Sparse Bayesian Hierarchical Prior Modeling Based Cooperative Spectrum Sensing in Wideband Cognitive Radio Networks

Feng Li and Zongben Xu

***Abstract*—This letter proposes a new method for cooperative spectrum sensing by exploiting sparsity. The novel scheme uses the theory of Bayesian hierarchical prior modeling in the framework of sparse Bayesian learning. This model has sparsity-inducing penalization terms leading to sparser solutions compared with typically l_1 norm based ones. Based on the factor graph that represents the signal model of the hierarchical prior models, the variational message passing (VMP) algorithm is implemented to estimate the power spectral density (PSD) map.**

***Index Terms*—Bayesian hierarchical model, cognitive radio, compressive sensing, cooperative spectrum sensing, sparse estimation, variational message passing.**

I. Introduction

CO-OPERATION among multiple cognitive radio (CR) users can exploit spatial diversity to enhance spectrum sensing performance [1]. However, long sensing delay and high complexity weigh heavily against the implementation of traditional spectrum sensing approaches in CR systems. Based on compressed sensing (CS) theory, the problem can be relaxed by exploiting the sparsity. In fact, the sparsity is twofold. Firstly, the frequency band that is used by primary users always occupies a tiny part of the system bandwidth resulting in the sparsity in frequency domain. Secondly, the number of active transmitters is always very small and the locations of them only occupy a tiny fraction of the possible locations, therefore we can exploit the sparsity in space domain [2].

The estimate of PSD does not need to be considerable accurate since it is only used to distinguish the used frequency bands from unused ones. Therefore, it is meaningful to study the basis expansion model (BEM) based PSD estimator [3]. Then the problem of PSD estimation is transformed to sparse signal reconstruction with regard to the expansion coefficients.

The problem of sparse signal representations is often transformed to the problem of the l_1 norm based algorithms [4]. Sparse Bayesian learning (SBL) is also a useful approach [5]. In [5], a conditional prior based two-layer (2-L$_a$) structure is constructed to calculate the unknowns. Recently, the layered prior model and variational message passing (VMP) based approaches are used to do sparse channel estimation [6].

The contributions of this letter are as follows: Firstly, sparse Bayesian hierarchical prior modeling based cooperative spectrum sensing is studied. Secondly, the estimation error of PSD at each CR user is considered. Lastly, the closed form of the solution is obtained via VMP.

Notation: Let $\mathbf{1}_N$ be the $N \times 1$ vector of all ones and $\otimes$ be the Kronecker product. Let $\mathcal{N}(\mathbf{E}, \mathbf{V})$ and $\mathcal{CN}(\mathbf{E}, \mathbf{V})$ be the real and complex multivariate Gaussian probability density function (PDF), respectively. Let $\mathrm{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$ be the Gamma PDF and $J_x(\bullet)$ be the modified Bessel function of the second kind with order x. Finally, $\mathbf{E}(h(x), f(x))$ is the expectation of function $h(x)$ with respect to the density $f(x)$.

II. System Model

A. Signal Model

We consider a system with N_c CR users and N_t candidate sources. The position coordinate of the candidate sources is denoted as $\mathbf{C}_t$, $t = 1, 2, \cdots, N_t$. Let $h_{tr}(k, l)$ and L_{tr} denote the channel impulse response (CIR) of lth path at time k and the number of multipaths between the tth source and the rth receiver, respectively. The received signal with position coordinate $\mathbf{C}_m$ at time k can be written as

$$r_m(k) = \sum_{t=1}^{N_t} \sum_{l=0}^{L_{tr}-1} h_{tr}(k, l) s_t(k - l) + n_m(k), \quad (1)$$

where $s_t(k)$ is the transmitted signal at time k of the tth source and $n_m(k)$ is the additive white Gaussian noise (AWGN).

Every L received symbol is collected to form a frame. A constant channel is assumed over the time interval of every frame. It is assumed that the CIR is stationary and uncorrelated across different blocks, across path lags, and across t and r in space domain. Let $H_{tr}(k, n)$ be the frequency response of $h_{tr}(k, l)$ with pathloss model $\mathbf{E}(|H_{tr}(k, n)|^2) = \beta(\|\mathbf{C}_r - \mathbf{C}_t\|) = \beta_{tr}$ where β_{tr} is a given function of the distance between the transmitter and the receiver.

We assume that $s_t(k)$ is stationary, mutually uncorrelated. The PSD of $s_t(k)$ is written as $P_t(f)$. Let B be the total system bandwidth. We use N_b to denote the number of non-overlapping unit height rectangular bases $b_n(f)$ of the basis expansion model (BEM), i.e., $b_n(f) = 1$, if and only if $(n-1)B/N_b < f \leq$

Manuscript received September 24, 2013; revised February 25, 2014; accepted March 07, 2014. Date of publication March 13, 2014; date of current version March 19, 2014. This work was supported in part by the National Natural Science Foundation of China under Grants 11131006, 61075054, and 11301413, by the International Cooperation and Exchanges Project under Grant S2010GR0902, and by the China Postdoctoral Science Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zoltan Safar.

