Variational Sparse Bayesian Learning for Estimation of Gaussian Mixture Distributed Wireless Channels

Lingjin Kong, Xiaoying Zhang, Haitao Zhao and Jibo Wei

School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China; konglingjin19@nudt.edu.cn (L.K.); haitaozhao@nudt.edu.cn (H.Z.); wjbhw@nudt.edu.cn (J.W.)
* Correspondence: zhangxiaoying@nudt.edu.cn

Abstract: In this paper, variational sparse Bayesian learning is utilized to estimate the multipath parameters for wireless channels. Due to its flexibility to fit any probability density function (PDF), the Gaussian mixture model (GMM) is introduced to represent the complicated fading phenomena in various communication scenarios. First, the expectation-maximization (EM) algorithm is applied to the parameter initialization. Then, the variational update scheme is proposed and implemented for the channel parameters’ posterior PDF approximation. Finally, in order to prevent the derived channel model from overfitting, an effective pruning criterion is designed to eliminate the virtual multipath components. The numerical results show that the proposed method outperforms the variational Bayesian scheme with Gaussian prior in terms of root mean squared error (RMSE) and selection accuracy of model order.

Keywords: channel measurement; Gaussian mixture model; variational Bayesian; sparse channel; channel estimation

1. Introduction

Channel modeling plays a pivotal role in the design and evaluation of wireless communication networks. In the fifth-generation (5G) mobile communication systems, the diversity of the communication scenarios leads to the urgent need for advanced channel modeling techniques with higher accuracy and lower implementation complexity. In the existing channel measurement campaigns, the channel sounding technique based on pseudo-random (PN) sequences has widely been adopted due to its higher processing gain and delay resolution. As for the estimation of the multipath characteristics such as power delay profile, Doppler power spectrum, and frequency response from the obtained channel responses, super-resolution algorithms such as space-alternating generalized expectation-maximization (SAGE) [1] and subtractive deconvolution technique (CLEAN algorithm) [2] are widely applied. However, since the SAGE and CLEAN algorithms are based on the maximum likelihood criterion, the selection of model order, i.e., the number of multipath, is a nontrivial task. The estimated order of the channel model tends to be overestimated because of the contamination of noise. The unnecessary “virtual paths” will not only affect the accuracy of the channel coefficients estimation but also increase the computational cost of the algorithm. To solve this problem, different techniques such as negative log-evidence (NLE) [3] and Bayesian information criterion (BIC) [4] have been proposed to estimate the model order. However, the estimation performance of these model estimation schemes is far from ideal, especially when considering the low signal-to-noise ratio scenarios [5]. Since considerable computation is needed to find the optimal model order value, the existing techniques also suffer from prohibitive complexity.

Due to the fact that the multipath components (MPC) are sparse in channel impulse responses (CIR) [6], the joint estimation problem of channel parameters and channel model order can be addressed by sparse Bayesian learning (SBL). SBL is a typical machine learning approach that leverages sparse priors to achieve model order selection and parameter estimation simultaneously. In this paper, we adopt a variational sparse Bayesian learning framework to estimate the multipath parameters under the GMM framework. The key contributions of this work are summarized as follows:

- **Gaussian Mixture Model (GMM):** The GMM is employed to model the complicated fading phenomena observed in practical wireless communications, allowing for a more flexible and accurate representation of the channel characteristics.
- **Variational Update Scheme:** A variational update scheme is proposed to approximate the posterior PDF of the channel parameters, facilitating the optimization process.
- **Effective Pruning Criterion:** An effective pruning criterion is designed to eliminate unnecessary virtual multipath components, reducing overfitting and improving the accuracy of the estimated channel model.

The proposed method is validated through numerical simulations, demonstrating its superior performance in terms of root mean squared error (RMSE) and selection accuracy of model order compared to the variational Bayesian scheme with Gaussian prior.
algorithm that was first proposed by Tipping in 2001 [7] and has been developed into an important branch of compressed sensing reconstruction algorithms. The basic idea of SBL is to introduce the sparsity parameter to establish a hierarchical Bayesian model. By defining the priors on the sparse vector, SBL can automatically determine the position and amount of the non-zero elements in the reconstructed signal without pre-specifying the sparsity number. SBL is a promising method for the joint estimation of channel model order and the channel response coefficients. However, even after the hierarchical Bayesian model is established, the posterior PDF cannot be directly derived due to an intractable integral of the joint probability density. Thus, the variational inference is introduced to obtain an approximation of the parameter posterior in the SBL framework. This method, also known as variational sparse Bayesian learning (VSBL) [8], has been widely applied in online spectrum estimation [9], orthogonal time-frequency space (OTFS) detector [10], multi-user joint decoding [11], joint signal detection [12], and high-resolution radar imaging [13], etc. For channel parameter estimation, the idea of complete hidden data is introduced to speed up the convergence in [5]. The simulation result shows that the VSBL scheme outperforms the BIC criterion in both root mean squared error (RMSE) and model order selection. In [14], the delay of MPCs is modeled as the Poisson distribution, which incorporates the Fast VSBL based channel estimation as a priori information and significantly improves the bit error ratio (BER) performance.

