

Sparse Channel Estimation for Large-Scale MISO-OFDM System - A Bayesian VMP Approach

Feng LI, Hailin LI, Shuyuan LI

School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

lf1981@mail.xjtu.edu.cn

Submitted September 28, 2018 / Accepted January 23, 2019

Abstract. *The problem of channel estimation, in large-scale multiple input single output orthogonal frequency division multiplexing (MISO-OFDM) systems, is studied in this paper. In order to take full advantage of the sparse property, an intermediate random vector is introduced to control the sparsity of the estimation of the channel state information (CSI) based on the maximum a posteriori estimator. After carefully designing the prior probability density function (PDF) of the intermediate random vector and the unknown CSI conditioned on it, the sparse optimization problem over the CSI is constructed. The Bayesian inference theory is applied to relax the optimization problem by calculating an approximated PDF with simpler form. After that, variational message-passing (VMP) is used to obtain the solution in iterative analytical form. Furthermore, block sparse structure is implemented to improve the performance. Simulation results demonstrate the merit of proposed algorithm over the traditional ones.*

Keywords

Channel estimation, multiple-input-single-output (MISO), OFDM, variational Bayesian inference, variational message-passing (VMP)

1. Introduction

One of the key problem of wireless communication is to improve the spectrum efficiency due to its scarcity [1], [2]. Recently, multiple-input-multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) has gained great development to address that problem. Up-to-date research efforts have found that MIMO systems using a large number of antennas can improve the transmission performance significantly. Against this background, the large-scale MISO (multiple-input-single-output) technology has been viewed as a promising technology for current and future wireless communication systems [3]. However, the complexity of obtaining the accurate channel state information (CSI) increases heavily as the number of antennas, which

makes it difficult to use this technology in practical systems [4–6]. One important reason is that the number of unknown variables becomes so large that too much spectrum resources need to be used. Therefore, we need to study on higher efficiency CSI estimation algorithms. Fortunately, the problem can be relaxed by the fact that the outdoor wireless channel always exhibits a sparse structure, which means that only a small fraction of the paths of the impulse response of the channel have feasible values [7]. In this situation, the problem of channel estimation can be transformed into a sparse estimation problem [8].

A lot of published works have addressed the sparse estimation problem. Compressed sensing (CS) is viewed as a promising method to solve this kind of problem. Based on the sparse structure of CSI, the problem of the large-scale MISO channel estimation can be constructed as an l_1 norm constrained optimization problem known as Least Absolute Shrinkage and Selection Operator (LASSO). Also, the greedy algorithm is a useful approach to solve this, such as Orthogonal Matching Pursuit algorithm and Subspace Pursuit algorithm [9], [10]. Besides, the sparse Bayesian learning (SBL) is another kind of approach to obtain sparse solution [11]. It aims at finding a sparse maximum a posteriori (MAP) estimate of unknowns, resulting in higher accuracy of estimation.

The Bayesian sparse estimation algorithm has been proven to be an effective tool to solve the problem of sparse channel estimation in single-input-single-output (SISO)-OFDM [12]. Besides SISO-OFDM systems, recent research indicates that MISO system would also have a sparse CSI structure. What's more, it always has a common channel support between different transmit-and-receive antenna pairs, due to the similar path delay induced by the so closely located antennas [13]. Furthermore, the channel statistical properties, such as spatial covariance and time correlation, can be used to improve the estimation performance. Taking this into consideration, a block sparse structure based model of the channel vector is also studied in this paper.

The contribution of this paper is as follows:

Firstly, the optimization problem over sparse CSI estimation by means of intermediate random variables is constructed.

Secondly, both the general case and the block-sparse-structure based algorithm is introduced, using the theory of variational message passing (VMP).

Finally, an iterative analytical form of the channel estimator is obtained.

Notation: Let $\mathcal{CN}(\mathbf{y}|\mathbf{J}, \mathbf{C})$ be the complex multivariate Gaussian probability density function (PDF). Also, $\text{diag}(\mathbf{x})$ denotes the diagonal matrix created from vector \mathbf{x} . Let \mathbf{A}_m be the matrix consisted of the first m columns of matrix \mathbf{A} . Let $Ga(x|a, b) = \frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx}$ be the Gamma PDF. Let $K_\nu(\bullet)$ be the modified Bessel function of the second kind with order ν , and $\mathbf{A} \otimes \mathbf{C}$ represent the Kronecker product of the two matrices \mathbf{A} and \mathbf{C} . Finally, let $\langle f(x) \rangle_{p(x)}$ denote the expectation of function $f(x)$ with respect to the density $p(x)$.

2. System Model

We consider a downlink transmission scenario with a base station and several terminals. The base station and the terminals are equipped with N_t transmit antennas and one single received antenna, respectively. The transmission symbols are organized in frames where the preamble signals and data symbols are time multiplexed. Each frame contains one preamble signal and several data symbols. At the beginning of each frame, the pilot symbols, which can be expressed as $\mathbf{r}_l = [r_l(0), r_l(1), \dots, r_l(M-1)]^T, 1 \leq l \leq N_t$, are assigned to the M subcarriers in every transmission block, where l denotes the l -th antenna.

These pilot signals in frequency domain can be transformed into time domain by the inverse discrete fourier transform (IDFT). Therefore, the time-domain preamble signals of the l -th antenna can be written as $\mathbf{s}_l = \frac{1}{\sqrt{N_t}}\mathbf{F}^H\mathbf{r}_l, 1 \leq l \leq N_t$, where $\frac{1}{\sqrt{N_t}}$ ensures the unit transmit power, and \mathbf{F} denotes the order- M normalization DFT matrix with

$$\mathbf{F}(a, b) = e^{-\frac{2j(a-1)(b-1)\pi}{M}}.$$

Without loss of generality, in order to distinguish the preamble signals from different antennas, it is assumed that $s_l \neq s_i$ for $l \neq i$. The cyclic prefix (CP) is also inserted to avoid inter symbol interference (ISI).