F. Li is with the Department of Information and Communication Engineering, Xian Jiaotong University, Shaanxi, China (e-mail: lf1981@mail.xjtu.edu.cn).

Z. Xu is with the Institute for Information and System Sciences, Xi'an Jiaotong University, Shaanxi, China (e-mail: zbxu@mail.xjtu.edu.cn).

Digital Object Identifier 10.1109/LSP.2014.2311902

$nB/N_b, n = 1, 2, \cdots, N_b$. Furthermore, $P_t(f)$ can be written as

$$P_t(f) = \sum\nolimits_{n=1}^{N_b} \alpha_{tn} b_n(f), \tag{2}$$

where α_{tn} is the expansion coefficient indicating the transmitted power of the transmitters.

Based on (1), the PSD of the received signal of the mth CR user can be written as

$$P_m(f) = \sum_{t=1}^{N_t} \beta_{tm} P_t(f) + \sigma_m. \tag{3}$$

By inserting (2) into (3), we will get

$$\begin{aligned} P_m(f) &= \sum_{t=1}^{N_t} \beta_{tm} \sum_{n=1}^{N_b} \alpha_{tn} b_n(f) + \sigma_m \\ &= \boldsymbol{\zeta}_m^T(f)\boldsymbol{\alpha} + \sigma_m, \end{aligned} \tag{4}$$

where σ_m is the receiver variance of the mth CR user, $\boldsymbol{\zeta}_m^T(f)$ and $\boldsymbol{\alpha}^T$ are $N_t N_b \times 1$ vectors with $\boldsymbol{\zeta}_m(f) = [\beta_{1m} b_1(f), \cdots, \beta_{1m} b_{N_b}(f), \cdots, \beta_{N_t m} b_{N_b}(f)]^T$ and $\boldsymbol{\alpha} = [\alpha_{11}, \cdots, \alpha_{1N_b}, \cdots, \alpha_{tn}, \cdots, \alpha_{N_t N_b}]^T$.

Various methods can be applied to obtain the estimate of $P_m(f)$ written as $\widehat{P}_m(f)$ [7]. We do not consider the process in detail in this letter. Instead, for simplicity, we assume that $\widehat{P}_m(f)$ has the form of

$$\widehat{P}_m(f) = P_m(f) + \delta_m(f), \tag{5}$$

where $\delta_m(f)$ is the additive white Gaussian noise with variance v_m^{-1}. In the rest of the paper, we will introduce an estimator of $\boldsymbol{\alpha}$ based on $\widehat{P}_m(f_i)$ at each CR user at frequencies $\{f_i = 2\pi i/L, 0 \le i \le L-1\}$. Based on (4) and (5), we have $\mathbf{y}_m = \boldsymbol{\Psi}_m \boldsymbol{\alpha} + \sigma_m \mathbf{1}_p + \boldsymbol{\delta}_m$, where $\mathbf{y}_m$ and $\boldsymbol{\delta}_m$ are $L \times 1$ vectors obtained by stacking $\widehat{P}_m(f_i)$ and $\delta_m(f_i)$, respectively. Also, $\boldsymbol{\Psi}_m$ is the matrix with rows $\boldsymbol{\zeta}_m^T(f_i)$. It is assumed that the sensing information obtained by all of the CR users is available at the central unit, and then the signal model at the central unit can be written as

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\Psi}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\sigma} + \boldsymbol{\delta} \\ &= \boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{\omega}, \end{aligned} \tag{6}$$

where $\mathbf{y}$ and $\boldsymbol{\delta}$ are $N_c L \times 1$ vectors obtained by stacking $\mathbf{y}_m$ and $\boldsymbol{\delta}_m$, respectively. The $N_c L \times N_t N_b$ matrix $\boldsymbol{\Psi}$ is obtained by stacking the N_c $\boldsymbol{\Psi}_m$'s. Also, $\mathbf{Z} = \mathbf{I}_{N_c} \otimes \mathbf{1}_L$, $\boldsymbol{\sigma} = [\sigma_1, \cdots, \sigma_{N_c}]^T$, $\boldsymbol{\omega}$ is a $N_c L \times 1$ white Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}_{\boldsymbol{\omega}}$, $\boldsymbol{\mu} = \mathbf{Z}\sigma$, $\mathbf{C}_{\boldsymbol{\omega}} = \mathrm{diag}^{-1}(\mathbf{Z}\mathbf{v})$ and $\mathbf{v} = [v_1, \cdots, v_{N_c}]^T$.