Several channel measurement campaigns conducted in high-speed railways [15], airport grounds [16,17], and inland rivers [18] have shown that Weibull and Nakagami-m distributions are more suitable for modeling channel characteristics compared with Rayleigh and Rice distribution. This phenomenon originates from the uneven distribution of scatterers in the mobile communication environment. However, in most VSBL models, the channel taps were assumed to be Gaussian distributed [19,20], which is not appropriate for parameter learning. In this paper, the Gaussian mixture model (GMM) is adopted to describe the statistical characteristics of the MPCs. A GMM is linearly composed of $K$ Gaussian PDFs with corresponding means and covariances. By using sufficient Gaussian components, the GMM can be flexible to approximate any given PDF [21], which is suitable for establishing various fading channels models.

The contributions of this paper are summarized as follows:

- The Gaussian mixture model as a powerful method for sparse parameter learning for wireless channels is introduced to the estimation problem of wireless channel parameters under the VSBL framework. The flexibility of GMM is capable of describing the statistical characteristics of both theoretical general channels and complex real-world channels.

- A new variational Bayesian inference scheme for the Gaussian mixture model (VB-GMM) is developed based on multiple channel observations. The corresponding graphical model is given, and the closed-form updates of the model variables are derived. By setting a pruning criterion on the sparsity priors, the joint estimation of channel parameters and model order is achieved with low complexity.

- The simulation results demonstrate that the performance of VB-GMM is superior to the existing algorithms in terms of the estimation error, the convergence rate, and the model order selection accuracy in most non-Gaussian channels.

Throughout the paper, we use the following notations. $(\cdot)^H$ and $(\cdot)^{-1}$ denote the Hermitian transpose and the inverse of matrices, respectively. The expression $CN(\mu, \Sigma)$ represents the multivariate complex Gaussian distribution with mean vector $\mu$ and covariance matrix $\Sigma$. $Ga(a|a,b)$ denotes the Gamma distribution $p(a) = \frac{a^b}{\Gamma(a)} \cdot a^{-1} e^{-ba}, a > 0$ in which $a, b$ and $\Gamma(\cdot)$ represent shape parameter, scale parameter and Gamma function, respectively. We use calligraphic uppercase letters to represent sets, e.g., $\mathcal{I}$ denotes the index set of MPCs and $\overline{\mathcal{I}(l)}$ is the complement set of $l \in \mathcal{I}$. The remainder of the paper is organized as follows. In Section 2, the signal model is introduced, and the graphical model of VB-GMM is presented. In Section 3, the derivations of the VB-GMM estimators are given.
followed by the initialization algorithm and the pruning criterion. Finally, in Section 4, the numerical results are presented to demonstrate the effectiveness of the proposed scheme.

2. System Model

2.1. Signal Model

Consider a single-input-single-output (SISO) wireless channel measurement system, the sounding signal is denoted as \( s(t) = \sum_{i=0}^{N_c-1} p_i b(t - iT_c) \), in which the pseudo-random sequence \( \{ p_i \} \), \( i = 0, \ldots, N_c - 1 \) has \( N_c \) chips and the shaped pulse \( b(t) \) has the duration \( T_c \). Assume that \( N_S \) copies of \( s(t) \) are transmitted periodically to obtain the time-varying channel responses. After passing through the multipath channel, the signal at the receiver is made of \( L \) duplications of \( s(t) \) with additive Gaussian white noise \( \eta(t) \), and the received signal corresponding to the \( n \)-th \( s(t) \) can be expressed as

\[
y_n(t) = \sum_{l=1}^{L} w_{ln} s(t, \theta_{ln}) + \eta(t), n = 1, \ldots, N_S
\]

where \( w_{ln} \) is the complex coefficient of the \( l \)-th multipath and \( \theta_{ln} \) denotes the dispersion parameter vector including multipath delay, DoA, DoD, etc. We define \( y_n = [y_n(0), y_n(T_s), \ldots, y_n((D - 1)T_s)]^T \) as the discrete samples of the \( n \)-th received signal in which \( T_s \) is the sampling interval at the receiver and \( D \) represents the number of discrete samples, i.e., \( DT_s = N_cT_c \). The discrete vector \( y_n \) is considered as an observation of the measured channel. In our proposed scheme, an Bayesian probabilistic model is established to infer the posterior probabilities of the channel parameters \( \{ w_{ln}, \theta_{ln}, L \} \) based on multiple observations.

Since the proposed algorithm is developed based on time-varying channels, it is crucial to determine the wide-sense stationary (WSS) region in which the statistical characteristics of channel parameters are assumed invariant. For vehicle-to-vehicle channel, the spatial distance of 20~40 times the wavelength is considered to be WSS [22]. For convenience, we denote the multiple observation matrix as \( y = [y_1, y_2, \ldots, y_N]_{D \times N} \), where \( N \) represents the number of channel responses obtained in a WSS region. If the spatial distance of the WSS is chosen properly according to the criterion, the dispersion parameter vector is assumed to be constant for the observations, i.e., \( \theta_{11} = \ldots = \theta_{1N} = \theta_1, n = 1, \ldots, N \). Thus Equation (1) can be rewritten in a discrete form as:

\[
y_n = \sum_{l=1}^{L} w_{ln} s(\theta_1) + \eta
\]

in which \( \eta \sim \mathcal{CN}(0, \Sigma) \) is the Gaussian white noise vector.

