Robust frequency-hopping spectrum estimation based on sparse bayesian method

Wang, Lu; Zhao, Lifan; Bi, Guoan; Zhang, Liren; Zhang, Haijian

2015

Zhao, L., Wang, L., Bi, G., Zhang, L., & Zhang, H. (2015). Robust frequency-hopping spectrum estimation based on sparse bayesian method. IEEE transactions on wireless communications, 14(2), 781-793.

<https://hdl.handle.net/10356/107401>

<https://doi.org/10.1109/TWC.2014.2360191>

© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [Article DOI: <http://dx.doi.org/10.1109/TWC.2014.2360191>].

Downloaded on 26 Apr 2023 13:39:24 SGT

Robust Frequency-Hopping Spectrum Estimation based on Sparse Bayesian Method

Lifan Zhao, Lu Wang, Guoan Bi, *Senior Member, IEEE*, Liren Zhang, *Senior Member, IEEE*,
Haijian Zhang, *Member, IEEE*

Abstract—This paper considers the problem of estimating multiple frequency hopping signals with unknown hopping pattern. By segmenting the received signals into overlapped measurements and leveraging the property that frequency content at each time instant is intrinsically parsimonious, a sparsity-inspired high-resolution time-frequency representation (TFR) is developed to achieve robust estimation. Inspired by the sparse Bayesian learning algorithm, the problem is formulated hierarchically to induce sparsity. In addition to the sparsity, the hopping pattern is exploited via temporal-aware clustering by exerting a dependent Dirichlet process prior over the latent parametric space. The estimation accuracy of the parameters can be greatly improved by this particular information-sharing scheme, and sharp boundary of the hopping time estimation is manifested. Moreover, the proposed algorithm is further extended to multi-channel cases, where task-relation is utilized to obtain robust clustering of the latent parameters for better estimation performance. Since the problem is formulated in a full Bayesian framework, labor-intensive parameter tuning process can be avoided. Another superiority of the approach is that high-resolution instantaneous frequency estimation can be directly obtained without further refinement of the TFR. Results of numerical experiments show that the proposed algorithm can achieve superior performance particularly in low signal-to-noise ratio scenarios compared with other recently reported ones.

Index Terms—Dirichlet process, multiple frequency-hopping signals, sparse Bayesian learning, stick-breaking process, time-frequency representation.

I. INTRODUCTION

FREQUENCY-hopping (FH) signals have been widely used in wireless communications and other information systems due to their inherent capability of low interception and anti-jamming [1]–[6]. In practical civilian and military applications, estimating the parameters of the FH signals is of great importance [2], [7]. In civilian wireless networks [2], [8], for example, estimating FH signals is vital to detect multiple users, avoid collision and achieve robust performance against interference [9]. In military communications, blind estimation is also required to estimate and intercept the intentional interference [10]. The main challenges of FH signal estimation are to robustly estimate the hopping time and frequency with unknown hopping pattern particularly in low signal-to-noise ratio (SNR) environments [2], [11]. Since the maximum likelihood

estimation requires a computational intractable combinatorial search [7], [11], many other alternative methods are developed.

Among many approaches that have been reported to cope with the FH spectrum estimation problem, time-frequency representation (TFR) is an intuitive and useful tool to exhibit the frequency contents of non-stationary signals. Spectrogram obtained by short-time Fourier transform (STFT) is a popular choice for estimating FH signals, since it can provide a TF distribution free of cross-terms compared to other distributions in Cohen's class [12]. However, the spectrogram of the signal inevitably suffers from limited resolution and poor signal energy concentration, which greatly restricts the performance of subsequent processing such as entropy or gradient based refinement [5], [13], [14].

In [15], a maximum likelihood approach is developed for estimating a single FH signal. In this method, the FH signal estimation problem has been formulated in a maximum likelihood framework, where hopping time and frequency are treated as unknown parameters. Empirical results show that the algorithm has achieved desirable performance of estimating hopping time and frequency in relatively high SNR environments. Since the formulation of this algorithm depends on the parametric modeling of a single FH signal, it cannot be straightforwardly extended to multiple FH signals due to the undesirable increase in dimensionality of the parameters and the induced intractable computational complexity [15].

In [7], [16], multiple FH signal estimation is considered in a sparse linear regression (SLR) framework, where a fused least absolute shrinkage and selection operator (LASSO) alike approach is used [17]. In this framework, two penalty terms are incorporated to encourage sparsity and smoothness in the frequency and TF domain, respectively. Since this approach is based on the compressive sensing concept, it can achieve better estimation accuracy in various scenarios. Despite its success in estimating FH signals, the experimental results show that this approach does not scale well under low SNR conditions. Although the bounds of regularization parameters are given in [7] for desirable recovery, parameter tuning process is still required to obtain robust performance of the algorithm, where optimal parameter selection is still an open problem.

Another approach is developed to cope with hopping time and frequency estimation in FH networks particularly for multi-channel systems [2], [18]. An expectation maximization (EM) approach is employed to iteratively estimate the amplitude, hopping time and frequency. In this approach, the success of the EM algorithm depends largely on the proper selection of initial value, where a de-noised STFT is often used to initialize

The corresponding author is Dr Guoan Bi. Lifan Zhao, Lu Wang, Guoan Bi and Haijian Zhang are with the Infocomm Centre of Excellence, School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798 (email: zhao0145@e.ntu.edu.sg; wa0001lu@e.ntu.edu.sg; egbi@ntu.edu.sg; haijian.zhang@whu.edu.cn).

Liren Zhang is with College of Information Technology, UAE University, United Arab Emirates (lzhang@uaeu.ac.ae).

the algorithm. In low SNR environments, however, it is hard to obtain a desirable initial TFR, which will inevitably result in degraded performance of the EM algorithm.

In order to accurately and robustly estimate multiple FH signals in low SNR environments, a novel enhanced high-resolution TFR is proposed by exploiting sparsity and piecewise smoothness in the TF domain. In this paper, we develop a linear and high-resolution TFR based on the sparse Bayesian framework, encouraging sparsity in spectral and smoothness in TF domain simultaneously. More concretely, rather than reconstructing the sparse spectrum at each time instant independently to achieve high resolution [19], the proposed formulation can also exploit the piecewise smoothness in the TF domain by utilizing the dependent Dirichlet process (dDP) for the latent parameter modeling. The proposed framework can be considered as a multi-task learning problem with unknown task relations. The main idea is to exploit the frequency hopping patterns in a statistical manner.

The rest of the paper is organized as follows. In Section II, the frequency-hopping problem is introduced and the sparse signal model is formulated. In Section III, FH signal estimation problem is formulated in a sparse Bayesian framework. Further discussions on related work are also presented in this section. Multi-channel extension of the proposed method is presented in Section IV. Numerical experiments and conclusion are given in Section V and Section VI, respectively.