The main task of spectrum sensing is to estimate the sparse vector $\boldsymbol{\alpha}$ through (6), and byproduct, the positions of the transmitting radios. This letter uses the virtual grid model proposed in [8]. The transmitting radios are assumed to be located at known candidate coordinates based on the virtual grid model.

III. Layered Hierarchical Prior Model for Sparse Estimation

The SBL theory aims at solving the sparse maximum a posterior (MAP) estimate problem with respect to $\boldsymbol{\alpha}$:

$$\begin{aligned} \boldsymbol{\alpha}_{MAP} &= \arg\max_{\boldsymbol{\alpha}} p\left(\boldsymbol{\alpha}|\mathbf{y}\right) \\ &= \arg\min_{\boldsymbol{\alpha}} \{-\log p\left(\mathbf{y}|\boldsymbol{\alpha}\right) - \log p\left(\boldsymbol{\alpha}\right)\}. \end{aligned} \tag{7}$$

The prior $p(\alpha)$ is modeled based on a hierarchical structure including a conditional prior $p(\alpha|\boldsymbol{\chi})$ and a hyperprior $p(\boldsymbol{\chi})$. The prior parameter $\boldsymbol{\chi}$ is called the sparsity parameter which is inversely proportional to the width of the PDF. Using this prior, the weight of each element of α_n is controlled by χ_n. The larger χ_n is, the closer α_n approaches zero. Then the estimation will be sparse. The advantage of this hierarchical structure is twofold. Firstly, by carefully designing the formulation of the prior PDFs, we can construct inference algorithms that can obtain enhanced sparsity solutions and analytical expressions. Secondly, the 2-L$_a$ hierarchical structure can be extended to three-layer (3-L$_a$) by viewing $\boldsymbol{\chi}$ as random vector. Therefore, the third layer has more degree of freedom to control the sparsity of the solutions of the inference approaches.

We consider both the 2-L$_a$ and 3-L$_a$ hierarchical models. The two models have different sparsity-inducing properties. At first, we will demonstrate the probabilistic model of the SBL algorithms for model (6).

A. Two-Layer Hierarchical Prior Model

For the 2-L$_a$ hierarchical model, the joint PDF of signal model (6) can be written as

$$p(\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\chi}, \boldsymbol{\mu}, \mathbf{v}) = p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{v}) p(\mathbf{v}) p(\boldsymbol{\mu}) p(\boldsymbol{\alpha}|\boldsymbol{\chi}) p(\boldsymbol{\chi}). \tag{8}$$

Based on (8), the likelihood function is Gaussian with $p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{v}) = \mathcal{N}(\boldsymbol{\Psi}\alpha + \boldsymbol{\mu}, \mathbf{C}_{\boldsymbol{\omega}})$ and $p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{v}) = \mathcal{CN}(\boldsymbol{\Psi}\alpha + \boldsymbol{\mu}, \mathbf{C}_{\boldsymbol{\omega}})$ for real and complex value systems, respectively. With regard to $\mathbf{v}$, we assume that the elements of it are independent identically distributed (i.i.d) with each $p(v_n)$ being selected as a Gamma prior with $p(v_n) = \mathrm{Ga}(v_n|a_n, b_n)$ and then $p(\mathbf{Z}\mathbf{v}) = \prod_{n=1}^{N_c} p^L(v_n) = \prod_{n=1}^{N_c} \mathrm{Ga}^L(v_n|a_n, b_n)$. Obviously, let a_n and b_n be zero resulting in a non-informative prior. We assume that the elements of $\boldsymbol{\sigma}$ are i.i.d Gaussian random variables with $p(\boldsymbol{\sigma}) = \mathcal{N}(\mathbf{E}_{\boldsymbol{\sigma}}, \mathbf{v}_{\boldsymbol{\sigma}})$ and $p(\boldsymbol{\sigma}) = \mathcal{CN}(\mathbf{E}_{\boldsymbol{\sigma}}, \mathbf{v}_{\boldsymbol{\sigma}})$ for real and complex value, respectively. The conditional prior $p(\boldsymbol{\alpha}|\boldsymbol{\chi}) = \prod_{n=1}^{N_t N_b} p(\alpha_n|\chi_n)$ is selected to be the product of Gaussian PDFs with $p(\alpha_n|\chi_n) = (\frac{\rho}{\pi\chi_n})^\rho e^{-\rho\frac{|\alpha_n|^2}{\chi_n}}$ where $\rho = 1/2$ and $\rho = 1$ for real and complex value, respectively. We assume that $p(\boldsymbol{\chi}) = \prod_{n=1}^{N_t N_b} p(\chi_n)$ where $p(\chi_n) = \mathrm{Ga}(\chi_n|\xi, \tau_n)$. The prior of $\boldsymbol{\alpha}$ can be computed as