In the SAGE algorithm, the complete parameter set \( \Omega \) is divided into several subsets \( \Omega_{\text{sub}} \) so that the joint estimation for \( \Omega \) is decomposed into optimizations for \( \Omega_{\text{sub}} \) in a sequential manner. Let \( x_{\text{sub}} \) denote the admissible hidden data with respect to \( \Omega_{\text{sub}} \), when the equation \( p(y|x_{\text{sub}}, \Omega) = p(y|x_{\text{sub}}, \Omega_{\text{sub}})p(x_{\text{sub}} | \Omega) \) is satisfied, the likelihood corresponding to the estimation monotonically increases and may converge to the local maximum [23]. In this paper, the concept of admissible hidden data \( x_l \) is also adopted. Consider an MPC \( l \) we have

\[
x_l = w_{ls}(\theta_1) + \eta
\]

in which \( \eta \) represents the noise that the \( l \)-th MPC contribute to the total noise \( \eta \) with

\[
E\{ \eta \eta^H \} = \Sigma_l = \beta \Sigma, 0 \leq \beta \leq 1, \quad \text{and the likelihood } \ p(x_l | \theta_l, w_l) = \mathcal{CN}(x_l; w_{ls}(\theta_l), \Sigma_l).
\]

Then the covariance of the noise that is not related to the \( l \)-th component is \( \Delta_l = (1 - \beta) \Sigma \), which can also affect the estimation of \( x_l \) when observations are obtained and fixed. The noise covariance has a significant impact on the estimation results of channel parameters. Since the length of the channel sounding frame is generally much longer than the channel delay spread, most of the channel response samples are considered to be measurement noise. In this paper, the noise covariance is computed by the tail of CIRs.
2.2. The GMM Model of Channel Coefficients

Denote the $N$ channel coefficients of the $l$-th multipath as $w_l = [w_{l1}, w_{l2}, \ldots, w_{LN}]_{1 \times N}$. Assume that $w_l$ obeys the Gaussian mixture distribution with $K$ components:

$$p(w_l | \pi_l, \mu_l, \Lambda_l) = \sum_{k=1}^{K} \pi_k \mathcal{N}(w_l | \mu_k, \Lambda_k)$$ (4)

where $\pi_l = [\pi_{l1}, \pi_{l2}, \ldots, \pi_{lK}]$, $\sum_{k=1}^{K} \pi_k = 1$ is defined as mixing coefficients. $\mu_k$ and $\Lambda_k$ represent the complex mean and covariance matrix of the $k$-th Gaussian distribution, respectively. For $w_l$, each sample can only be derived from one component of the Gaussian mixture distribution. Therefore, the discrete hidden variable $c_l$ is introduced. $c_l$ is a 0–1 matrix in which $c_{l0k} = 1$ means the channel coefficient $w_{ln}$ corresponds to the $k$-th Gaussian component. Given the mixing coefficients, the probability of hidden variables takes a product form:

$$p(c_l | \pi_l) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_{lnk} c_{lnk}$$ (5)

For the variable $\pi_l$, a Dirichlet distribution prior is applied:

$$p(\pi_l | \gamma_l) = \text{Dir}(\pi_l | \gamma_l) = C(\gamma_l) \prod_{k=1}^{K} \pi_{lk}^{\gamma_{lk} - 1}$$ (6)

where $C(\gamma_l) = \frac{\Gamma(\gamma_{l0})}{\prod_{k=1}^{K} \Gamma(\gamma_{lk})}$, $\gamma_{l0} = \sum_{k=1}^{K} \gamma_{lk}$, and $\gamma_l = [\gamma_{l1}, \ldots, \gamma_{lK}]$ is assumed non-informative, i.e., $\gamma_{l1} = \gamma_{l2} = \cdots = \gamma_{lK}$. The Dirichlet distribution is introduced here as the conjugate prior of the polynomial likelihood function $p(c_l | \pi_l)$, which means the posterior of $c_l$ has the same form of Dirichlet distribution. In variational inference, the application of conjugate priors makes it convenient to merely update the parameters of the density functions with their form unchanged.

The graphical model for the variational Bayesian-based GMM is presented in Figure 1. Denote the parameter set as $\Omega = \{\theta, w, \mu, \alpha, \pi\}$ the joint probability density can be rewritten according to the structure of the graph:

$$p(y, \Omega) = p(y | w, \theta, \eta) p(w | \alpha, \mu) p(c | \pi) p(\alpha) p(\mu)$$ (7)

in which

$$p(y | w, \theta, \eta) = \prod_{n=1}^{N} \mathcal{N}(y_n; \sum_{l=1}^{L} w_{ln} s(\theta_l), \eta)$$ (8)

$$p(w | \alpha, \mu) = \prod_{l=1}^{L} p(w_l | \alpha_l, c_l, \mu_l)$$ (9)