Notation : Vectors and matrices are denoted by bold symbol. For a matrix $\mathbf{A}$, $\mathbf{A}^H$ and $\mathbf{A}^{-1}$ denote the conjugate and inverse of the matrix, respectively. ℓ_p norm of a vector is defined by $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$. $\mathcal{N}(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathcal{CN}(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote multivariate real and complex Gaussian distributions, $\text{Bern}(x|p)$ denotes Bernoulli distribution and $\Gamma(x|a, b)$ denotes Gamma distribution.

II. PRELIMINARIES AND PROBLEM FORMULATION

A. Signal Model

Let us consider a noise corrupted and interference free FH signal,

$$\mathbf{y}(t) = \sum_{k=1}^{K_i} \sigma_{i,k} e^{j2\pi f_{i,k} t} + \mathbf{v}(t), \quad t_{i-1} < t < t_i \quad (1)$$

where K_i is the number of hopping frequencies during the i -th system dwell time, $\sigma_{i,k}$ is the amplitude of the signal associated with the i -th system dwell time and k -th hopping frequency. The noise vector is denoted by $\mathbf{v}(t)$, which is an independent identical distributed (i.i.d.) Gaussian random variable. The discrete form of $\mathbf{y}(t)$ is expressed as

$$\mathbf{y}(n) = \sum_{k=1}^{K_i} \sigma_{i,k} e^{j2\pi f_{i,k} n/f_s} + \mathbf{v}(n/f_s), \quad n_{i-1} < n < n_i \quad (2)$$

where f_s is the sampling frequency larger than the Nyquist sampling frequency, i.e., $f_s > 2 \cdot \max\{f_{i,j} : i = 1, \dots, N_d \text{ and } j = 1, \dots, K_i\}$, and N_d is the number of system-wise dwells. In practical systems, the number of signals K_i , the system dwell time n_i and hopping frequency

Fig. 1. The segmentation of the signal into overlapped measurements, where P is set to be an odd number for computational convenience.

$f_{i,k}$ are all unknown to the user and are required to be estimated.

B. The Sparsity Based Estimator via SLR

The frequency hopping signal estimation problem can be formulated within a sparse framework by exploiting the sparsity and spectral smoothness [7]. It can be formulated as the following convex optimization problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left[\frac{1}{2} \|\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{D}\mathbf{x}\|_1 \right] \quad (3)$$

where the first term is for data fitting, the second and third terms are for encouraging sparsity in frequency and differential-time domain, respectively [7]. The λ_1 and λ_2 are used as regularization parameters.

It has been shown that sparsity-induced approach can achieve state-of-the-art performance over many other methods. However, an important shortcoming of the SLR based approach is that it does not provide robust estimation performance in low SNR environments, due to the sensitivity of the differential operator $\|\mathbf{D}\mathbf{x}\|_1$ used in fused LASSO [17]. Furthermore, the convex optimization based approach requires appropriate tuning of parameters, where optimum solution is still an open problem.

C. Proposed Sparse Bayesian Model

In order to alleviate the drawbacks, a new approach is proposed to make use of the TF processing scheme by segmenting the received signals into overlapped segments, meanwhile exploiting sparsity and smoothness in time and TF domain, respectively. Let us assume a uniformly dense grid $[f_1 : \Delta f : f_N]$ including all the hopping frequencies, where Δf is the frequency interval and N is the number of frequency bins. Similar to the spectrogram based method, the signal is segmented into M over-lapped measurements, where P is the window length for the segmentation, as shown in Fig. 1. The length of the segment is often chosen to be appropriate for stable sparse recovery [20] as well as low bias in hopping transition time estimation [12], which will be formally defined in Section III-C. Then, (2) becomes

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M] = \boldsymbol{\Phi} [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M] + \mathbf{V} \quad (4)$$

where $\mathbf{y}_i \in \mathbb{C}^{P \times 1}$ denotes the i -th segment of the received signal, $\mathbf{x}_i \in \mathbb{C}^{N \times 1}$ represents the spectrum of the i -th segment to be estimated and $\mathbf{V} \in \mathbb{C}^{P \times M}$ represents the unknown noise. The matrix $\boldsymbol{\Phi} = [\Phi(f_1), \dots, \Phi(f_N)] \in \mathbb{C}^{P \times N}$ is constructed as

time index). As observed from the above-described model, the mixing coefficient is temporally related.

In addition, the weight coefficient $\mathbf{w}_k$ in (10) is further modeled as Gaussian-Gamma distribution to encourage sparsity [23],

$$p(\mathbf{w}_k|\lambda_k) = \mathcal{N}(\mathbf{w}_k|\mathbf{0}, \lambda_k) \quad (12)$$

$$p(\lambda_k|e, f) = \prod_{i=1}^M \Gamma(\lambda_{ki}|e, f). \quad (13)$$

This hierarchical modeling will induce $\mathbf{w}_k$ to be a sparse vector [23], [24]. The induction of sparsity for $\mathbf{w}_k$ is to achieve smooth segmentation since the dominant entry in each $\mathbf{w}_k$ can produce a smooth segment as indicated in [21] and our experimental results. The intuition of this particular modeling on mixing weight $\pi_k(t_i)$ in (8) is to make the clustering procedure be temporal dependent, where time-frequency smoothness of the FH signals can be well exploited.

Finally, the noise precision parameter α_0 is modeled to follow a Gamma distribution for convenient inference,

$$p(\alpha_0|a, b) = \Gamma(a, b) \quad (14)$$

where convenient inference of the noise precision parameter α_0 can be enabled for modeling conjugacy [20].

The above-described model can encourage sparsity of the spectra by hierarchical sparse modeling and temporal aware clustering, respectively. Combining the likelihood function in (5) with the above-described hierarchical priors, the joint probability can be expressed by

$$\begin{aligned} p(\mathbf{y}, \mathbf{x}, \alpha, \mathbf{z}, \mathbf{w}, \lambda, \alpha_0) = & \prod_{i=1}^M \prod_{k=1}^K p(\mathbf{y}_i|\mathbf{x}_i, \alpha_0) \cdot [p(\mathbf{x}_i|\alpha_k)]^{\xi_{ik}} \\ & \cdot p(\alpha_j|c, d) \cdot p(\mathbf{z}_{ik}|\mathbf{w}_k) \cdot p(\mathbf{w}_k|\lambda_k) \\ & \cdot p(\lambda_k|e, f) \cdot p(\alpha_0|a, b). \end{aligned} \quad (15)$$

Therefore, the estimation of these parameters can be obtained by calculating the following posterior,

$$p(\mathbf{x}, \alpha, \mathbf{z}, \mathbf{w}, \lambda, \alpha_0|\mathbf{y}) = p(\mathbf{x}, \alpha, \mathbf{z}, \mathbf{w}, \lambda, \alpha_0|\mathbf{y})/p(\mathbf{y}). \quad (16)$$

The calculation of marginal distribution $p(\mathbf{y})$, however, demands the computationally intractable multi-dimensional integral. Since the closed-form solution of the posterior is generally not accessible, approximated or sampling approach is required for Bayesian inference [26].