$$p(\boldsymbol{\alpha}) = \int_0^\infty p(\boldsymbol{\alpha}|\boldsymbol{\chi}) p(\boldsymbol{\chi}) d\boldsymbol{\chi} = \prod_{n=1}^{N_t N_b} p(\alpha_n), \tag{9}$$

where

$$p(\alpha_n) = \frac{2\rho^{\frac{\xi+\rho}{2}}}{\pi^\rho \Gamma(\xi)} \tau_l^{\frac{\xi+\rho}{2}} |\alpha_n|^{\xi-\rho} J_{\xi-\rho}(2\sqrt{\rho\tau_n}|\alpha_n|). \tag{10}$$

When $\xi = 3/2$ and $\rho = 1$ for complex value, using the identity $J_{1/2}(x) = (\pi/2x)^{1/2} e^{-x}$ we will get $p(\alpha_n) = \frac{2\tau_n}{\pi} e^{-2\sqrt{\tau_n}|\alpha_n|}$. While for real value with $\xi = 1$ and $\rho = 1/2$, $p(\alpha_n) = \sqrt{\frac{\tau_n}{2}} e^{-\sqrt{2\tau_n}|\alpha_n|}$. Obviously, Lasso cost function is obtained in this situation. Furthermore, it is found that as ξ decreases to zero, we will obtain a sparser solution [6].

B. Three-Layer Hierarchical Prior Model

The 2-L$_a$ model can be extended to the 3-L$_a$ model directly by introducing the regularization parameter $\boldsymbol{\tau}$ into the inference framework. It is assumed that $p(\boldsymbol{\tau}) = \prod_{n=1}^{N_t N_b} p(\tau_n)$ where $p(\tau_n) = \mathrm{Ga}(\tau_n|d_n, q_n)$. Let $\mathbf{d} = [d_1, \cdots, d_n]^T$ and $\mathbf{q} = [q_1, \cdots, q_n]^T$. Then $p(\boldsymbol{\alpha})$ can be computed as $p(\boldsymbol{\alpha}) = \prod_{n=1}^{N_t N_b} p(\alpha_n)$ with

$$\begin{aligned} p(\alpha_n) &= \int_0^\infty p(\alpha_n|\tau_n)p(\tau_n)d\tau_n \\ &= \left(\frac{\rho}{\pi q_n}\right)^\rho \frac{\Gamma(\xi+d_n)\Gamma(d_n+\rho)}{\Gamma(\xi)\Gamma(d_n)}\left(\rho\frac{|\alpha_n|^2}{q_n}\right)^{\xi-\rho} \\ &\quad \times U\left(\xi+d_n;\xi-\rho+1;\rho\frac{|\alpha_n|^2}{q_n}\right), \end{aligned} \tag{11}$$

where $U(\cdot;\cdot;\cdot)$ is the confluent hypergeometric function [9].

IV. Variational Message Passing Based Sparse Estimate

Define $\boldsymbol{\Omega} = \{\boldsymbol{\alpha}, \boldsymbol{\chi}, \boldsymbol{\mu}, \boldsymbol{\tau}, \mathbf{v}\}$ as the set of unknown parameters. Let $p(\mathbf{y}, \boldsymbol{\Omega})$ be the joint PDF of (8) with the factor graph as shown in Fig. 1. In fact, $p(\boldsymbol{\alpha}|\mathbf{y})$ in (7) can be written as $p(\boldsymbol{\alpha}|\mathbf{y}) = \int p(\boldsymbol{\Omega}|\mathbf{y})d\boldsymbol{\chi}d\boldsymbol{\mu}d\boldsymbol{\tau}d\mathbf{v}$ by regarding $\boldsymbol{\chi}$, $\boldsymbol{\mu}$, $\boldsymbol{\tau}$ and $\mathbf{v}$ as nuisance parameters and "integrating them out". The core of the theory of variational inference is to find a simple distribution $f(\boldsymbol{\Omega})$ that is very close to $p(\boldsymbol{\Omega}|\mathbf{y})$. Furthermore, $f(\boldsymbol{\Omega})$ is always assumed to have simplified form and can be factorized as $f(\boldsymbol{\Omega}) = f(\boldsymbol{\chi})f(\boldsymbol{\mu})f(\boldsymbol{\tau})f(\mathbf{v})f(\boldsymbol{\alpha})$. Kullback - Leibler (KL) divergence is used as a measure of the dissimilarity between two probability distributions. In order to obtain the $f(\boldsymbol{\Omega})$ that most resembles $p(\boldsymbol{\Omega}|\mathbf{y})$, we need to solve the problem of minimize the KL divergence between them. VMP is a useful tool to deal with this problem. The points of the VMP theory are shown as follows [10]. The functions $f(\Omega_i), \Omega_i \in \boldsymbol{\Omega}$ are computed iteratively and updated as the product of incoming messages from the neighbour factor nodes g_n to the variable node Ω_i, i.e.,

