 - \lambda_{\lambda}(i)).$$

The proposed channel estimation algorithm is termed as "EM-BBG-VAMP", as shown by Tab. I. At the beginning of iterations, the $a$-priori variance of noise is guessed to be $\sigma(i) = ||y||_2^2/(SNR) + 1$, where the signal-to-noise ratio (SNR) is empirically set to 0 dB. Following the suggestion given by [21], the initial sparsity ratio could be set to

$$\lambda_{s}(i) = N_{max} \min_{i} \left(1 - \frac{2M}{\xi_{s}(i)} + \frac{M}{\xi_{s}(i)} \right),$$

where $B(\cdot)$ and $b(\cdot)$ denote the commutative distribution function and the probability density function of the standard Gaussian distribution, respectively. The $a$-priori variance $\tau_i$ is set to $\tau_i = (||y||_2^2 - \sigma_0(0))/\lambda_{s}(0)/||\Phi_i||_F^2 M$. After $M_{max}$ iterations, the algorithm gives $\hat{h}_f(I_{max})$ as the $a$-posteriori estimate of the MIMO channel. Tab. II compares the complexity defined by the number of floating point operations for the proposed EM-BBG-VAMP, for the subspace pursuit (SP) [22], for the distributed sparsity based adaptive matching pursuit (DSAMP) of Gao et al. [23], for the EM-BBG-EP [19], and for the EM-BG-GAMP [21], where $N_s$ is the number of supports of $h_m$, and $i_s \leq N_s$ is the index of the current stage. Note that since the SVD only has to be executed offline once, the complexity of SVD is not included in Tab. II.

### Table I: The proposed channel estimation algorithm.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Number of floating point operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP</td>
<td>$O(MNL + M^2N^2 + N^3)$</td>
</tr>
<tr>
<td>DSAMP</td>
<td>$O(MNL + M^2N^2 + N^3)$</td>
</tr>
<tr>
<td>EM-BBG-EP</td>
<td>$O(M^3 + MNL)$</td>
</tr>
<tr>
<td>EM-BG-GAMP</td>
<td>$O(MNL)$</td>
</tr>
<tr>
<td>EM-BBG-VAMP</td>
<td>$O(MNL)$</td>
</tr>
</tbody>
</table>

### IV. Numerical Simulations

Consider a FDD massive MIMO-OFDM system having $M$ BS TAs and $K = 4096$ OFDM subcarriers. The spatially correlated 3D channel model proposed in [18] is employed, and the detailed setting of the model parameters is identical to that of [19], which is omitted here owing to the page-limit. Moreover, the idealized Oracle-LMMSE relying on perfectly known supports and perfectly known noise variance is considered as a benchmark to provide the best performance bound.

Fig. 2(a) shows the relationship between the normalized mean square error (NMSE) of channel estimation and the SNR, when $M = 64$ and the pilot overhead is $N = 1024$. It is shown that the NMSE of EM-BBG-VAMP may match the oracle-bound provided by Oracle-LMMSE, when the SNR varies between 5 dB and 35 dB. On the other hand, the EM-BBG-EP joins the oracle-bound at SNR = 10 dB, while the EM-BG-GAMP requires 32.5 dB for approaching the oracle-bound. When the target of NMSE is -15 dB, the EM-BBG-VAMP gets about 12.5 dB gain over the EM-BG-GAMP. Fig. 2(b) gives the results for the case of $N = 1024$ and $M = 128,$
where the performance of EM-BBG-EP has not been evaluated due to its excessive complexity. We can find that the proposed EM-BBG-VAMP still matches the oracle-bound, considerably outperforming the SP, the DSAMP, and the EM-BG-GAMP. Both the DSAMP and EM-BG-GAMP algorithm, and the DSAMP algorithm.

Fig. 2: The NMSE versus SNR.

Fig. 3(a) shows the NMSE versus the pilot overhead for $M = 64$ and $N = 30$ dB. The EM-BBG-VAMP matches the oracle-bound even when the pilot overhead is as low as $N = 512$, while the EM-BBG-EP merges with the oracle-bound at $N = 640$. Both the DSAMP and EM-BG-GAMP merge with the oracle-bound at $N = 1152$, and the SP at $N = 1408$. As shown in Fig. 3(b), the required pilot overhead for EM-BBG-VAMP to get the targeted NMSE = -15dB is about 50% of that of the DSAMP and EM-BG-GAMP.

Fig. 4(a) shows the performance of EM-BBG-VAMP and EM-BG-AMP with different iteration times in the case of $N = 2048$ and $M = 64$. When SNR = 25 dB, the number of iterations that the EM-BBG-VAMP requires to converge is less than 20, which is the same as that of the case of SNR = 35 dB. However, the EM-BG-AMP requires more than 60 iterations to converge at SNR = 25 dB, and more than 90 iterations at SNR = 35 dB. In the case of $N = 4096$ and $M = 128$, as shown by Fig. 4(b), the EM-BBG-VAMP requires less than 20 iterations to converge, while other algorithms require more than 60 iterations to converge.

V. CONCLUSION

We have proposed an EM-BBG-VAMP algorithm relying on the vector message passing technique for the DL channel estimation of massive MIMO systems. The proposed algorithm is capable of reducing the number of pilots in the massive MIMO system by about 50%, compared to the SP algorithm, the EM-BG-GAMP algorithm, and the DSAMP algorithm.

REFERENCES