After removing CP, the matrix-form received signal in time domain can be written as

$$\mathbf{q} = \sum_{l=1}^{N_t} \mathbf{H}_l \mathbf{s}_l + \mathbf{n}$$

where \mathbf{q} is the received preamble signal of length M , and \mathbf{H}_l is the cyclic matrix composed by the l -th CSI vector, i.e. l -th transmit-receive antenna pair.

The l -th CSI vector can be expressed as $\mathbf{h}_l = [h_l(0), h_l(1), \dots, h_l(L-1)]^T$, where L denotes the maximum delay spread of the CSI. Additionally, \mathbf{n} denotes the sampled noise vector, the elements of which are independent and identically distributed complex Gaussian vector with zero mean and variance σ_n^2 .

After performing DFT operation on the received signal \mathbf{q} , its frequency-domain form vector \mathbf{y} can be written as

$$\begin{aligned} \mathbf{y} &= \mathbf{F}\mathbf{q} \\ &= \sum_{l=1}^{N_t} \frac{1}{\sqrt{N_t}} \mathbf{F}\mathbf{H}_l(\mathbf{F}^H\mathbf{r}_l) + \mathbf{w}. \end{aligned} \tag{1}$$

Since \mathbf{H}_l is a cyclic matrix, (1) can be rewritten as

$$\mathbf{y} = \sqrt{\frac{N}{N_t}} \sum_{l=1}^{N_t} \text{diag}(\mathbf{F}_m \mathbf{h}_l) \mathbf{r}_l + \mathbf{w} \tag{2}$$

where $\mathbf{w} = \mathbf{F}\mathbf{n}$ is the noise vector in frequency domain. Considering the property of diagonal matrices, $\text{diag}(\mathbf{F}_m \mathbf{h}_l) \mathbf{r}_l$ in (2) can be further rewritten as $\mathbf{R}_l \mathbf{F}_m \mathbf{h}_l$ with $\mathbf{R}_l = \text{diag}(\mathbf{r}_l)$. Furthermore, by stacking all of the channel coefficients as an aggregate channel vector $\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \dots, \mathbf{h}_{N_t}^T]^T$, (2) can be expressed as

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{w} \tag{3}$$

where

$$\mathbf{A} = \sqrt{\frac{N}{N_t}} [\mathbf{R}_1 \mathbf{F}_m, \mathbf{R}_2 \mathbf{F}_m, \dots, \mathbf{R}_{N_t} \mathbf{F}_m].$$

The equation (3) describes the linear relationship between the unknown vector \mathbf{h} and the observation vector \mathbf{y} . Since $M < LN_t$, it is of high difficulty to obtain the unknowns from the observation vectors exactly.

3. Sparse Channel Estimation

In this section, we introduce a new method based on VMP algorithm to solve the problem of (3).

3.1 Bayesian Prior Model for Sparse Channel Estimation

Using maximum a posterior (MAP) algorithm, the sparse channel estimation \mathbf{h} can be written as the solution of

$$\begin{aligned} \mathbf{h}_{\text{MAP}} &= \arg \max_{\mathbf{h}} p(\mathbf{h} | \mathbf{y}) \\ &= \arg \min_{\mathbf{h}} \{-\log p(\mathbf{y} | \mathbf{h}) - \log p(\mathbf{h})\}. \end{aligned} \tag{4}$$

Since \mathbf{h} is sparse, we need to introduce some constraint conditions to utilize the sparse structure of \mathbf{h} in (4) in order to get better performance. What we need is a kind of mechanism that can control the sparsity of the solution of (4). Inspired by [12], the Bayesian prior model can be used as a useful tool to achieve this. According to this model, we introduce an intermediate random vector χ first.

Based on χ , the prior $p(\mathbf{h})$ can be demonstrated by a two-layer hierarchical structure, including a conditional prior $p(\mathbf{h}|\chi)$ and a $p(\chi)$, as shown in

$$p(\mathbf{b}) = \int p(\mathbf{b}|\chi)p(\chi)d\chi.$$

The core of this model is to find some proper distribution $p(\chi)$ and $p(\mathbf{h}|\chi)$. By carefully designing $p(\mathbf{h}|\chi)$ and $p(\chi)$, the sparsity of the solution can be controlled by χ , e.g., the larger the element of χ is, the more closer to zero the corresponding element of \mathbf{h} is. Meanwhile, when $p(\chi)$ and $p(\mathbf{h}|\chi)$ have some proper form, we can construct inference algorithms to obtain analytical expressions. Furthermore, it contributes to simplification of the computation.