$$f(\Omega_i) \propto \prod_{g_n \in S_{\Omega_i}} m(g_n \to \Omega_i), \tag{12}$$

where S_{Ω_i} denotes the set composed by the factor nodes that neighbour the variable node Ω_i. Also, $m(g_n \to \Omega_i)$ is the message from the factor node g_n to the variable node Ω_i with

$$m(g_n \to \Omega_i) = \exp\left(\mathbf{E}(\ln g_n, \prod_l q(\Omega_l)), \Omega_l \in S_{g_n}\backslash\{\Omega_i\}\right), \tag{13}$$

where S_{g_n} is the set of the neighbouring variable nodes of the factor node g_n. Since the 2-L$_a$ model can be viewed as the 3-L$_a$ model with $f(\tau_i) = \delta(\tau_i - \hat{\tau}_i)$ where $\delta(\cdot)$ is the Dirac delta function and $\hat{\tau}_i$ is a fixed number, we will compute the messages for the 3-L$_a$ VMP algorithms. We refer to the proposed VMP based algorithms using the 2-L$_a$ and 3-L$_a$ prior models as VMP-2L$_a$ and VMP-3L$_a$ respectively.

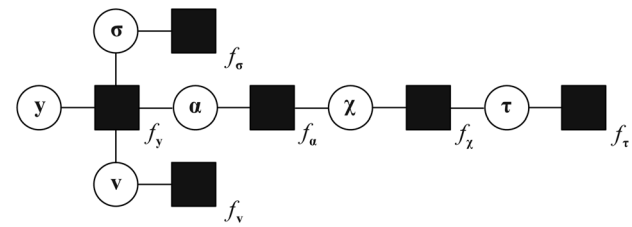

Fig. 1. Factor graph of signal model of 3-L$_a$ prior model.

1) $f(\boldsymbol{\alpha})$: Let $\mathbf{A}_1 = \mathbf{y} - \boldsymbol{\Psi}\alpha - \mathbf{Z}\mathbf{E}(\boldsymbol{\sigma})$. Updating $f(\boldsymbol{\alpha})$ involves computing the product of messages $m(g_\mathbf{y} \to \boldsymbol{\alpha})$ and $m(g_{\boldsymbol{\alpha}} \to \boldsymbol{\alpha})$ with

$$\begin{aligned} m(g_\mathbf{y} \to \boldsymbol{\alpha}) &= \exp\{\mathbf{E}(\ln p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{v}), f(\mathbf{v})f(\boldsymbol{\sigma}))\} \\ &\propto \exp\{-\rho\mathbf{A}_1^H[\mathrm{diag}(\mathbf{Z}\mathbf{v}_E)]\mathbf{A}_1\} \end{aligned} \tag{14}$$

$$\begin{aligned} m(g_{\boldsymbol{\alpha}} \to \boldsymbol{\alpha}) &= \exp\{\mathbf{E}(\ln p(\boldsymbol{\alpha}|\boldsymbol{\chi}), f(\boldsymbol{\chi}))\} \\ &\propto \exp\{-\rho\boldsymbol{\alpha}^H\mathbf{D}(\boldsymbol{\chi})\boldsymbol{\alpha}\}, \end{aligned} \tag{15}$$

where $\mathbf{v}_E = [\mathbf{E}(v_1, f(\mathbf{v})), \cdots, \mathbf{E}(v_{N_c}, f(\mathbf{v}))]^T$ and $\mathbf{D}(\boldsymbol{\chi}) = \mathrm{diag}\{\mathbf{E}({\chi_1}^{-1}, f(\boldsymbol{\chi})), \cdots, \mathbf{E}({\chi_{N_t N_b}}^{-1}, f(\boldsymbol{\chi}))\}$. Multiplying (14) and (15) yields Gaussian PDF with covariance $\mathbf{C}_{\boldsymbol{\alpha}}$ and mean $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ shown in (16) and (17) where $\mathbf{A}_2 = (\boldsymbol{\Psi}^H\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^H\mathrm{diag}^{-1}(\mathbf{Z}\mathbf{v}_E)\boldsymbol{\Psi}(\boldsymbol{\Psi}^H\boldsymbol{\Psi})^{-1}$.

$$\mathbf{C}_{\boldsymbol{\alpha}} = [\mathbf{A}_2 + \mathbf{D}(\boldsymbol{\chi})]^{-1} \tag{16}$$

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \mathbf{C}_{\boldsymbol{\alpha}}\mathbf{A}_2^{-1}(\boldsymbol{\Psi}^H\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^H[\mathbf{y} - \mathbf{Z}\mathbf{E}(\boldsymbol{\sigma})] \tag{17}$$