Remark 2: In FH signals estimation, the latent parameter α cannot be explicitly assumed to be fully independent or dependent in sparse Bayesian learning (SBL) or block SBL (BSBL). In SBL [20], [23], [24], [27], [28], the sparse signals $\mathbf{x}_i, i = 1, \dots, M$, are hierarchically modeled to impose sparsity. The prior of $\mathbf{x}_i$ is modeled as a complex Gaussian distribution with independent variance vector α_i , $p(\mathbf{x}_i|\alpha_i) = \mathcal{CN}(\mathbf{0}, \alpha_i^{-1})$, $i = 1, \dots, M$. In contrast, the block SBL (BSBL) approach [29] uses a shared α defined by exploiting the support-sharing property of the measurements. The prior of $\mathbf{x}_i$ is modeled as, $p(\mathbf{x}_i|\alpha) = \mathcal{CN}(\mathbf{0}, \alpha^{-1})$ for $i = 1, \dots, M$.

Remark 3: Rather than imposing the sparsity in frequency domain and hop point explicitly as in SLR approach, the proposed model implicitly exploits the temporal clustering

property of the linear time frequency distribution along with sparsity. In this manner, the sensitivity of differential operator can be relaxed by the proposed multi-task learning framework [21], [30]. It is seen that the sparsity helps to recover high-resolution TF distribution and the temporal clustering property enhances the recovery of TF distribution. Our empirical results have validated the robustness of the proposed method.

B. The Bayesian Inference

Since direct inference of the graphic model is generally intractable, a straightforward approach is to sample the distribution by standard Markov Chain Monte Carlo (MCMC) approach. However, the MCMC approach suffers from high computational complexity in high dimensional scenarios and its general convergence is difficult to diagnose. Instead, we use a computationally efficient approximation procedure known as variational Bayesian inference (VBI) to obtain efficient inference and the guaranteed convergence. With the mean-field assumption, the factorized posterior can be approximated via the following expression,

$$p(\mathbf{x}, \alpha, \mathbf{z}, \mathbf{w}, \lambda, \alpha_0|\mathbf{y}) \approx q(\mathbf{x})q(\alpha)q(\mathbf{z})q(\mathbf{w})q(\lambda)q(\alpha_0) \quad (17)$$

where the parameter is defined as $\Theta = \{\mathbf{x}_i, \alpha_k, \mathbf{z}_{ik}, \mathbf{w}_k, \lambda_k, \alpha_0\}$. With this assumption, let us derive the optimal approximation to minimize the Kullback Leibler (KL) divergence between the approximated distribution and true posterior,

$$q^*(\Theta) = \arg \min_{\Theta} \int q(\Theta) \log \frac{p(\Theta|\mathbf{y})}{q(\Theta)} d\Theta. \quad (18)$$

In summary, the optimal distribution for each Θ_k can be expressed as,

$$q^*(\Theta_i) \propto \exp [\mathbb{E}_{i \neq j} [\ln p(\mathbf{Y}, \Theta)]] \quad (19)$$

Since the graphic model is formulated hierarchically based on the conjugacy of the distributions, we can conveniently derive the following updating formula for the individual parameter based on the mean value of their approximated posterior distribution. The derivations of updating rules for each of the parameters are given as follows.

- 1) *Updating rule of noise precision parameter α_0 :* According to the Markov blanket of α_0 in the graphic model, the optimal approximated distribution $q(\alpha_0)$ should obey,

$$q(\alpha_0) \propto \prod_{i=1}^M p(\mathbf{y}_i|\Phi \mathbf{x}_i; \alpha_0) p(\alpha_0|a, b). \quad (20)$$

Since the conjugacy of Gaussian likelihood $\prod_{i=1}^M p(\mathbf{y}_i|\Phi \mathbf{x}_i; \alpha_0)$ and Gamma prior $p(\alpha_0|a, b)$ will result in a Gamma posterior, the update equation for the noise precision parameter α_0 can be derived by substituting (5) and (14) into (20),

$$\begin{aligned} \ln q(\alpha_0) \propto & -\alpha_0 \sum_{i=1}^M \mathbb{E} \|\mathbf{y}_i - \mathbf{A} \mathbf{x}_i\|_2^2 + nM \ln \alpha_0 \\ & + (a-1) \ln \alpha_0 - b\alpha_0. \end{aligned} \quad (21)$$

Therefore, the approximated posterior for α_0 obeys a Gamma distribution with parameters $\hat{a}$ and $\hat{b}$ expressed as,

$$\hat{a} = M \cdot n + a \quad (22)$$

$$\hat{b} = - \sum_{i=1}^M \left[\|y_i - \Phi \mu_i\|_2^2 + \text{trace}(\Phi \Sigma_i \Phi^H) \right]. \quad (23)$$

Since only the first order momentum of the parameter is utilized in the updating rules for other parameters, the mean of α_0 is given accordingly as $\mathbb{E}[\alpha_0] = \hat{a}/\hat{b}$.

- 2) *Updating rule for sparse coefficient $\mathbf{x}_i$* : According to the Markov blanket of the $\mathbf{x}_i$ in the graphic model, the optimal approximated distribution $q(\mathbf{x}_i)$ should also obey,

$$q(\mathbf{x}_i) \propto \prod_{j=1}^K [p(y_i|\mathbf{x}_i)p(\mathbf{x}_i|\alpha_j)]^{\xi_{ij}}. \quad (24)$$

Due to the conjugacy of Gaussian distribution, the optimal distribution for $\mathbf{x}_i, i = 1, \dots, M$, is given by substituting (5) and (6) into (24),

$$\begin{aligned} \ln q(\mathbf{x}_i) &\propto -\alpha_0 \cdot \|y_i - \Phi \mathbf{x}_i\|_2^2 + \sum_{k=1}^K -\xi_{ik} \mathbf{x}_i^H \alpha_k \mathbf{x}_i \\ &\propto -\mathbf{x}_i^H \left(\sum_{k=1}^K \xi_{ik} \alpha_k + \alpha_0 \Phi^H \Phi \right) \mathbf{x}_i \\ &\quad + \mathbf{x}_i^H \Phi^H y_i + y_i^H \Phi \mathbf{x}_i. \end{aligned} \quad (25)$$

Based on the above equation, the mean and covariance matrix of the approximated posterior distribution can be given as,

$$\mu_i = \alpha_0 \Sigma_i^{-1} \Phi^H y_i \quad (26)$$

$$\Sigma_i = \left[\text{diag} \left[\sum_{k=1}^K \langle \xi_{ik} \rangle \cdot \alpha_k \right] + \alpha_0 \Phi^H \Phi \right]. \quad (27)$$