Then the joint PDF can be decomposed as

$$p(\mathbf{y}, \mathbf{h}, \chi, \sigma) = p(\mathbf{y}|\mathbf{h}, \sigma)p(\sigma)p(\mathbf{h}|\chi)p(\chi) \quad (5)$$

where

$$p(\mathbf{y}|\mathbf{h}, \sigma) = \mathcal{CN}(\mathbf{y}|\mathbf{P}\mathbf{h}, \sigma^2\mathbf{I})$$

based on (3). Furthermore, let the conjugate prior

$$p(\sigma) = \text{Ga}(\sigma|c, d)$$

and

$$p(\mathbf{h}|\chi) = \prod_{l=1}^{LN_t} p(h_l|\chi_l)$$

with

$$p(h_l|\chi_l) = \left(\frac{\rho}{\pi\chi_l}\right)^\rho \exp\left(-\rho\frac{|h_l|^2}{\chi_l}\right),$$

respectively. Also, it is assumed that $p(\chi) = \prod_{l=1}^{LN_t} p(\chi_l)$ where $p(\chi_l) = \text{Ga}(\chi_l|\varepsilon, \eta_l)$. Then we can use the product of the generalized inverse Gaussian (GIG) to compute the prior of \mathbf{h} as

$$p(\mathbf{h}; \chi) = \int_0^\infty p(\mathbf{h}|\chi)p(\chi)d\chi = \prod_{l=1}^{LN_t} p(h_l; \varepsilon, \eta_l)$$

with

$$p(h_l; \varepsilon, \eta_l) = \frac{2}{\pi\Gamma(\varepsilon)} \eta_l^{\frac{\varepsilon+1}{2}} |h_l|^{\varepsilon-1} K_{\varepsilon-1}(2\sqrt{\eta_l}|h_l|).$$

3.2 VMP Based Sparse Channel Estimation

In this section we present a VMP algorithm for estimating \mathbf{h} in (3).

In fact, the PDF $p(\mathbf{h}|\mathbf{y})$ in (4) can be written as

$$p(\mathbf{h}|\mathbf{y}) = \int p(\Omega|\mathbf{y})d\chi d\eta d\sigma,$$

by regarding χ , η , and σ as nuisance parameters and “integrating them out”, where Ω is a set including all unknown variables, i.e., $\Omega = \{\mathbf{h}, \chi, \sigma\}$. However, it is always difficult to compute the integration to obtain the joint probability density function $p(\Omega, \mathbf{y})$. Fortunately, the variational theory is an effective way to relax this. The core of the variational theory is to find auxiliary probability $q(\Omega)$ to substitute the true probability $p(\Omega, \mathbf{y})$ by VMP method [12]. The VMP algorithm is an iterative scheme that attempts to compute the auxiliary PDF $q(\Omega)$ by minimizing the Kullback-Leibler

(KL) divergence $KL(q(\Omega)||p(\Omega, \mathbf{y}))$. The KL divergence is used to evaluate the extent of approximation between $q(\Omega)$ and $p(\Omega, \mathbf{y})$. Furthermore, we assume that

$$q(\Omega) = q(\mathbf{h})q(\chi)q(\eta)q(\sigma).$$

To clearly encode the factorization of $p(\Omega, \mathbf{y})$, the factor graph is shown in Fig. 1 [14]. In Fig. 1, the factor nodes g_i can be expressed as $g_y = p(\mathbf{y}|\mathbf{h}, \sigma)$, $g_h = p(\mathbf{h}|\chi)$, $g_\chi = p(\chi)$, and $g_\sigma = p(\sigma)$, respectively.

Based on Fig. 1, $q(\Omega)$ can be computed by VMP algorithm as follows

$$q(\Omega_i) \propto \prod_{g_n \in S_{\Omega_i}} m(g_n \rightarrow \Omega_i). \quad (6)$$

In (6), S_{Ω_i} is the set including all factor nodes whose neighbor is the factor node Ω_i and $m(g_n \rightarrow \Omega_i)$ denotes the message from factor node g_n to variable node Ω_i with

$$m(g_n \rightarrow \Omega_i) = \exp(E(\ln g_n, \prod_j q(\Omega_j)), \Omega_j \in S_{g_n} \setminus \{\Omega_i\}) \quad (7)$$

where S_{g_n} is the set composed by the variable nodes that neighbour the factor node g_n .

Using (6), the unknowns can be calculated iteratively by viewing the other variables as constant, which are shown as follows.

1) Update of $q(\mathbf{h})$: Combining (6) and Fig. 1, the computation of update of $q(\mathbf{h})$ can be written as

$$q(\mathbf{h}) \propto m(g_y \rightarrow \mathbf{h})m(g_h \rightarrow \mathbf{h})$$

where

$$m_{g_y \rightarrow h} = \exp(E(\ln p(\mathbf{y}|\mathbf{h}, \sigma), q(\sigma))) \propto \exp(-\rho\langle\sigma\rangle_{q(\sigma)} \|\mathbf{y} - \mathbf{A}\mathbf{h}\|_2^2)$$

and

$$m_{g_h \rightarrow h} = \exp(\langle\ln p(\mathbf{h}|\chi)\rangle_{q(\chi)}) \propto \exp(-\rho\mathbf{h}^H\mathbf{V}(\chi)\mathbf{h}),$$

respectively. Multiply these messages, and the auxiliary PDF $q(\mathbf{h})$ can be written as

$$q(\mathbf{h}) = \mathcal{CN}(\mathbf{h}|\hat{\mathbf{h}}, \hat{\Sigma}_{\mathbf{h}}) \quad (8)$$

$$\text{where } \hat{\Sigma}_{\mathbf{h}} = \left(\langle\sigma\rangle_{q(\sigma)}\mathbf{A}^H\mathbf{A} + \mathbf{V}(\chi)\right)^{-1} \quad (9)$$

$$\hat{\mathbf{h}} = \langle\mathbf{h}\rangle_{q(\mathbf{h})} = \langle\sigma\rangle_{q(\sigma)}\hat{\Sigma}_{\mathbf{h}}\mathbf{A}^H\mathbf{y}. \quad (10)$$

In equation (9), $\mathbf{V}(\chi)$ is defined as

$$\mathbf{V}(\chi) = \text{diag}\{E(\chi_1^{-1}, q(\chi)) \cdots E(\chi_N^{-1}, q(\chi))\}.$$