2) $f(\boldsymbol{\chi})$: We need to compute $m(g_{\boldsymbol{\alpha}} \to \boldsymbol{\chi})$ and $m(g_{\boldsymbol{\chi}} \to \boldsymbol{\chi})$ to update $f(\boldsymbol{\chi})$ with

$$\begin{aligned} m(g_{\boldsymbol{\alpha}} \to \boldsymbol{\chi}) &= \exp\{\mathbf{E}(\ln p(\boldsymbol{\alpha}|\boldsymbol{\chi}), f(\boldsymbol{\alpha}))\} \\ &\propto \prod_n {\chi_n}^{-\rho}\exp\{-\rho\chi_n^{-1}\mathbf{E}(|\alpha_n|^2, f(\boldsymbol{\alpha}))\} \end{aligned} \tag{18}$$

$$\begin{aligned} m(g_{\boldsymbol{\chi}} \to \boldsymbol{\chi}) &= \exp\{\mathbf{E}(\ln p(\boldsymbol{\chi}), f(\boldsymbol{\chi}))\} \\ &\propto \prod_n {\chi_n}^{\xi-1}\exp\{-\chi_n\mathbf{E}(\tau_n, f(\boldsymbol{\tau}))\}. \end{aligned} \tag{19}$$

By multiplying (18) and (19), we will obtain

$$f(\boldsymbol{\chi}) \propto \prod_n {\chi_n}^{\xi-\rho-1}\exp(-{\chi_n}^{-1}\rho\mathbf{E}(|\alpha_n|^2, f(\boldsymbol{\alpha}))), \tag{20}$$

which can be viewed as the product of generalized inverse Gaussian (GIG) PDFs with order $\xi - \rho$. Based on the properties of GIG distribution [11], we have

$$\begin{aligned} \mathbf{E}({\chi_n}^m, f(\boldsymbol{\chi})) &= \left(\frac{\rho\mathbf{E}(|\alpha_n|^2, f(\boldsymbol{\alpha}))}{\mathbf{E}(\tau_n, f(\boldsymbol{\tau}))}\right)^{\frac{m}{2}} \\ &\quad \times \frac{J_{\xi-\rho+m}(2\sqrt{\rho\mathbf{E}(|\alpha_n|^2, f(\boldsymbol{\alpha}))\mathbf{E}(\tau_n, f(\boldsymbol{\tau}))})}{J_{\xi-\rho}(2\sqrt{\rho\mathbf{E}(|\alpha_n|^2, f(\boldsymbol{\alpha}))\mathbf{E}(\tau_n, f(\boldsymbol{\tau}))})}. \end{aligned} \tag{21}$$

Using (21), $\mathrm{D}(\boldsymbol{\chi})$ in (16) can be computed directly.

3) $f(\boldsymbol{\tau})$: In order to update $q(\boldsymbol{\tau})$, $m(g_{\boldsymbol{\tau}} \to \boldsymbol{\tau})$ and $m(g_{\boldsymbol{\chi}} \to \boldsymbol{\tau})$ need to be computed with

$$f(\boldsymbol{\tau}) \propto \prod_n {\tau_n}^{\xi+d_n-1}\exp(-(\mathbf{E}(\chi_n, f(\boldsymbol{\chi})) + q_n))\tau_n \tag{22}$$

being viewed as the product of Gamma PDFs. In (21), we have

$$\mathbf{E}(\tau_n, f(\boldsymbol{\tau})) = \frac{\xi + d_n}{(\mathbf{E}(\chi_n, f(\boldsymbol{\chi})) + q_n}. \tag{23}$$

4) $f(\mathbf{v})$: The update of $f(\mathbf{v})$ is the product of messages $m(g_{\mathbf{y}} \to \mathbf{v})$ and $m(g_{\mathbf{v}} \to \mathbf{v})$ with

$$\begin{aligned} m(g_{\mathbf{y}} \to \mathbf{v}) &= \exp(\mathbf{E}(\ln p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{v}), f(\boldsymbol{\sigma}) f(\boldsymbol{\alpha}))) \\ &\propto |\mathbf{C}_{\boldsymbol{\omega}}|^{-\rho} \exp\{-\rho[\mathbf{t}]^H |\mathbf{C}_{\boldsymbol{\omega}}|^{-1}[\mathbf{t}]\}, \end{aligned} \tag{24}$$

where $\mathbf{t} = \mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\mu}_\alpha - \mathbf{Z}\mathbf{E}(\boldsymbol{\sigma})$ and t_n is the nth element of $\mathbf{t}$. By assuming that $f(\mathbf{v}) = \prod_n f(v_n)$, we will have $f(v_n) = \mathrm{G}a(v_n|a'_n, b'_n)$ where $a'_n = a_n - \rho$ for $(n-1)N_c \le n < lN_c$ and $b'_n = b_n - t_n^2$ for $(n-1)N_c \le n < lN_c$. In (16) and (17), $\mathbf{E}(v_n, f(\mathbf{v})) = a'_n / b'_n$.