It is seen that the calculation of (26) requires the inversion of matrix Σ_i . To reduce the computational complexity of the matrix inversion Σ_i^{-1} , the equation for Σ_i^{-1} can be expanded by matrix inverse lemma [20],

$$\begin{aligned} \Sigma_i^{-1} &= \text{diag} \left[\sum_{k=1}^K \langle \xi_{ik} \rangle \cdot \alpha_k \right]^{-1} - \text{diag} \left[\sum_{k=1}^K \langle \xi_{ik} \rangle \cdot \alpha_k \right]^{-1} \Phi^H \\ &\quad \cdot \mathbf{C}^{-1} \Phi \text{diag} \left[\sum_{k=1}^K \langle \xi_{ik} \rangle \cdot \alpha_k \right]^{-1} \end{aligned} \quad (28)$$

where the matrix $\mathbf{C}$ is further expressed as,

$$\mathbf{C} = \alpha_0^{-1} \mathbf{I} + \Phi \text{diag} \left[\sum_{k=1}^K \langle \xi_{ik} \rangle \cdot \alpha_k \right]^{-1} \Phi^H. \quad (29)$$

It can be observed from the above equations that the updating rule for μ_i will incorporate all the latent parameter $\alpha_k, k = 1 : K$, and the corresponding indicator variables ξ_{ik} . It is noted that when $\xi_{ik} = 1$, the i -th sparse coefficient will only be influenced by the support of latent parameter α_k . More concretely,

the sparse coefficient with $\xi_{ik} = 1, i = 1 : M$, will be influenced by the clustering of the latent parameter α_k .

- 3) *Updating rule for the precision vector of the sparse coefficient α* : Based on the Markov blanket of the α in Fig. 2, the optimal approximated distribution $q(\alpha_k)$ should obey,

$$q(\alpha_k) \propto \prod_{i=1}^M [p(\mathbf{x}_i|\alpha_k)]^{\xi_{ik}} p(\alpha_k|c, d). \quad (30)$$

Substituting (6) and (7) into (30), we have,

$$\begin{aligned} \ln q(\alpha_k) &\propto -\alpha_0 \cdot \sum_{i=1}^M \xi_{ik} [-\mathbf{x}_i^H \alpha_k \mathbf{x}_i + \sum_{j=1}^N \ln \alpha_{kj}] \\ &\quad + (c-1) \ln \alpha_k - d \alpha_k. \end{aligned} \quad (31)$$

Based on the above equation, the mean and covariance matrices of the approximated posterior distribution are given as,

$$\hat{c}_{kj} = c + \sum_{i=1}^M \langle \xi_{ik} \rangle \quad (32)$$

$$\hat{d}_{kj} = d + \sum_{i=1}^M \langle \xi_{ik} \rangle \cdot [\mu_{ij}^2 + \Sigma_i(j, j)]. \quad (33)$$

Similar to the updating rule for $\mathbf{x}_i$ who utilizes all the latent parameters α_k , estimation of each α_k also incorporates all the estimation of sparse coefficient $\mathbf{x}_i$ and the corresponding indicator parameters ξ_{ik} .

- 4) *Updating rule for the indicator variable z_{ik}* : Because the posterior of z_{ik} is only dependent on $\mathbf{x}_i$, we can obtain

$$q(z_{ik}) \propto \prod_{j=k}^K [p(\mathbf{x}_i|\alpha_j)]^{\xi_{ij}} p[z_{ik}|\sigma(g_k(t_i))] \quad (34)$$

where $\mathbf{w}_k^T \psi_i$ is defined as $g_k(t_i)$ for notational brevity. Substituting (6) and (10) into (34), the approximated posterior is given in (35) in next page. Since the prior of z_{ik} is not conjugate to its likelihood, the derivation of the posterior is given in the following proposition.

Proposition 1: The probability of $p(z_{ik}) = 1$ and $p(z_{ik}) = 0$ are given by (55) and (56) in Appendix A, respectively.

Proof: The proof is given in Appendix B. ■

It is observed from the derivation that the updating rule for z_{ik} involves both the sparse signals and their corresponding latent parameters.

- 5) *Updating rule for the weight coefficient $\mathbf{w}_k$* : Based on the Markov blanket of $\mathbf{w}_k$ in Fig. 2, the optimal approximated distribution $q(\mathbf{w}_k)$ should obey,

$$q(\mathbf{w}_k) \propto p(\mathbf{w}_k|\lambda_k^{-1}) \prod_{i=1}^M p[g_k(t_i)]. \quad (36)$$

In general, the inference is intractable due to the complexity in the logistic function, where we seek a

$$\ln q(z_{ik}) \propto \sum_{j=k+1}^K z_{ik} [\xi_{ij}] \left[\sum_{n=1}^N \ln \alpha_{jn} - \mathbf{x}_i^H \text{diag}(\boldsymbol{\alpha}_j) \hat{\mathbf{x}}_i \right] + z_{ik} \ln \sigma(g_k(t_i)) + (1 - z_{ik}) [-\mathbf{w}_k^T \boldsymbol{\psi}_i + \ln \sigma(g_k(t_i))] \quad (35)$$

variational bound for logistic function approximation.

Proposition 2: Based on the quadratic variational bound of the logistic function², the approximated posterior of $\mathbf{w}_k$ obeys a Gaussian distribution with mean and covariance matrices given by,

$$\hat{\mathbf{w}}_k = \boldsymbol{\Sigma}_{w_k} \cdot \sum_{i=1}^M [\hat{z}_{ik} - 0.5] \cdot \boldsymbol{\psi}_i \quad (37)$$

$$\boldsymbol{\Sigma}_{w_k} = [\lambda_k \mathbf{I} + \sum_{i=1}^M 2 \cdot \boldsymbol{\psi}_i \boldsymbol{\psi}_i^H f(\eta_{ik})]^{-1}. \quad (38)$$

Proof: The proof is given in Appendix C. ■

- 6) *Updating rule for λ_k :* Based on the Markov blanket of λ_k in Fig. 2, the optimal approximated distribution $q(\lambda_k)$ should obey,

$$q(\lambda_k) \propto p(\mathbf{w}_k | \lambda_k^{-1}) p(\lambda_k | e, f). \quad (39)$$

Similarly, the parameter λ_k obeys a Gamma distribution,

$$\ln q(\lambda_k) \propto -\frac{1}{2} \mathbf{w}_k^T \lambda_k \mathbf{w}_k + \frac{1}{2} \ln |\lambda_k| + \sum_{i=1}^M [(e - 1) \cdot \ln \lambda_{ki} - f \cdot \lambda_{ki}]. \quad (40)$$

The approximated posterior for $q(\lambda_k)$ obeys a Gamma distribution with parameter $\hat{e}$ and $\hat{f}$ given as,

$$\hat{e}_k = \left(\frac{1}{2} + e \right) \cdot \mathbf{1}_g \quad (41)$$

$$\hat{f}_k = \frac{1}{2} \cdot \text{diag}(\boldsymbol{\Sigma}_{w_k} + \mathbf{w}_k \mathbf{w}_k^T) + f \quad (42)$$

where $\mathbf{1}_g$ is a g -dimensional vector composed of 1.

In summary, the proposed algorithm works in a round-robin manner to iteratively update these parameters, which is given in Algorithm 1. Remarkably, the convergence of the algorithm is guaranteed within this full variational Bayesian framework [31].