The sparsity prior of $l$-th path $p(w_l | \alpha_l, c_l, \mu_l)$ is a key to restrain the gain of MPCs and implement model order selection. The sparsity prior is usually chosen as Laplace [24] or Gaussian distributed [7]. The Laplace prior leads to a $l_1$-type sparsity and Gaussian prior a $l_2$-type sparsity [5]. In the proposed scheme each component of the GMM is specified with an independent Gaussian sparsity prior, i.e., $p(w_{ln} | \alpha_k; c_{lnk} = 1) = \mathcal{N}(\mu_k, \alpha_k^{-1})$. Thus consider all the observations for the $l$-th path, we define

$$p(w_l | \alpha_l, c_l, \mu_l) = \prod_{n=1}^{N} \prod_{k=1}^{K} \left\{ \frac{1}{\sqrt{2\pi \alpha_{lk}^2}} \exp \left( -\frac{(w_{ln} - \mu_k)^2}{2\alpha_{lk}^2} \right) \right\} c_{lnk}$$ (10)

The nonnegative sparsity parameter $\alpha_{lk}$ is inversely proportional to the width of the Gaussian function. Consider a situation that $\alpha_{lk} \to \infty$, the Gaussian function will
be concentrated at the mean value \( \mu_{lk} \), which renders the MAP estimations of channel coefficients to be constant. Due to the time-varying nature of wireless channels, the channel coefficients cannot be a constant value over time, unless they are concentrated at the value of zero, i.e., the \( l \)-th MPC is considered to be the measurement noise and should be removed. By setting a large threshold for \( a_l \), the irrelevant MPCs are pruned and the sparsity is obtained. The prior of the sparsity parameter \( p(a_{lk}) \) is set to be a Gamma distribution \( p(a_{lk}) = Ga(a_{lk}|a_{lk}, b_{lk}) \) with a non-informative hyperprior for all branches, i.e., \( a_{lk} = b_{lk} = 10^{-7}, k = 1, 2, \ldots, K \). Such choice is proper when no prior information is available, and it has been widely applied in various Bayesian hierarchical models, e.g., [5,7,8].

Figure 1. Graphical Model for VB-GMM algorithm.

3. GMM-Based Variational Bayesian Learning

In this section, the VSBL is applied to the estimation problem of the Gaussian mixture channel model. First, the basic principle of variational Bayesian inference is introduced. Then the update expression of each variable in the graphical model is derived. Lastly, the appropriate initialization algorithm and corresponding parameter settings are given.

3.1. Variational Bayesian Inference

Consider a probabilistic model defined by observed variables \( O \), hidden variable sets \( H \), and the corresponding joint probability density \( p(H,O) \). Our target is to obtain the posterior \( p(H|O) = p(H,O)/p(O) \). However, the computation of the marginal likelihood function \( p(O) \) is intractable since it involves a multi-dimensional and complicated integral \( \int p(H,O)dH \). Therefore, a PDF with a simpler structure is introduced to approximate the posterior distribution, i.e.,

\[
q(H) \approx p(H|O) \tag{11}
\]

where \( q(H) \) is the variational posterior distribution (VPD). In the variational Bayesian framework, the Kullback-Leibler (KL) divergence is defined as

\[
\text{KL}(q || p) = \int q(H) \ln \left( \frac{q(H)}{p(H|O)} \right) dH \tag{12}
\]

To obtain a deterministic approximation of the posterior, (12) should be minimized, and the optimal VPD \( q^*(H) \) is achieved only if \( \text{KL}(q || p) = 0 \). Using Bayes theorem to replace \( p(H|O) \), it makes the derivation of VPD a tractable optimization problem:

\[
q^*(H) = \arg \max_{q(H)} \int q(H) \ln \left( \frac{p(H,O)}{q(H)} \right) dH \tag{13}
\]
For simplicity, we adopt the mean-field theory and assume the joint PDF factorizes as
\[
q(\mathcal{H}) = \prod_{s=1}^{S} q_s(\mathcal{H}_s) \tag{14}
\]

For most of the variables including \{\{w, a, c, \pi, x\}\}, (13) is optimized over the exponential family of distributions. In this case, the KL divergence vanishes, and the optimal solution of (13) is assumed to be achievable. For \(\mu\) and \(\theta\), the optimization on \(q_s(\mathcal{H}_s)\) is restricted on Dirac measures to obtain point estimates. By optimizing (13), the solution that minimizes the KL divergence is:
\[
q_s^*(\mathcal{H}_s) \propto \exp\left( E_{q(MB(\mathcal{H}_s))} \ln\{p(\mathcal{H}, \mathcal{O})\} \right) \tag{15}
\]

In (15), \(E_{q(x)}(\cdot)\) denotes the expectation with respect to the distribution \(q(x)_{MB(\mathcal{H}_s)}\) represents the Markov blanket \([25]\) of the variable \(\mathcal{H}_s\). In the Bayesian network, the Markov Blanket of a node is the set of the nodes comprised of the node’s parents, its children, and its co-parents. I.e., \(MB(\mathcal{H}_s)\) is the minimal set around \(\mathcal{H}_s\) which makes \(\mathcal{H}_s\) independent of the rest of the variables.