5) $f(\boldsymbol{\sigma})$: Let $\mathbf{A}_3 = \mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\mu}_{\boldsymbol{\alpha}} - \mathbf{Z}\boldsymbol{\sigma}$. The update of $f(\boldsymbol{\sigma})$ is the product of messages $m(g_{\mathbf{y}} \to \boldsymbol{\sigma})$ and $m(g_{\boldsymbol{\sigma}} \to \boldsymbol{\sigma})$ where

$$\begin{aligned} m(g_{\mathbf{y}} \to \boldsymbol{\sigma}) &= \exp(\mathbf{E}(\ln p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{v}), f(\mathbf{v}) f(\boldsymbol{\alpha}))) \\ &\propto \exp\{-\rho \mathbf{A}_3^H \mathrm{diag}(\mathbf{Z}\mathbf{v}_{\mathbf{E}}) \mathbf{A}_3\}. \end{aligned} \tag{25}$$

Based on (25), $f(\boldsymbol{\sigma})$ follows Gaussian distribution with covariance $\mathbf{C}_{\boldsymbol{\sigma}}$ and mean $\boldsymbol{\mu}_{\boldsymbol{\sigma}}$. Define $\boldsymbol{\mu}' = (\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H(\mathbf{y} - \Psi\boldsymbol{\mu}_{\boldsymbol{\alpha}})$ and $\mathbf{C}' = (\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H \mathrm{diag}^{-1}(\mathbf{Z}\mathbf{v}_{\mathbf{E}})\mathbf{Z}(\mathbf{Z}^{\mathbf{H}}\mathbf{Z})^{-1}$, and we will have $\mathbf{C}_{\boldsymbol{\sigma}} = (\mathbf{C}'^{-1} + \mathbf{v}_{\boldsymbol{\sigma}}^{-1})^{-1}$ and $\boldsymbol{\mu}_{\boldsymbol{\sigma}} = \mathbf{C}_{\boldsymbol{\sigma}}(\mathbf{C}'^{-1}\boldsymbol{\mu}' + \mathbf{v}_{\boldsymbol{\sigma}}^{-1}\mathbf{E}_{\boldsymbol{\sigma}})$.

V. Simulation Results

In the simulation, $N_c = 4$ CR users make cooperative estimation of the PSD in both frequency and space domain. There are 3 sources with unknown positions on a grid of $N_t = 36$ candidate locations. The CR users scan $L = 1024$ frequencies from 20 to 80 MHz. Also, $N_b = 64$ rectangles are used to be frequency bases. The pathloss model obeys the inverse polynomial law with $\beta_{tr} = \min\{1, (d/d_0)^{-h}\}$, where $d = \|\mathbf{C}_t - \mathbf{C}_r\|$, $d_0 = 200$m and $h = 3.5$.

Fig. 2 shows the transmit PSDs of the first 20000 frames of the three sources. The solid line, dashed line and dashdotted line indicate the PSD of the low-frequency, mid-frequency and high-frequency, respectively. Each of them corresponds to the PSD of a CR user which is spanned by six base. For the following 20000 frames the mid-frequency source shuts off. For the last 20000 frames both of the mid-frequency and high-frequency sources shut off, and only the low-frequency remains. In fact, the vector $\boldsymbol{\alpha}$ has $N_t N_b = 2304$ elements of which only 18, 12 and 6 are non-zero for the first, the second and the last 20000 frames, respectively. At the initial step, $\mathbf{E}(v_i, f(\mathbf{v}))$ and $\mathbf{E}({\chi_i}^{-1}, q(\boldsymbol{\chi}))$ are set to be the inverse of the variance of $\mathbf{y}_i$ and the inverse of $N_t N_b$, respectively. Also, a_n and b_n are set to be 0 leading to a non-informative prior for $\mathbf{v}$. After that, the PDFs of $f(\boldsymbol{\alpha})$, $f(\boldsymbol{\chi})$, $f(\boldsymbol{\tau})$, $f(\mathbf{v})$ and $f(\boldsymbol{\sigma})$ are computed iteratively until convergence is achieved. For the VMP-2L$_a$ model, computation of $f(\boldsymbol{\tau})$ is not necessary and each element of it is set to be $N_c L$. For both VMP-2L$_a$ and VMP-3L$_a$, $d_n = 1$ and $q_n = 3 \times 10^{-5}$. A 10-taps Rayleigh fading channel model is adopted. The power delay profile is exponentially decreasing with a delay constant of four taps. In Fig. 3, the variance of the additive white Gaussian noise at each CR user is 0 dB, -5 dB, -5 dB and -10 dB, respectively. The parameter ξ is set to be