C. Analysis and Discussion

Sharing information among related tasks will benefit the estimation. However, sharing information among unrelated tasks will induce undesirable bias, which is known as negative transfer in machine learning literature [32]. In this paper, the idea is to encourage the segmented measurements with the same frequency content to cluster in the latent parameter space and to sufficiently separate unrelated tasks.

It is noted that the proposed approach is a biased one by introducing hierarchical sparse modeling and logistic stick

²The variational bound of logistic function is given as $\sigma(x) \geq \sigma(\eta) \exp[0.5(x - \eta) - f(\eta)(x^2 - \eta^2)]$, where η is the variational parameter and $f(\eta)$ is defined as $\tanh(\eta/2)/(4\eta)$ [26].

Algorithm 1 Frequency hopping signals estimation based on the proposed sparse Bayesian method

```

1: Input:  $\mathbf{y}_i, \Phi, \boldsymbol{\alpha}, a, b, c, d, e, f$ .
2: while  $\sim$  Converge do
3:   Update  $\alpha_0$  by (22) and (23)
4:   for  $i = 1 : M$  do
5:     Update  $\boldsymbol{\mu}_i$  and  $\boldsymbol{\Sigma}_i$  by (26) and (28);
6:   end for
7:   for  $k = 1 : K$  do
8:     Update  $\boldsymbol{\alpha}_k$  by (32) and (33);
9:     Update  $\mathbf{w}_k$  by (37) and (38);
10:    Update  $\lambda_k$  by (41) and (42);
11:    for  $i = 1 : M$  do
12:      Update  $z_{ik}$  by (55) and (56);
13:    end for
14:  end for
15: end while
16: Output:  $\mathbf{x}_i, i = 1, \dots, M$ .
```

breaking process into a multi-task compressive sensing framework. However, this bias can encourage to obtain smooth frequency profile in the time-frequency domain as well as accurate hopping pattern estimation.

The intuitive idea of the proposed approach is to utilize sparse signal recovery technique to obtain high resolution time frequency estimation, where temporal aware latent space clustering is imposed to obtain enhanced signal recovery.

Definition 1: Without loss of generality, let us consider a signal segment that has one frequency hop between two active frequencies. Let us denote the hopping time index and window length by T and P , respectively. The hopping transition time index I with respect to window length P is then defined as,

$$I = \{I : I \in T - (P - 1)/2 : T + (P - 1)/2\}. \quad (43)$$

Theorem 1: In noiseless case, i.e., $\mathbf{V} = 0$ in (3), under the assumption that $\boldsymbol{\alpha}_k$ and the corresponding indicator function ξ_{ik} are correctly estimated, the algorithm is guaranteed to give correct time-frequency representation of the measurements $\mathbf{x}_i$ during hopping transition time.

Proof: The proof is given in Appendix D. ■

D. Computational Complexity

The STFT requires the least computational complexity to achieve representation performance that should be further improved. The computational complexity of the recently reported SLR method implemented by alternating direction method of multiplier (ADMoM) is in the order of $O(M^2 N^2)$ for one iteration. The computational complexity of the proposed algorithm mainly depends on the matrix inverse operation in

Fig. 3. An illustrative example of frequency hopping signal estimation. The FH signal estimation results obtained by spectrogram, SLR, BCS-DP and proposed algorithm are given in (a), (c), (e) and (g), and their corresponding hopping time statistics are given in (b), (d), (f) and (h), respectively.

calculating (28) and (38). Therefore, it is in the order of $O(MP^3 + Kg^3)$ for one iteration. In general, P is chosen to be a small value compared to M and g is also chosen to be smaller than M . It is seen that the computational complexities of the proposed algorithm and SLR are in a similar order.

IV. MULTI-CHANNEL FH ESTIMATION

In practical information systems, such as FH wireless network, multiple channel systems are often deployed in order to obtain robust performance against frequency collision and interference [2], [8]. In the estimation process, the direction-of-arrival (DOA), hopping time and frequency are all required to be estimated. In this section, we extend the previously proposed algorithm to multi-channel scenarios.

Assuming K_s sources are impinging on a uniform linear array (ULA) with L sensors, the obtained measurement for each snapshot can be given as

$$\mathbf{y}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{v}(n), n = 1, \dots, N_s \quad (44)$$

where $\mathbf{y}(n) \in \mathbb{C}^{L \times 1}$, $\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_{N_0})] \in \mathbb{C}^{L \times N_0}$, $\mathbf{s}(n) \in \mathbb{C}^{N_0 \times 1}$ contains the FH signals, N_s is the number of snapshots. The noise $\mathbf{v}(n)$ is assumed to be spatially and temporally uncorrelated,

$$\mathbb{E}[\mathbf{v}(n_1)\mathbf{v}^H(n_2)] = \alpha_0^{-1}\mathbf{I} \cdot \delta_{n_1, n_2}. \quad (45)$$

To estimate the DOA from a sparsity perspective, matrix $\tilde{\mathbf{A}} \triangleq [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_{N_0})] \in \mathbb{C}^{L \times N_0}$ is an over-complete dictionary, whose i -th atom

$$\mathbf{a}(\theta_i) = [e^{j\pi \cos \theta_i}, e^{j2\pi \cos \theta_i}, \dots, e^{jL\pi \cos \theta_i}]^H \quad (46)$$

corresponds to an ideal steering vector from spatial angle θ_i .

An integrated model can be formulated by iteratively estimating the DOAs and hopping frequencies via a variational sparse Bayesian framework. Rather than the complicated integrated framework, a simple scheme is proposed in this paper to carry out the DOA estimation and FH signal estimation in two separate stages. The intuition behind this scheme is

that clustering on the latent parameter space can be more easily carried out with a smaller number of clusters, because the projection manipulation in the DOA estimation stage will simplify the clustering procedure in the FH estimation stage. More concretely, we argue that the proposed approach can particularly benefit from this scheme for the simplification of latent parameter clustering. In Section V, our empirical results show that the proposed scheme can give desirable performances, even in the scenarios that DOA estimation is biased.

In order to efficiently estimate the DOA, ℓ_1 -svd algorithm [33] is utilized by exploiting sparsity in the spatial domain, where other popular sparse regularized approach for DOA estimation can also be used [34], [35]. Empirical results have demonstrated that sparsity-driven algorithms can provide robust estimation of DOA whether the sources are uncorrelated or correlated.

Assuming N_s snapshots are collected and denoting the measurement $\tilde{\mathbf{Y}}$ as $[\mathbf{y}(1), \dots, \mathbf{y}(N_s)]$, it is obtained

$$\tilde{\mathbf{Y}} = \mathbf{a}^H(\hat{\theta}_i)\mathbf{Y}, \quad i = 1, \dots, K_s. \quad (47)$$

After the beam-forming manipulation and substituting (44) into (47), we obtain

$$\tilde{\mathbf{Y}} = \mathbf{a}^H(\hat{\theta}_i)\mathbf{A}\mathbf{s} + \mathbf{V} \quad (48)$$

$$= \mathbf{s}_i + \sum_{j=1, j \neq i}^{K_s} \mathbf{a}^H(\hat{\theta}_i) \cdot \mathbf{a}(\theta_j) \cdot \mathbf{s}_j + \mathbf{V}. \quad (49)$$

It is noted that the projection of signal onto the i -th direction can be decomposed as the single FH signal in the i -th direction and the projection of other FH signals from other directions into the i -th direction. It can be properly assumed that because the term $\mathbf{a}^H(\hat{\theta}_i) \cdot \mathbf{a}(\theta_j)$ is relatively trivial compared with $\mathbf{s}_i$, the second and the third terms in the right hand side of (49) can be treated as noise sources. Therefore, the proposed algorithm for multi-channel frequency hopping signal estimation is summarized in Algorithm 1.