2) Update of $q(\chi)$: According to (7) and Fig. 1, the update of $q(\chi)$ can be expressed as

$$m_{g_h \rightarrow \chi} = \exp(\langle\ln p(\mathbf{h}|\chi)\rangle_{q(\mathbf{h})}) \propto \prod_{l=1}^{LN_t} \chi_l^{-\rho} \exp(-\rho\chi_l^{-1}\langle|h_l|^2\rangle_{q(\mathbf{h})})$$

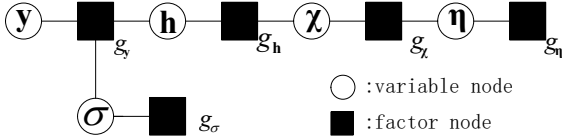


Fig. 1. A factor graph of the signal model of the joint PDF of (5).

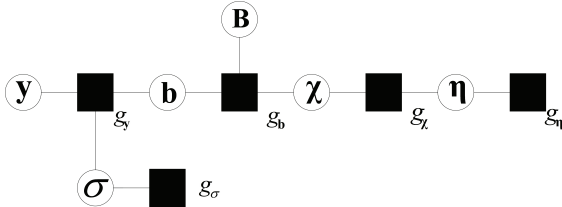


Fig. 2. A factor graph of the signal model of the joint PDF of (17).

and

$$m_{g_{\chi} \rightarrow \chi} = \exp(\langle p(\chi|\eta) \rangle_{q(\eta)}) \propto \prod_{l=1}^{LN_t} \chi_l^{\varepsilon-1} \exp(-\chi_l \langle \eta_l \rangle_{q(\eta)}).$$

Then the expression of $q(\chi)$ can be obtained as

$$q(\chi) = \prod_{l=1}^{LN_t} \chi_l^{\varepsilon-\rho-1} \exp(-\chi_l^{-1} \rho \langle |h_l|^2 \rangle_{q(\mathbf{h})} - \chi_l \eta_l) \quad (11)$$

where the auxiliary probability $q(\chi)$ can be viewed as the product of GIG PDF with order $\varepsilon - \rho$. Therefore, the $\langle \chi_l^j \rangle_{q(\chi)}$ is given by [15]

$$\left(\frac{\langle |h_l|^2 \rangle_{q(\mathbf{h})}}{\langle \eta_l \rangle_{q(\eta)}} \right)^{\frac{j}{2}} \frac{K_{\varepsilon-\rho+j}(2\sqrt{\langle \eta_l \rangle_{q(\eta)} \langle |h_l|^2 \rangle_{q(\mathbf{h})}})}{K_{\varepsilon-\rho}(2\sqrt{\langle \eta_l \rangle_{q(\eta)} \langle |h_l|^2 \rangle_{q(\mathbf{h})}})}. \quad (12)$$

3) Update of $q(\sigma)$: From (7) and Fig. 1, it can be easily obtained that

$$q(\sigma) = Ga(\sigma|\rho M + c, \rho \langle \|\mathbf{y} - \mathbf{P}\mathbf{b}\|_2^2 \rangle_{q(\mathbf{h})} + d).$$

Then we have

$$\langle \sigma \rangle_{q(\sigma)} = \frac{\rho M + c}{\langle \|\mathbf{y} - \mathbf{P}\mathbf{b}\|_2^2 \rangle_{q(\mathbf{h})} + d}. \quad (13)$$

4. Bayesian Block-Sparse Channel Estimation

4.1 Block-Structured Channel Model

As the number of transmit antennas increases, channel estimation performance may be degraded significantly, due to the increment of the length of the aggregate channel vector and the limited pilot signal size.

Specially, due to the close antenna spacing at the base station, the times of arrival from different transmit antennas are similar to each other. With the limited sampling resolution at the receiver, the nonzero tap locations of each CSI vector are considered identical [16], that is

$$\text{supp}(\mathbf{h}_i) = \text{supp}(\mathbf{h}_j), \quad i \neq j \quad (14)$$

where $\text{supp}(\mathbf{h}_i)$ is defined as the length- L support vector for the i -th vector \mathbf{h}_i , and

$$\text{supp}(\mathbf{h}_i)(k) = \begin{cases} 1 & \mathbf{h}_i(k) \neq 0 \\ 0 & \mathbf{h}_i(k) = 0 \end{cases}, \quad 1 \leq k \leq L \quad (15)$$

where $\text{supp}(\mathbf{h}_i)(k)$ is the k -th element in \mathbf{h}_i . Based on this analyses, let G be the number of the nonzero values in \mathbf{h}_l . Without loss of generality, it is assumed that the numbers of nonzero value of \mathbf{h} satisfies $(G \times N_t) \ll (L \times N_t)$ and $(G \times N_t) < M$ in (3). Thus, the aggregate channel vector \mathbf{h} can be rearranged into $\mathbf{b} = [\mathbf{b}_0^T, \dots, \mathbf{b}_{L-1}^T]^T$ with $\mathbf{b}_l = [\mathbf{h}_1(l), \dots, \mathbf{h}_{N_t}(l)]^T$. Furthermore, the equation (3) can be rewritten as

$$\mathbf{y} = \mathbf{P}\mathbf{b} + \mathbf{w} \quad (16)$$

where

$$\mathbf{P} = [\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_{L-1}]$$

and $\mathbf{P}_l = [\mathbf{a}_l, \mathbf{a}_{l+L}, \dots, \mathbf{a}_{(N_t-1)L+l}]$. Also, \mathbf{a}_l is the $(l+1)$ -th column vector of matrix \mathbf{A} . In this situation, the new vector \mathbf{b} exhibits block sparsity. It means that among the L blocks of vector \mathbf{b} , only a small number of \mathbf{b}_l in the multiple channel vectors $\{\mathbf{b}_l\}_{l=0}^{L-1}$ are nonzero vectors.