### 3.2. Update of the Estimation Expressions

As discussed in Section 2, a hierarchical Bayesian model is set up to find sparse MPCs and infer the parameter sets \(\Omega = \{\theta, w, \mu, c\}\) from the measurements \(y\). A detailed derivation of the VPDs for each factor is given below. The VPDs can be updated in any order during the iteration process.

1. **Estimation of \(q(x_l)\):** From the graphical model in Figure 1 and (15), the VPD of \(x_l\) is expressed as
\[
q(x_l) \propto \exp\left( E_{q(wl)} q(\theta_l) (\ln\{p(x_l|y, \theta_\mathcal{I}, w_\mathcal{I})\}) \right), \tag{16}
\]

In (16), \(\mathcal{I}\) is the set of all MPCs. Note that scattering parameters \(\hat{\theta}_l\) should be determined for a specific MPC, thus, we define \(q(\hat{\theta}_l)\) to be a point estimation, i.e., \(q(\hat{\theta}_l) = \delta(\theta_l - \hat{\theta}_l)\), where \(\hat{\theta}_l\) denotes the estimated value for \(\theta_l\) in a former update iteration. Based on the Gaussian noise assumption in Section 2.1, we have
\[
p(x_l|y, \theta_\mathcal{I}, \omega_{\mathcal{I}}) = CN(x_l|y - \sum_{p \in \mathcal{I}} \omega_p s(\theta_p), \Delta_l) \tag{17}
\]

By substituting \(p(x_l|\theta_p, \omega_p)\) and (17) into (16), the admissible hidden data of the \(n\)-th observation obeys a Gaussian distribution with the mean and covariance matrix as follows:
\[
\hat{x}_{ln} = (1 - \beta) \hat{w}_{ln}s(\hat{\theta}_l) + \beta \left( y_n - \sum_{p \in \mathcal{I}} \hat{w}_p s(\hat{\theta}_p) \right) \tag{18}
\]
\[
\hat{\Sigma}_l' = \left( \Delta_l^{-1} + \Sigma_l^{-1} \right)^{-1} \tag{19}
\]

2. **Estimation of \(q(c_l)\):** Evaluating (15) with (5) and (10), the VPD of \(c_l\) takes the form
\[
q(c_l) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} r_{lnk} \tag{20}
\]

in which
\[
r_{lnk} = \exp\left( \psi(\hat{\gamma}_{lk}) - \psi\left( \sum_{k=1}^{K} \hat{\gamma}_{ik} \right) + \frac{1}{2} \left[ \psi(\hat{\beta}_{lk}) - \ln(\hat{\beta}_{lk}) \right] - \hat{\gamma}_{lk}^2 \right) - \hat{\beta}_{lk}^2 \tag{21}
\]
Entropy 2021, 23, 1268

7 of 13

and \( \psi(\cdot) \) represents the di-gamma function. Since \( c_1 \) is a 0–1 matrix with a discrete VPD, the posterior expectation of \( c_{ink} \) is \( E_{q(c_1)}[c_{ink}] = q(c_{ink} = 0) \times 0 + q(c_{ink} = 1) \times 1 = r_{ink} \).

3. **Estimation of \( q(w_i) \):** From the graph the Markov blanket of \( w_i \) is \( \{x_i, a_i, \theta_i, c_i, \mu_i\} \), thus

\[
q(w_i) \propto \exp \left( E_{q(\text{MB}\{w_i\})} \left( \ln \{p(x_i|\theta_i, w_i) p(w_i|a_i, c_i, \mu_i)\} \right) \right)
\]

(22)

By integrating the factors in \( \text{MB}\{w_i\} \), the VPD can be written as \( q(w_i) = \prod_{n=1}^{N} q(w_{in}) = \prod_{n=1}^{N} \mathcal{CN}(\delta_{in}', \Phi_{in}') \) with

\[
\delta_{in}' = \Phi_{in}' \left( s(\hat{\theta}_l)^H \Sigma_l^{-1} \bar{x}_{in} + \sum_{k=1}^{K} \hat{r}_{ink} \hat{\theta}_{lk} \right)
\]

(23)

\[
\Phi_{in}' = \left( \sum_{k=1}^{K} \hat{r}_{ink} \hat{\theta}_{lk} + s(\hat{\theta}_l)^H \Sigma_l^{-1} s(\hat{\theta}_l) \right)^{-1}
\]

(24)

4. **Estimation of \( q(\theta_i) \):** Similarly, evaluating (15) leads to \( \bar{p}(\theta_i) \propto p(\theta_i) \exp \left( E_{q(x_i)} q(w_i) \left( \ln \{p(x_i|w_i, \theta_i)\} \right) \right) \). As mentioned above, the proxy PDF is defined as \( q(\theta_i) = \delta(\hat{\theta}_l - \hat{\theta}_i) \) in order to obtain a point estimation. Therefore, we have

\[
\hat{\theta}_i' = \arg\max_{\hat{\theta}_i} \left( \ln p(\theta_i) + \sum_{n=1}^{N} |\delta_{in}' s(\hat{\theta}_l)^H \Sigma_l^{-1} \bar{x}_{in}'| \right)
\]

(25)