Fig. 4. The correct hopping time detection ratios for frequency hopping signal with two and three components are given in (a) and (c), and their corresponding incorrect IF detection ratios are given in (b) and (d), respectively.

Remark 4: In the multi-channel scenarios, the DOA and hopping frequency estimation are carried out in separate stages, where each hopping signal can be estimated individually. A remarkable advantage of this scheme is that the proposed approach can achieve much better frequency hopping estimation accuracy by the preprocessing that is similar to the beam-forming manipulation, since clustering of the latent parameter can be more accurately carried out for a single frequency hopping signal within this scheme.

Remark 5: This approach can also be flexibly extended to the scenario where different dwells involve different DOAs. In the modification of the algorithm, the DOA estimation is required to be carried out to account for the directional hopping in the spatial domain, which is similar to the frequency hopping estimation in the frequency domain. We do not discuss it further for brevity because the required modification is relatively straightforward.

Algorithm 2 Multi-channel FH signal estimation

- 1: Input: $\mathbf{y}_i, \Phi, \alpha, a, b, c, d, e, f$.
- 2: **I. DOA Estimation Stage;**
- 3: Use l_1 -svd algorithm or OMP to estimate DOA [33];
- 4: **II. FH Signal Estimation Stage**
- 5: **for** $j = 1 : K_s$ **do**
- 6: Beam-forming $\tilde{\mathbf{Y}}_j = \mathbf{a}^H(\hat{\theta}_j)\mathbf{Y}$ in (47);
- 7: Signal Segmentation; $\hat{\mathbf{Y}}_j = [\tilde{\mathbf{Y}}_{j1}, \dots, \tilde{\mathbf{Y}}_{jN_s}]$;
- 8: Use Algorithm 1;
- 9: **end for**
- 10: Output: $\mathbf{X}_j, j = 1, \dots, K_s$.

V. NUMERICAL EXPERIMENTS

In this section, the numerical experimental results are given to evaluate the performance of the proposed algorithm in comparison with other reported ones in both single and multiple channel cases. In the following experiments, the SNR is defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{\|\mathbf{s}\|_2^2}{N_s \sigma^2} \right) \quad (50)$$

where $\mathbf{s}$ denotes the signal vector, N_s is the number of time indices and σ^2 is the noise power.

In particular, two performance measures are defined for comparison of hopping time and instantaneous frequency (IF) detection, respectively. The correct hopping time detection ratio is defined as,

$$P_t = \left(\sum_{i=1}^{M_c} D_t(i) \right) / M_c \quad (51)$$

where M_c is the number of Monte Carlo trails and $D_t(i)$ is the number of correct detections in each Monte Carlo trial. The hopping time statistic is defined as $\Delta_n = \|x_{n+1} - x_n\|_2^2$. A correct hopping time detection is declared if the estimated hopping instant is less than 3 samples away from the associated true hopping instant, which is defined in the same way as in [7].

Fig. 6. An illustrative example of frequency hopping signal estimation for multi-channel scenarios (SNR = -10 dB and $L = 6$). The multi-channel FH signal (component 1) estimation based on STFT, SLR and proposed algorithm are given from (a.1) to (a.3) and their corresponding hopping time statistics are given from (b.1) to (b.3). The multi-channel FH signal (component 2) estimation based on STFT, SLR and proposed algorithm are given from (c.1) to (c.3) and their corresponding hopping time statistics are given from (d.1) to (d.3).

Fig. 5. The correct hopping time detection ratio based on STFT, SLR and proposed algorithm in terms of number of reconstructed frequency bins. (SNR = 2 dB)

The incorrect IF detection ratio is further defined as,

$$E_f = 1 - \left(\sum_{i=1}^{M_c} D_f(i) \right) / M_c \quad (52)$$

where $D_f(i)$ is the correct frequency detection rate in the i -th Monte Carlo trial.

A. Single Channel FH Estimation

In single channel FH estimation, experimental results of the proposed algorithm are qualitatively and quantitatively com-

pared with other recently reported ones in terms of incorrect hopping time detection ratio and incorrect IF detection ratio.

The hopping signals used in the following experiments are generated as follows: the first hopping component is active within the range of time index $[0 : 15]$ and the carrier frequency hops from 13 KHz to 18 KHz within the range of time index $[16 : 63]$. The second hopping component is active within the range of time index $[0 : 31]$ and the carrier frequency hops from 28 KHz to 23 KHz within the range of time index $[32 : 63]$. The third hopping component is active within the range of time index $[0 : 47]$ and the carrier frequency hops from 3 KHz to 6 KHz within the range of time index $[48 : 63]$. The sampling frequency f_s is 64 KHz.

The regularization parameter is chosen to be $\lambda_1 = \lambda_1^*/10$ and $\lambda_2 = \lambda_2^*/20$ to achieve desirable hopping time and frequency detection in the tested SNRs, where optimal λ_1^* and λ_2^* are given in [7]. The number of Monte Carlo trails is chosen to be 200.

An illustrative example is given in Fig. 3. In this experiment, the segment length of the signal is chosen to be 12 and SNR is 0 dB. It is seen that the spectrogram obtained in Fig. 3 can barely give a clear concentration of the signal, particularly in the first system dwell time. In contrast, all the three sparsity-driven method can give better estimation and concentration due to the utilization of sparsity regularized strategy in the time-frequency domain. Notably, the SLR and Bayesian compressive sensing with Dirichlet prior (BCS-DP) [22] approaches can only detect the second system dwell time while the first dwell time cannot be detected correctly as

Fig. 7. The correct hopping time detection of (a) two component signals and (b) three component signals, based on SLR and proposed algorithm in multi-channel scenario.

shown in Fig. 3(d) and 3(f). The reason for the degraded performance of the SLR is possibly the sensitivity of fused LASSO with heavy noise effects. Meanwhile, since the BCS-DP [32] only operates the clustering procedure for latent parameter without temporal smoothness constraint, the noise might be easily categorized into the existing clusters. The success of the proposed algorithm depends on the temporal aware clustering by utilizing the logistic stick-breaking process, which can encourage spectral smoothness and provide sharp partition simultaneously as indicated from Fig. 3(g) and (h).