Based on the assumption of the wide sense stationary uncorrelated scattering (WSSUS) model, the blocks can be viewed as independent of each other, which implies that we only need to consider the internal correlation of blocks.

We aim at obtaining the estimation of \mathbf{b} in (16). Since \mathbf{b} has the feature of block sparsity, it is meaningful to introduce the block-sparse structure into the sparse estimation problem (16) to improve the system performance. In this section, we will propose a variational message passing (VMP) based Bayesian-block-sparse channel estimation algorithm to estimate \mathbf{b} with the given observation \mathbf{y} .

4.2 Bayesian Prior Model for Block-Sparse Estimation

Since \mathbf{b} has a block-sparse structure, the covariance matrix \mathbf{B}_l is introduced to describe the character of the correlation between the elements of \mathbf{b}_l . Similar to [17], it is assumed that $\mathbf{B} = \mathbf{B}_l$ for all l . In this situation,

$$p(\mathbf{b}_l|\chi_l, \mathbf{B}) = \left(\frac{\rho}{\pi\chi_l}\right)^{N_t\rho} |\mathbf{B}|^{-\rho} \exp(-\rho \mathbf{b}_l^H (\chi_l \mathbf{B})^{-1} \mathbf{b}_l).$$

Then the PDF of \mathbf{b} can be written as

$$p(\mathbf{b}|\chi, \mathbf{B}) = \prod_{l=1}^L \left(\frac{\rho}{\pi\chi_l}\right)^{N_t\rho} |\mathbf{B}|^{-\rho} \exp(-\rho \mathbf{b}_l^H (\chi_l \mathbf{B})^{-1} \mathbf{b}_l). \quad (17)$$

Also, $p(\chi)$ is written as

$$p(\chi) = \prod_{l=1}^L p(\chi_l)$$

with $p(\chi_l) = Ga(\chi_l|\varepsilon, \eta_l)$. Thus the computation of the prior of \mathbf{b} is

$$p(\mathbf{b}; \mathbf{B}) = \prod_{l=1}^L \int p(\mathbf{b}_l|\chi_l, \mathbf{B})p(\chi_l)d\chi_l.$$

The factor graph of the block-sparse model is shown in Fig. 2. The factor nodes g_i can be expressed as $g_y = p(\mathbf{y}|\mathbf{b}, \sigma)$, $g_b = p(\mathbf{b}|\chi, \mathbf{B})$, $g_\chi = p(\chi)$, and $g_\sigma = p(\sigma)$, respectively.

4.3 VMP-Block-Sparse Channel Estimation

Using VMP algorithm, the calculation of each variable is as follows.

1) Update of $q(\mathbf{b})$: Since

$$q(\mathbf{b}) \propto m(g_y \rightarrow \mathbf{b})m(g_b \rightarrow \mathbf{b})$$

where

$$m_{g_y \rightarrow b} = \exp(E(\ln p(\mathbf{y}|\mathbf{b}, \sigma), q(\sigma))) \\ \propto \exp(-\rho \langle \sigma \rangle_{q(\sigma)} \|\mathbf{y} - \mathbf{P}\mathbf{b}\|_2^2)$$

and

$$m_{g_b \rightarrow b} = \exp(\langle \ln p(\mathbf{b}|\chi, \mathbf{B}) \rangle_{q(\chi)}) \propto \exp(-\rho \mathbf{b}^H \mathbf{V}(\chi) \mathbf{b}),$$

respectively. Multiply these messages, and the auxiliary PDF $q(\mathbf{b})$ can be written as :

$$q(\mathbf{b}) = CN(\mathbf{b}|\hat{\mathbf{b}}, \hat{\Sigma}_{\mathbf{b}}) \quad (18)$$

where $\hat{\Sigma}_{\mathbf{b}} = \left(\langle \sigma \rangle_{q(\sigma)} \mathbf{P}^H \mathbf{P} + \mathbf{V}(\chi) \right)^{-1}$ (19)

$$\hat{\mathbf{b}} = \langle \mathbf{b} \rangle_{q(\mathbf{b})} = \langle \sigma \rangle_{q(\sigma)} \hat{\Sigma}_{\mathbf{b}} \mathbf{P}^H \mathbf{y}. \quad (20)$$

In equation (19), $\mathbf{V}(\chi)$ is defined as

$$\mathbf{V}(\chi) = \text{diag}\{B^{-1}E(\chi_1^{-1}, q(\chi)) \cdots B^{-1}E(\chi_L^{-1}, q(\chi))\}.$$

2) Update of $q(\chi)$: Since

$$m_{g_b \rightarrow \chi} = \exp(\langle \ln p(\mathbf{b}|\chi, \mathbf{B}) \rangle_{q(\mathbf{b})}) \\ \propto \prod_{l=1}^L \chi_l^{-\rho N_t} |\mathbf{B}|^{-\rho} \exp(-\rho \langle \mathbf{b}_l^H \chi_l^{-1} \mathbf{B}^{-1} \mathbf{b}_l \rangle_{q(\mathbf{b})})$$

where

$$m_{g_\chi \rightarrow \chi} = \exp(\langle p(\chi|\eta) \rangle_{q(\eta)}) \propto \prod_{l=1}^L \chi_l^{\varepsilon-1} \exp(-\chi_l \langle \eta_l \rangle_{q(\eta)}).$$