5. **Estimation of \( q(a_i) \):** From the graphic model the Markov blanket of \( a_i \) is \( \text{MB}\{a_i\} = \{w_i, c_i, \mu_i\} \) and \( q(a_i) \propto p(a_i) \exp \left( E_{q(w_i), c_i, \mu_i} \left( \ln \{p(w_i|a_i, c_i, \mu_i)\} \right) \right) \). Due to the gamma hyperprior \( p(a_i) = \mathcal{Ga}(a_{ik}|\alpha_{ik}, \beta_{ik}) \) and the nature of conjugate prior, the VPD also satisfies the Gamma distribution, i.e., \( q(a_{ik}) \propto \mathcal{Ga}(a_{ik}|\hat{a}_{ik}', \hat{\beta}_{ik}') \) with

\[
\hat{a}_{ik}' = a_{ik} + \frac{1}{2} \sum_{n=1}^{N} \hat{r}_{ink}
\]

(26)

\[
\hat{\beta}_{ik}' = b_{ik} + \frac{N}{2} \sum_{n=1}^{N} \left| \delta_{in}' + \Phi_{in}' - 2\hat{r}_{ink} \delta_{in}' + \frac{\hat{r}_{ink}^2}{2} \right|
\]

(27)

6. **Estimation of \( q(\pi_i) \):** The only variable related to \( \pi_i \) is \( c_i \), thus, \( q(\pi_i) \propto p(\pi_i) \exp \left( E_{q(c_i)} \left( \ln p(c_i|\pi_i) \right) \right) \). By substituting (5) and (6) into (15) we obtain

\[
q(\pi_i) \propto \prod_{k=1}^{K} \pi_k \prod_{n=1}^{N} r_{ink}^{-1}
\]

(28)

which indicates that the VPD also satisfies the Dirichlet distribution. The parameter of the VPD can be updated by

\[
\hat{\gamma}_{ik}' = \gamma_{ik} + \sum_{n=1}^{N} r_{ink}
\]

(29)

7. **Estimation of \( q(\mu_i) \):** The Markov blanket of \( \mu_i \) is \( \text{MB}\{\mu_i\} = \{w_i, c_i, a_i\} \), so that

\[
\bar{p}(\mu_i) \propto p(\mu_i) \exp \left( E_{q(w_i, c_i, a_i)} \ln \{p(w_i|a_i, c_i, \mu_i)\} \right)
\]

(30)
For simplicity, the prior of the mean vector $\mu$ is set to be flat, and a delta VPD $q(\mu_l) = \delta(\mu_l - \tilde{\mu}_l)$ is also applied. By maximizing $\tilde{p}(\mu_l)$, a point estimation can be obtained as follows:

$$\hat{\mu}_l = \frac{\sum_{n=1}^{N} p_{ln} \hat{w}_{ln}}{\sum_{n=1}^{N} \hat{w}_{ln}}$$

(31)

3.3. Initialization Algorithm

The initialization algorithm is proposed to infer the initial variational variables, including the preliminary estimation of the parameter set $\Omega$ and the GMM model parameters. As shown in Algorithm 1, the estimation for multipath parameters $w_l$ and $\theta_l$ is obtained by cyclically finding and removing the corresponding template function $s(\theta_l)$ from the remaining observations. The estimation process stops when the number of multipath reaches the preset amount $L_0$. This method will inevitably result in duplicated estimations with the same scatter parameter vectors. Therefore, the channel taps with the same $\theta_l$ are checked and combined to be one. After obtaining the multipath coefficients $w_l$, we use the expectation-maximization (EM) algorithm [21] to estimate the initial parameters of $K$ Gaussian clusters. In the E-step, the probability that $w_l$ is classified as the $k$-th Gaussian component, i.e., $p(c_{lnk} = 1|w_{ln})$, is computed by the relative weight of the Gaussian likelihoods. In the M-step, $\mu_{lk}$ and $\Lambda_{lk}$ are updated by the statistics of observations. When the EM process converges, we obtain $r_{lnk} = p(c_{lnk} = 1|w_{ln})$. Notice that when initializing the EM algorithm, improper mean and variance values may cause the estimation of $p(c_{lnk} = 1|w_{ln})$ failing to converge. Thus in this paper, the K-means algorithm [26] is applied to estimate the initial mean and variance of each cluster.

Algorithm 1: Initialization

Set $L = 1$, $\alpha_l = 0$, $r_{mk} = 0$, $I = O$; $x_l \leftarrow z$

for $l = 1 \rightarrow L_0$

\begin{itemize}
  \item Initialize $q(\theta_l)$ from (25)
  \item Initialize $q(w_l)$ from (23)
  \item Initialize $q(\alpha_l)$ from (26) and (27)
  \item $\hat{x}_{ln} \leftarrow y_n - \sum_{l=1}^{L} \hat{w}_{ln} s(\hat{\theta}_l)$
  \item $I = I \cup l$
  \item $L = L + 1$
\end{itemize}

end for

// remove duplicate estimates

for $l \in I$

\begin{itemize}
  \item if $\forall d, \theta_d = \theta_l$
    \begin{itemize}
      \item $w_l = w_l + w_d$
      \item $I = I - d$
      \item $L = L - 1$
    \end{itemize}
\end{itemize}