In Fig. 4, Monte Carlo experiments are conducted to give quantitative evaluation of the proposed algorithm and the previously reported ones with both two components and three components in terms of hopping time and frequency estimation, respectively. In Fig. 4 (a) and (b), the performance comparison of signal with two components (component 1 and component 2) are given. It can be concluded that with the increase of SNR, SLR and the proposed algorithm can obtain better results than STFT. In particular, the correct hopping time detection ratio of SLR method is even lower than that of STFT particularly when $\text{SNR} < -2\text{dB}$. However, our proposed algorithm can achieve the best hopping time and frequency detection ratios. In Fig. 4 (c) and (d), the performance comparison of signal with three components is given. When three components are present, all of the algorithms suffer from

Fig. 8. The correct hopping time detection ratio and incorrect IF detection ratio based on FHSE-EM, SLR and proposed algorithm in multi-channel scenario.

degraded performances to a certain extent. However, among all the algorithms, the proposed approach can achieve desirable results compared to the other ones. In summary, it can be observed from these figures that the proposed algorithm can achieve the best hopping time and frequency estimation in low SNR scenarios.

Finally, the performance of the algorithm is compared in terms of the number of reconstructed frequency bins, where the hopping signal used in the experiment is the same as previous ones. The correct hopping time detection ratios obtained by different algorithms are compared, where incorrect IF detection ratios are omitted due to off-grid phenomenon in frequency domain. Figure 5 shows that the proposed algorithm can provide robust performance to achieve the best hopping time detection ratio.

B. Multi-Channel FH Estimation

The performance of algorithm with multi-channel setting is given in this section. We test both two-component and three-component signals. The signal with two components is from DOAs $\theta = [40, 60]$, while the signal with three components is from three DOAs $\theta = [40, 58, 78]$. The hopping patterns are the same as that used in the single channel experiment.

As shown in the illustrative example in Fig. 6, it can be shown that the frequency hopping signal can be separately estimated after beam-forming procedure in (47). Since the

SNR is relatively low in this experiment, the STFT cannot obtain a good signal energy concentration, while the SLR and proposed algorithm give better estimation performance. It is noted that all the three algorithms cannot completely eliminate the influence of the other signal component after beam-forming. This undesirable signal component degrades the hopping time detection based on spectrogram and SLR method as indicated in Fig. 6. It is also observed that both the spectrogram and SLR cannot give the correct hopping time estimation. In contrast, the proposed algorithm can give correct detection for both signal components.

In order to demonstrate the robustness of the proposed algorithm against DOAs estimation bias, a Monte Carlo experiment for frequency hopping signal estimation is carried out in terms of DOA estimation bias. In Fig. 7 (a) and (b), we can observe that the hopping time detection rate decreases with the increase of the DOA bias. The proposed algorithm can give better results than the SLR algorithm in terms of DOA estimation bias. In particular, the proposed algorithm can give more robust results even when the bias is as large as 5 degrees. With the increase in the DOA bias, the SLR algorithm suffers more degradation than that of the proposed algorithm. Moreover, compared with the single channel scenario, the proposed scheme in multi-channel system has superior performance over the SLR based method due to the efficiency of clustering procedures. In particular, the estimation performance of first FH signal is worse than the second one, since component 2 can be more easily estimated and detected with hopping time instant in the center of the signal. It can be concluded that the proposed algorithm gives better performance in multi-channel systems than the previously reported ones, even with a biased DOA estimation.

To validate the proposed multi-channel estimation scheme, the performance comparison is carried out in terms of SNRs, based on frequency hopping signal estimation expectation maximization (FHSE-EM) [18], SLR and the proposed one. In Fig. 8, it can be observed that the proposed method can achieve best performance in both the hopping time detection ratio and incorrect IF detection ratio over other methods. This can validate our argument that the two-stage sparsity-driven scheme is effective for multi-channel FH signal estimation.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we propose a robust frequency hopping signal estimation method based on sparse Bayesian framework. In this approach, we formulate the frequency estimation problem in a multi-task learning manner with signal being partitioned into overlapped measurements. To enhance the estimation accuracy of the sparse spectral coefficients, the latent parameters are encouraged to cluster in a temporal-aware sense by utilizing the logistic stick-breaking process. In this way, more accurate estimation can be obtained due to the use of multi-task learning framework. This novel approach not only avoids the tedious parameter tuning process, but also provides robust performance in low SNR environments.

APPENDIX

A. Review of Dependent Dirichlet Process

Dirichlet process (DP) has been defined to model random measure on measures [25]. This process has been widely used for flexible mixture modeling with an unknown number of mixtures. A DP process can be expressed as,

$$G = \sum_{i=1}^{\infty} \pi_i \delta_{\alpha_i} \quad (53)$$

where π_i is the mixing probability for the i -th mixture and δ_{α_i} is point mass located at α_i [25]. Since the explicit formulation of DP from the definition in (53) is not attainable, various constructions of DP have been proposed [36]. In the original DP, exchangeability of the data is often assumed. However, many real data exhibiting correlation does not satisfy this assumption. In order to model the temporal or spatial dependency on the mixture modeling, dependent Dirichlet process [37] is further defined by imposing a stochastic process on mixing weight and base measure,

$$G_x = \sum_{i=1}^{\infty} \pi_i(x) \delta_{\alpha_i(x)} \quad (54)$$

where $\pi_i(x)$ is the dependent mixing weight, $\delta_{\alpha_i(x)}$ is the dependent point mass at $\alpha_i(x)$ and x denotes the temporal or spatial related index. In this way, dependency is imposed on the Dirichlet process.

B. Proof of Proposition 1

When $z_{ik} = 1$, the corresponding indicator parameter $\xi_{ik} = 1$. Consequently, the approximated posterior for $z_{ik} = 1$ is given as

$$q(z_{ik} = 1) \propto \exp [\ln |\Lambda_k| - \langle \mathbf{x}_i^H \Lambda_k \mathbf{x}_i \rangle + \ln \sigma[g_k(t_i)]] \quad (55)$$

When $z_{ik} = 0$, the corresponding indicator parameters $\xi_{ij} = 0, j < k$. Consequently, the approximated posterior for $z_{ik} = 0$ can be given as

$$q(z_{ik} = 0) \propto \exp \left\{ \sum_{j=k+1}^K [\xi_{ij} \ln |\Lambda_j| - \langle \mathbf{x}_i^H \Lambda_j \mathbf{x}_i \rangle] - \mathbf{w}_k^T \psi_i + \ln \sigma[g_k(t_i)] \right\} \quad (56)$$

The resulted approximated posterior can be therefore expressed as

$$q(z_{ik} = j) = \frac{\mathbb{Z}_{ikj}}{\mathbb{Z}_{ik0} + \mathbb{Z}_{ik1}}, \quad j = 0, 1 \quad (57)$$

where $\mathbb{Z}_{ik1}$ and $\mathbb{Z}_{ik0}$ are the values calculated from the right hand side of (55) and (56).