Then the expression of $q(\chi)$ can be written as

$$q(\chi) = \prod_{l=1}^L \chi_l^{\varepsilon-\rho N_t-1} \exp(-\rho \langle \mathbf{b}_l^H \chi_l^{-1} \mathbf{B}^{-1} \mathbf{b}_l \rangle_{q(\mathbf{b})} \\ - \chi_l \langle \eta_l \rangle_{q(\eta)}) \quad (21)$$

where $q(\chi)$ can be viewed as the product of GIG PDF with order $\varepsilon - \rho N_t$. Therefore, the $\langle \chi_l^i \rangle_{q(\chi)}$ is given by [15]

$$\langle \chi_l^i \rangle_{q(\chi)} = \left(\frac{e}{f} \right)^{\frac{i}{2}} \frac{K_{\varepsilon-\rho N_t+i}(2\sqrt{ef})}{K_{\varepsilon-\rho N_t}(2\sqrt{ef})} \quad (22)$$

where $e = \langle \mathbf{b}_l^H \mathbf{B}^{-1} \mathbf{b}_l \rangle_{q(\mathbf{b})}$ and $f = \langle \eta_l \rangle_{q(\eta)}$. For modified Bessel function, we have

$$K_m(x) \approx \frac{1}{2} \Gamma(m) \left(\frac{x}{2} \right)^{-m} \text{ for } m > 0, x \rightarrow 0 \quad (23)$$

and

$$\frac{dK_m(x)}{dx} + \frac{m}{x} K_m(x) = -K_{m-1}(x). \quad (24)$$

Therefore, $\langle \chi_l^{-1} \rangle_{q(\chi)}$ with $i = -1$ can be written as

$$\langle \chi_l^{-1} \rangle_{q(\chi)} = \frac{\rho N_t - \varepsilon}{e}$$

where e can be further deduced as

$$\langle \mathbf{b}_l^H \mathbf{B}^{-1} \mathbf{b}_l \rangle_{q(\mathbf{b})} = \text{Tr}[\mathbf{B}^{-1}(\hat{\Sigma}_{\mathbf{b}}^l + \hat{\mathbf{b}}_l(\hat{\mathbf{b}}_l)^H)].$$

3) Update of \mathbf{B} : In order to improve the estimation performance, the covariance matrix \mathbf{B} can be constrained, which is inspired by [17]. A possible form in [17] assumes that the covariance matrix is a Toeplitz matrix. The matrix \mathbf{B} is assumed to have a non-stochastic variable in this paper because of its statistical characteristic. According to equation (17), the expression of updating \mathbf{B} can be derived as

$$\log p(\mathbf{b}|\chi, \mathbf{B}) \propto -\rho \log(|\mathbf{X}|^{N_t} |\mathbf{B}|^L) - \rho \mathbf{b}^H (\mathbf{X}^{-1} \otimes \mathbf{B}^{-1}) \mathbf{b} \quad (25)$$

where $\mathbf{X} = \text{diag}(\chi_1, \dots, \chi_L)$. The derivative of (25) with respect to \mathbf{B} , is given by

$$\frac{\partial(\log p(\mathbf{b}|\chi, \mathbf{B}))}{\partial \mathbf{B}} = -\rho L \mathbf{B}^{-1} \\ + \rho \sum_{l=1}^L \frac{1}{\langle \chi_l \rangle_{q(\chi)}} \mathbf{B}^{-1} (\hat{\Sigma}_{\mathbf{b}}^l + \hat{\mathbf{b}}_l(\hat{\mathbf{b}}_l)^H) \mathbf{B}^{-1}.$$

Thus, we obtain the rule of updating of \mathbf{B} as

$$\mathbf{B} = \frac{1}{L} \sum_{l=1}^L \frac{\hat{\Sigma}_{\mathbf{b}}^l + \hat{\mathbf{b}}_l(\hat{\mathbf{b}}_l)^H}{\langle \chi_l \rangle_{q(\chi)}} \quad (26)$$

where $\hat{\mathbf{b}}_l$ is the l -th block in $\hat{\mathbf{b}}$ (with the size of $N_t \times 1$), and $\hat{\Sigma}_{\mathbf{b}}^l$ is the l -th diagonal block in $\hat{\Sigma}_{\mathbf{b}}$ (with the size of $N_t \times N_t$). Furthermore, \mathbf{B} can be substituted by a Toeplitz matrix with

$$\mathbf{B} = \text{Toeplitz}([1, r, \dots, r^{N_t-1}]) \\ = \begin{bmatrix} 1 & r & r^2 & \dots & r^{N_t-1} \\ r & 1 & r & \dots & r^{N_t-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^{N_t-1} & r^{N_t-2} & r^{N_t-3} & \dots & 1 \end{bmatrix} \quad (27)$$

where

$$r \triangleq \text{sign}\left(\frac{a_1}{a_0}\right) \min\left\{\left|\frac{a_1}{a_0}\right|, 0.99\right\}.$$

Also, a_0 and a_1 are the average value of the elements along the main diagonal and the sub-diagonal of the matrix in (26), respectively.