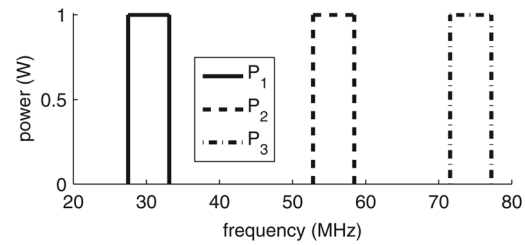

Fig. 2. Transmit PSDs of the three sources.

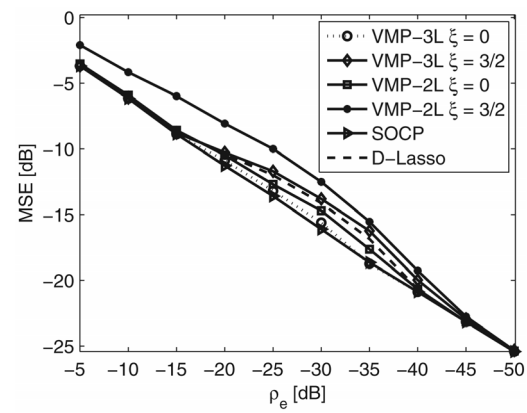

Fig. 3. Performance comparison of MSE of the algorithms.

0 and 3/2 for both VMP-2L$_a$ and 3L$_a$. Let ρ_e be the the variance of the PSD estimation error. Also, each CR user has the same value of ρ_e in the simulation. The curves demonstrates the normalized MSE performance versus ρ_e. When ρ_e is very high, the large estimation error of PSD decreases system performance seriously. The performance of the algorithms is very poor and is similar to each other. When ρ_e is very small, the estimation error of PSD can practically be neglected, and the performance of the algorithms is almost the same. Furthermore, when $\xi = 0$ the performance of the algorithms is better than $\xi = 3/2$ for both VMP-2L$_a$ and VMP-3L$_a$ algorithms. For the same ξ, VMP-3L$_a$ outperforms VMP-2L$_a$. In fact, $\xi = 3/2$ is equivalent to l_1 norm parameter constraint. In the simulation, when $\rho_e = -25$dB, VMP-2L$_a$ converges in about 50-60 iterations for both $\xi = 0$ and $\xi = 3/2$. While VMP-3L$_a$ converges in about 80-100 iterations. The performance of the second-order cone programming (SOCP) based algorithm of [12] and batch D-Lasso algorithm of [3] is also shown.

VI. Conclusion

This letter studies the problem of cooperative spectrum sensing in wideband cognitive radio networks. The sparsity of the PSD in both frequency and space domains is exploited. The PSD estimation error at each CR user and different variance of the noise at each receiver are considered. Both 2-L$_a$ and 3-L$_a$ sparse Bayesian hierarchical prior model based on BEM is constructed. VMP theory is used to solve the problem of sparse estimation and the closed form solution of the iterative estimator is obtained. We find that the proposed VMP-3L$_a$ approach with $\xi = 0$ outperforms the others in terms of MSE and sparsity. The VMP based Bayesian hierarchical model turned out to be an effective approach to solve the sparse estimation problem in CR networks.

REFERENCES

[1] E. C. Peh, Y. C. Liang, Y. L. Guan, and Y. H. Zeng, "Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 5294–5299, Nov. 2009.

[2] Z. Tian, "Compressed wideband sensing in cooperative cognitive radio networks," in *Proc. IEEE Global Communications Conf.*, New Orleans, LA, USA, 2008, pp. 1–5.

[3] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.

[4] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Statist. Soc. B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[5] D. Wipf and B. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[6] N. L. Pedersen, C. N. Manchon, D. Shutin, and B. H. Fleury, "Application of Bayesian hierarchical prior modeling to sparse channel estimation," in *Proc. Int. Conf. Communications (ICC)*, Ottawa, ON, Canada, 2012, pp. 3487–3492.

[7] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[8] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radios by exploiting sparsity," in *42nd Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct. 26-29, 2008.

[9] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, NY, USA: Dover, 1972.

[10] J. Winn and C. M. Bishop, "Variational message passing," *J. Mach. Learn. Res.*, vol. 6, pp. 661–694, Apr. 2005.

[11] B. Jorgensen, *Statistical Properties of the Generalized Inverse Gaussian Distribution (Lecture Notes in Statistics 9)*. New York, NY, USA: Springer-Verlag, 1982.

[12] J. F. Strm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optim. Meth. Softw.*, vol. 12, no. 11, pp. 625–653, 1999.