end if

// initialize GMM parameters

for $l \in I$

\begin{itemize}
  \item initialize $\mu^0_l, \Lambda^0_l$ from $w_l$ by K-means algorithm
  \item estimate $\mu_l, \Lambda_l$, $p(c_{lnk} = 1|w_{ln})$ by EM algorithm
  \item $r_{lnk} = p(c_{lnk} = 1|w_{ln})$
\end{itemize}

end

3.4. Pruning and Convergence Condition

A close examination of the sparsity prior (10) for the $l$-th multipath reveals that the sparsity parameter $\alpha_{lk}$ controls the width of the Gaussian prior. When $\alpha_{lk} \rightarrow \infty$ the cluster will be concentrated around the center which tends to be zero for a fading channel. Thus, large values of $\alpha_l$ will make the component “irrelevant” and be removed from the estimation. In the proposed method, we set $\sum_{k=1}^{K} \alpha_{lk} > TH_p$ as a pruning condition where
$TH_p$ is a preset threshold. In order to halt the iterative variational process, another preset threshold $\Delta_0$ is used to check whether the estimation converges. Define the update rate for the $i$-th iteration

$$\Delta^{(i)} = \frac{\sum_{l=1}^{L} \sum_{n=1}^{N} |w_{ln}^{(i)} - w_{ln}^{(i-1)}|}{\sum_{l=1}^{L} \sum_{n=1}^{N} |w_{ln}^{(i)}|}$$

(32)

The iteration process stops when $\Delta^{(i)} \leq \Delta_0$. The general flow chart of the VB-GMM algorithm is shown in Figure 2.

![Flow chart of VB-GMM algorithm](image)

**Figure 2.** The flow chart of the VB-GMM algorithm.

### 3.5. Computational Complexity

The computational complexity of the proposed algorithm is discussed in this section. For the benchmark comparison, we introduce the Gaussian prior-based variational Bayesian estimation scheme proposed in [5] (VB-G). The complex multiplication times (CMT) is used as the comparison metric for the evaluation of the computational complexity in VB-G and VB-GMM. Note that the multiplication of a complex matrix and a complex matrix is $4N^3$ [27]. The computational complexity of each procedure is given in Table 1, in which the iterations taken by the variational updates in VB-G, the EM process, the K-means algorithm, and the initialization process of VB-GMM has greater complexity due to an iterative estimation for GMM parameters, in which the computational complexity of the K-means algorithm is $O(T_3 NK)$ [28].

In the variational inference procedure, the coefficient of the main complexity term $ND^3$ in VB-GMM is much smaller, resulting in the running time of about $1/5$ of the VB-G in numerical simulations.
In the numerical simulation, the setup of the sounding parameters is given below. We use the constant amplitude zero auto-correlation sequence as the sounding sequence with 256 chips, and the interval is $T_c = 5 \mu s$. The sample rate at the receiver is set as $R_s = 1$ MHz which implies that the number of discrete samples per CIR is $D = 1280$. The number of observations per WSS region is assumed as $N = 500$. For the multipath channel, a 4-tap model is adopted in which the mean power of each tap is set as $0, -3, -5, -10$ dB, and the multipath delay equals 10, 12, 15, 20 $\mu$s, respectively. For testing the performance of VB-GMM in various complex environments, three different fading types are considered. In the first scenario, the coefficients of the MPC are generated from Gaussian mixture distributions with $K = 3$. The real and imaginary parts of Gaussian means, covariances, and the mixing coefficients are randomly selected from the interval $[-10, 10],[0, 1]$ and $[0,1]$, respectively. In the second scenario, all four multipaths are modeled as Weibull distribution with the shape parameter set as 3, 4, 5, 6, respectively. In the third scenario, we define the channel taps to be Nakagami-m distributed, and the $m$-parameters are set as 2, 3, 4, 5, respectively, to simulate different fading depths.

The numerical setup for the VB-GMM algorithm is as follows. The preset multipath amount is $L_0 = 20$, and the pruning threshold is set as $TH_p = 10^8$ for all SNR conditions. In the VBE-step, the initial parameters are: $a_i = 0, \beta = 1$ and a flat prior probability $p(\theta_i)$. In the VBM-step, the initial values of the parameters are: $a_{ij} = b_{ik} = 10^{-2}, \gamma_{ik} = 1/K$. Note that all the numerical simulations are based on over 5000 observations, and the iteration process is terminated when the update rate $A(i)$ is smaller than 0.001.

Figure 3a shows a scatter plot of the estimated Gaussian mixture channel taps with the noise level SNR = $-2$ dB, in which the circles represent cluster centroids and the dots represent the estimated channel taps. The graph indicates that when the channel coefficients are Gaussian mixture distributed, the proposed VB-GMM is capable of classifying channel tap samples that come from different Gaussian components. Meanwhile, the cluster centroids can be precisely located. In Figure 3b,c, we examined the classification capability of the proposed algorithm when the channel taps are assumed to be Weibull and Nakagami distributed, respectively. As shown in Figure 3b,c, the channel coefficients in Weibull and Nakagami scenarios with random phases are clearly clustered into three regions, which indicates that the proposed method can learn and fit the distribution of random channels.