4) Update of $q(\sigma)$: It can be easily obtained that

$$q(\sigma) = Ga(\sigma | \rho M + c, \rho < \|\mathbf{y} - \mathbf{P}\mathbf{b}\|_2^2 >_{q(\mathbf{h})} + d).$$

Then we have

$$\langle \sigma \rangle_{q(\sigma)} = \frac{\rho M + c}{\langle \|\mathbf{y} - \mathbf{P}\mathbf{b}\|_2^2 \rangle_{q(\mathbf{h})} + d}. \tag{28}$$

5. Simulation

In this section, the performance of the proposed VMP algorithm is evaluated. Meanwhile, we compare the normalized mean square error (NMSE) of the proposed algorithm and the traditional ones. The measurement matrix \mathbf{P} in section II is given with dimension $M = 560$, $N_t = 8$, $L = 128$, and the number of the nonzero values in \mathbf{h}_l is $G = 20$ in Fig. 3.

Figure 3 shows the performance of NMSE among OMP algorithm, SP algorithm and the proposed VMP algorithm versus the signal-to-noise ratio (SNR). It can be observed that the proposed VMP channel estimate algorithm outperforms the other traditional estimation methods as SNR increases. As shown in Fig. 3, the proposed VMP algorithm shows better performance compared with SP algorithm when SNR is relative small. Besides, with the increase of SNR, the gap between the proposed algorithm and the SP algorithm becomes smaller, for the sake that the affection of the noise is almost negligible when SNR becomes large enough. In Fig. 3, it can be seen that the SP algorithm performs not only better than OMP algorithm but also than VMP algorithm when $\varepsilon = 1$. However, with appropriate choice of the value of ε , the proposed VMP algorithm achieves significant performance gain over the others.

Additionally, the performance with different ε is also shown in Fig. 4. It can be seen that with the decrease of ε the performance of the VMP algorithm is improved. That is because the parameter ε controls the sparsity properties, and the case $\varepsilon = 0$ encourages a sparser solution than $\varepsilon > 0$. Specifically, by considering the sparse block structure of the channel vector, the complexity of proposed method can be reduced significantly.

In Fig. 5, we set the dimension of P with $M = 560$, $N_t = 60$, $L = 30$. And the number of the nonzero values in \mathbf{h}_l is $G = 4$. Comparing Fig. 4 and Fig. 5, it can be easily found that with the increasing of the number of the transmission antennas, the performance of all of the algorithms decreases. However, the proposed algorithm shows better robustness than the others. The performance of the proposed VMP algorithm achieves significant performance than the others even when $\varepsilon = 1$.

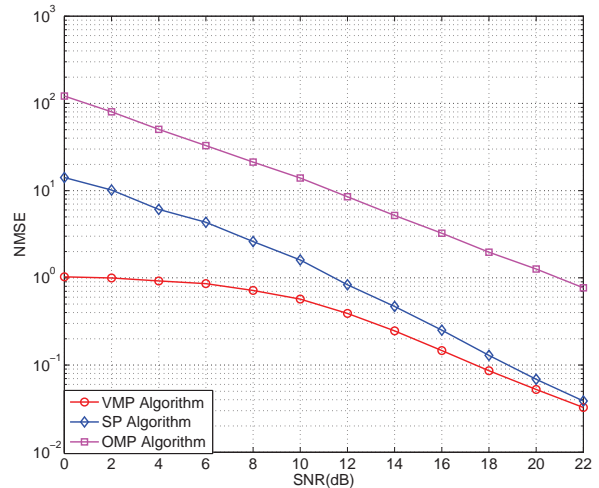


Fig. 3. Performance comparison of NMSE of the algorithms with 8 antennas.

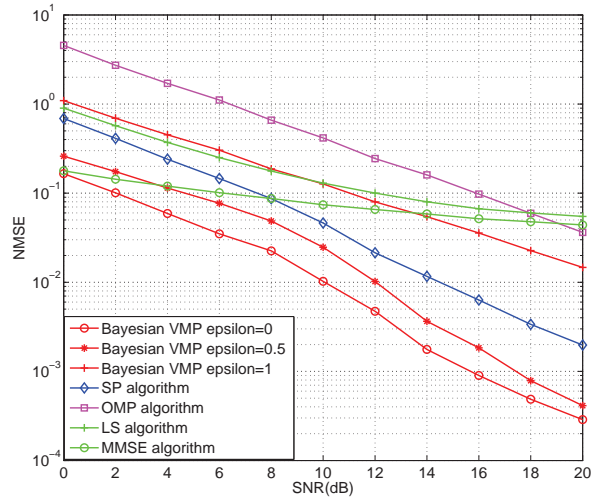


Fig. 4. Performance comparison of NMSE of the Block-sparse algorithms with 8 antennas.

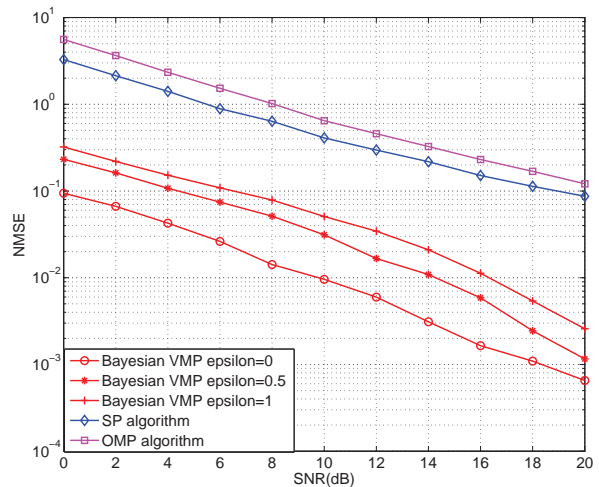


Fig. 5. Performance comparison of NMSE of the Block-sparse algorithms with 60 antennas.

6. Conclusion

This paper studies a Bayesian-VMP sparse channel estimation algorithm for large scale MISO-OFDM systems. By introducing an intermediate random vector and carefully choosing the PDF based on it, the sparsity-inducing prior of the unknown vector is modelled. After that, an optimization problem over CSI is constructed. The VMP algorithm is proved to be an effective way to solve the optimization problem. Simulations indicate the merits of the proposed algorithm over traditional ones. The proposed algorithm is proved to be an effective solution of estimating the sparse channel parameters in large MISO-OFDM systems.

Acknowledgments

Part of this work was published in IEEE 83rd Vehicular Technology Conference (IEEE VTC), 2016. This work was supported by the National Natural Science Foundation of China under Grant No. 11301413.

References

- [1] KHAN, M., JAMIL, M., HUSSAIN, M. Adaptation of hybrid FSO/RF communication system using puncturing technique. *Radioengineering*, 2016, vol. 6, no. 4, p. 644–651. DOI: 10.13164/re.2016.0644
- [2] KHAN, M., GILANI, S., JAMIL, M., et al. Maximizing throughput of hybrid FSO-RF communication system: An algorithm. *IEEE Access*, 2018, vol. 6, p. 30039–30048. DOI: 10.1109/ACCESS.2018.2840535
- [3] RUSEK, F., PERSSON, D., LAU, B., et al. Scaling up MIMO: Opportunities and challenges with very large arrays. *IEEE Signal Processing Magazine*, 2013, vol. 30, no. 1, p. 40–60. DOI: 10.1109/MSP.2011.2178495
- [4] JORGE, I., ISMAIL, G., SENNER, D., et al. Efficient noise variance estimation under pilot contamination for massive MIMO systems. *IEEE Transactions on Vehicular Technology*, 2018, vol. 67, no. 4, p. 2982–2996. DOI: 10.1109/TVT.2017.2766226
- [5] SHASHINDRA, S., ARUMA, B., MASOUD, A. Performance analysis of massive MIMO two-way relay networks with pilot contamination, imperfect CSI, and antenna correlation. *IEEE Transactions on Vehicular Technology*, 2018, vol. 67, no. 6, p. 4831–4842. DOI: 10.1109/TVT.2018.2812724
- [6] PAULRAJ, A., NABAR, R., and GORE, D. *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003. ISBN: 9780521826150
- [7] BAJWA, W., HAUPT, J., SAYEED, A., et al. Compressed channel sensing: A new approach to estimating sparse multipath channels. *Proceedings of the IEEE*, 2010, vol. 98, no. 6, p. 1058–1076. DOI: 10.1109/JPROC.2010.2042415
- [8] TAUBOCK, G., HLAWATSCH, F. A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Las Vegas (USA), 2008, p. 2885–2888. DOI: 10.1109/ICASSP.2008.4518252
- [9] COTTER, S., RAO, B. Sparse channel estimation via matching pursuit with application to equalization. *IEEE Transactions on Communications*, 2002, vol. 50, no. 3, p. 374–377. DOI: 10.1109/26.990897
- [10] BERGER, C., WANG, Z., HUANG, J., et al. Application of compressive sensing to sparse channel estimation. *IEEE Communications Magazine*, 2010, vol. 48, no. 11, p. 164–174. DOI: 10.1109/MCOM.2010.5621984
- [11] TIPPING, M. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 2001, vol. 1, p. 211–244. DOI: 10.1162/15324430152748236
- [12] PEDERSEN, N., MANCHON, C., SHUTIN, D., et al. Application of Bayesian hierarchical prior modeling to sparse channel estimation. *IEEE International Conference on Communications (ICC)*. Ottawa (Canada), 2012, p. 3487–3492. DOI: 10.1109/ICC.2012.6363847
- [13] BARBOTIN, Y., HORMATI, A., RANDAN, S., et al. Estimation of sparse MIMO channels with common support. *IEEE Transactions on Communications*, 2012, vol. 60, no. 12, p. 3705–3716. DOI: 10.1109/TCOMM.2012.091112.110439
- [14] KSCHISCHANG, F., FREY, B., LOELIGER, H. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 2001, vol. 47, no. 2, p. 498–519. DOI: 10.1109/18.910572
- [15] JORGENSEN, B. *Statistical Properties of the Generalized Inverse Gaussian Distribution (Lecture Notes in Statistics 9)*. Springer-Verlag New York Inc, 1982. ISBN: 9780387906652
- [16] HORMATI, A., VETTERLI, M. Compressive sampling of multiple sparse signals having common support using finite rate of innovation principles. *IEEE Signal Processing Letters*, 2011, vol. 18, no. 5, p. 331–334. DOI: 10.1109/LSP.2011.2131649
- [17] ZHANG, Z., RAO, B. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE Journal of Selected Topics in Signal Processing*, 2011, vol. 5, no. 5, p. 912–926. DOI: 10.1109/ICASSP.2011.5947212

About the Authors ...

Feng LI (corresponding author) was born in 1981. He received his B.Sc. and Ph.D. degrees in Information and Communication Engineering in 2003 and 2009, respectively, both from Xi'an Jiaotong University, China. He is now an Associate Professor in the same university. His research interests are wireless communications including OFDM, MIMO, cognitive radio.

Hailin LI was born in 1990. She received the B.Sc. in 2013 from Xi'an Jiaotong University, China. She is currently pursuing the M.Sc degree in the field of Information and Communication Engineering in Xi'an Jiaotong University, China. Her research interests are sparse signal processing in wireless communication systems.

Shuyuan LI was born in 1990. He received his B.Sc. in 2013 from Xi'an Jiaotong University, China. He is currently pursuing the M.Sc degree in the field of Wireless Communication in Xi'an Jiaotong University, China. His research interests are channel estimation for MIMO and OFDM systems.