Figure 4a represents the RMSE between the synthetic and reconstructed CIRs versus SNR in different scenarios. Notice that the VB-G scheme adopts a fixed sensitivity level of $SNR_t = 10$ dB. For low SNR of $-5$ dB, the VB-GMM exhibits comparable performance with the VB-G. As SNR increases, the VB-GMM clearly outperforms the VB-G in all three scenarios due to the flexibility in fitting non-Gaussian data. The reason is that for VB-SAGE the estimation errors of $\theta_i$, i.e., the delay of MPCs in part of the observations will result in a significant RMSE between the synthetic and reconstructed channel responses. While for VB-GMM, the multipath observations make the estimate of $\theta_i$ more robust because $\theta_i$ is assumed invariant in the WSS region. Figure 4b shows the curves of the estimated mean model order versus SNR. In the SNR range of $-5$ dB to 4 dB, the value of sparsity parameters tends to diverge, resulting in an obvious suppression effect on the MPCs. Under most SNR conditions, the VB-GMM shows accurate model selection for the 4-tap multipath channel, while the VB-G has a positive model order bias. Because for some of the observations if $\theta_i$ is wrongly estimated, there will be residual in the received signal after removing the template function, which leads to the falsely identified components. From

4. Simulation Results

Table 1. The CMT required by VB-G and VB-GMM algorithms.

| Procedure                  | VB-G                                      | VB-GMM                                   |
|---------------------------|-------------------------------------------|-------------------------------------------|
| Initialization            | $(6D^3 + 4D^2 + \frac{1}{2}DL_0)N_{DL_0}$ | $(ND(D + 1)^2 + 4N^2K + \frac{1}{2}DT_0)N_{DL_0} + (D^2 + K)NK_{T_2} + O(T_3NK)$ |
| Variational inference     | $\sum_{i=1}^T NDL_i + (6D^3 + 3D^2 + 3D)NT_i$ | $\sum_{i=1}^T NDL_i + (ND(D + 1)^2 + N(K^2 + 2K + 4) + 4N^2+2K)\bar{T}_4$ |

## Table 1.

| Procedure                  | VB-G                                      | VB-GMM                                   |
|---------------------------|-------------------------------------------|-------------------------------------------|
| Initialization            | $(6D^3 + 4D^2 + \frac{1}{2}DL_0)N_{DL_0}$ | $(ND(D + 1)^2 + 4N^2K + \frac{1}{2}DT_0)N_{DL_0} + (D^2 + K)NK_{T_2} + O(T_3NK)$ |
| Variational inference     | $\sum_{i=1}^T NDL_i + (6D^3 + 3D^2 + 3D)NT_i$ | $\sum_{i=1}^T NDL_i + (ND(D + 1)^2 + N(K^2 + 2K + 4) + 4N^2+2K)\bar{T}_4$ |
the results of Figures 3 and 4, the proposed method demonstrates its superiority in terms of model complexity and estimation accuracy. In Figure 5 we demonstrate curves of RMSEs that vary with the increasing iteration number in different scenarios for an SNR of 10 dB. The proposed method performs a similar convergence rate to the Gaussian prior-based scheme with smaller RMSEs in all the scenarios.

Figure 3. Scatter plot of estimated channel coefficients in different scenarios. (a) Gaussian mixture distributed channels; (b) Weibull distributed channels; (c) Nakagami distributed channels.

Figure 4. (a) RMSE between the synthetic and reconstructed channel responses; (b) The mean value of estimated model order.

Figure 5. The RMSE of estimated channel responses versus iteration numbers.
To study the influence of the number of Gaussian components $K$, we generated over 30,000 channel responses from the Weibull and Nakagami channels and the performance of the proposed VB-GMM is given in Figure 6. In the simulation, the SNR is fixed at 10 dB. As depicted in Figure 6a, the RMSEs of the reconstructed channel responses have no significant change as $K$ increases in both Weibull and Nakagami channels. From Figure 6b, it can be seen that the estimated model orders are accurate when $K \geq 2$. Consider that a large $K$ cannot bring considerate improvement and may increase the complexity of GMM, we propose to set $K = 3$ for most channel conditions.

![Figure 6. The performance of VB-GMM versus the number of Gaussian components. (a) RMSE between the synthetic and reconstructed channel responses; (b) The mean value of estimated model order.](image)

5. Conclusions

In this paper, we presented a variational Bayesian-based channel parameter estimation approach for Gaussian mixture distributed wireless channels, which combined the variational Bayesian estimation [5] and the variational GMM training [21]. Compared with the existing Bayesian estimation schemes, the proposed scheme can exploit the sparse solutions of the multipath channel and represent the complicated distribution patterns by the application of GMMs. Experiment results showed that the VB-GMM method can obtain the channel estimates with a small reconstruction RMSE and determine the optimal number of MPCs through a fixed pruning criterion under various SNR conditions. For future works, we will further investigate factors that may affect the algorithm performance in terms of the sparsity and the implementation complexity.

**Author Contributions:** Conceptualization, J.W.; Data curation, L.K.; Investigation, H.Z. and J.W.; Methodology, X.Z.; Supervision, J.W.; Writing—original draft, L.K.; Writing—review & editing, X.Z. and H.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China under Grant 61931020.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.