C. Proof of Proposition 2

In the following, a variational lower bound of logistic function, i.e. Gaussian lower bound, will be applied to obtain tractable inference. According to [26] and variational bound,

the lower bound of the approximated posterior can be obtained by

$$\begin{aligned} \ln p(\mathbf{w}_k) &\propto -\frac{1}{2}\mathbf{w}_k^T \text{diag}(\lambda_k)\mathbf{w}_k \\ &+ \sum_{i=1}^M \log [\sigma(g_k(t_i))]^{z_{ik}} + \log [\sigma(g_k(t_i))]^{1-z_{ik}} \quad (58) \\ &\geq -\frac{1}{2}\mathbf{w}_k^T [\text{diag}(\lambda_k) + \sum_{i=1}^M 2\psi_i \psi_i^T f(\eta_{ik})]\mathbf{w}_k \\ &+ \sum_{i=1}^M [z_{ik} - \frac{1}{2}]\psi_i^T \mathbf{w}_k. \quad (59) \end{aligned}$$

Because it is seen that the lower bound of the approximated posterior is Gaussian distributed, the mean and covariance matrix can be easily given by rearranging the above equation to the canonical Gaussian distribution form, which is not further discussed for brevity.

Therefore, the updating rule of the mean and covariance matrix can be obtained as in (42) and (43).

D. Proof of Theorem 1

Let us consider the segmented measurements in the hopping transition time, i.e., the time $i \in [T - (P - 1)/2 : T + (P - 1)/2]$. It is assumed that the k -th latent parameter α_k is correctly estimated for sparse signal $\mathbf{x}_i$ in the i -th time index. In other words, the indicator variable ξ_{ik} equals 1. Thus, update equation for sparse signal $\mathbf{x}_i$ can be expressed as,

$$\boldsymbol{\mu}_i = \alpha_0 [\text{diag}(\alpha_k) + \alpha_0 \boldsymbol{\Phi}^H \boldsymbol{\Phi}]^{-1} \boldsymbol{\Phi}^H \mathbf{y}_i. \quad (60)$$

The i -th measurement can be decomposed as

$$\mathbf{y}_i = \boldsymbol{\Phi}_{i,1} \mathbf{x}_{i,1} + \boldsymbol{\Phi}_{i,2} \mathbf{x}_{i,2}. \quad (61)$$

The matrix $\boldsymbol{\Phi}_{i,1}$ is constructed from $\boldsymbol{\Phi}$ by replacing the last $(i - T) + 1 + (P - 1)/2$ rows of $\boldsymbol{\Phi}$ with $\mathbf{0}$, and the matrix $\boldsymbol{\Phi}_{i,2}$ is also constructed from $\boldsymbol{\Phi}$ by replacing first $(T - i) - 1 + (P + 1)/2$ rows of $\boldsymbol{\Phi}$ with $\mathbf{0}$. The position of the non-zero element in sparse signal $\mathbf{x}_{i,1}$ corresponds to the first carrier frequency index and the position of non-zero element in sparse signal $\mathbf{x}_{i,2}$ corresponds to the second carrier frequency index.

Substituting (61) into (60) and taking the limit as $\alpha_0 \rightarrow \infty$, we can obtain,

$$\begin{aligned} \lim_{\alpha_0 \rightarrow \infty} \boldsymbol{\mu}_i &= \text{diag}(\alpha_k)^{-1} \boldsymbol{\Phi}^H (\boldsymbol{\Phi}^H \text{diag}(\alpha_k)^{-1} \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}_1 \mathbf{x}_{i,1} \\ &+ \text{diag}(\alpha_k)^{-1} \boldsymbol{\Phi}^H (\boldsymbol{\Phi}^H \text{diag}(\alpha_k)^{-1} \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}_2 \mathbf{x}_{i,2}. \quad (62) \end{aligned}$$

It can be seen from the above equation that the sparsity profile is determined by the sparsity profile of $\text{diag}(\alpha_k)^{-1}$. Since support of α_k only corresponds to the correct carrier frequency bin in time i , only one of the summand terms on the right-hand side of (62) will be non-zero. More concretely, the estimated signal will only depend on α_k and the sparse signal $\mathbf{x}_{ij}$, $j = 0$ or 1 , whose support is the same as α_k^{-1} . As long as α_k and z_{ik} are correctly estimated, the algorithm can guarantee to provide correct time-frequency estimation.

REFERENCES

- [1] A.J. Viterbi, "Spread spectrum communications: myths and realities," *IEEE Communications Magazine*, vol. 40, no. 5, pp. 34–41, May 2002.
- [2] X. Liu, N. D. Sidiropoulos, and A. Swami, "Joint hop timing and frequency estimation for collision resolution in FH networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 3063–3074, 2005.
- [3] D. J. Torrieri, "Mobile frequency-hopping CDMA systems," *IEEE Transactions on Communications*, vol. 48, no. 8, pp. 1318–1327, 2000.
- [4] L. Aydin and A. Polydoros, "Hop-timing estimation for FH signals using a coarsely channelized receiver," *IEEE Transactions on Communications*, vol. 44, no. 4, pp. 516–526, 1996.
- [5] S. Barbarossa and A. Scaglione, "Parameter estimation of spread spectrum frequency-hopping signals using time-frequency distributions," in *First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, 1997, pp. 213–216.
- [6] S. Srinivasa and S. A. Jafar, "Cognitive radios for dynamic spectrum access-the throughput potential of cognitive radio: A theoretical perspective," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 73–79, 2007.
- [7] D. Angelosante, G.B. Giannakis, and N.D. Sidiropoulos, "Estimating multiple frequency-hopping signal parameters via sparse linear regression," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5044–5056, 2010.
- [8] S. Glisic, Z. Nikolic, N. Milosevic, and A. Pouttu, "Advanced frequency hopping modulation for spread spectrum WLAN," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 1, pp. 16–29, Jan. 2000.
- [9] L. Zhang, H. Wang, and T. Li, "Anti-jamming message-driven frequency hopping - part I: System design," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 70–79, 2013.
- [10] L. Zhang and T. Li, "Anti-jamming message-driven frequency hopping - part II: Capacity analysis under disguised jamming," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 80–88, 2013.
- [11] X. Liu, Nicholas D. Sidiropoulos, and A. Swami, "Blind high-resolution localization and tracking of multiple frequency hopped signals," *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 889–901, 2002.
- [12] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 1989.
- [13] M. K. Simon, U. Cheng, L. Aydin, A. Polydoros, and B. K. Levitt, "Hop timing estimation for noncoherent frequency-hopped M-FSK intercept receivers," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 1144–1154, 1995.
- [14] T. Chen, "Joint signal parameter estimation of frequency-hopping communications," *IET Communications*, vol. 6, no. 4, pp. 381–389, Mar. 2012.
- [15] C. C. Ko, Wanjun Zhi, and F. Chin, "ML-based frequency estimation and synchronization of frequency hopping signals," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 403–410, 2005.
- [16] D. Angelosante, G.B. Giannakis, and N.D. Sidiropoulos, "Sparse parametric models for robust nonstationary signal analysis: Leveraging the power of sparse regression," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 64–73, 2013.
- [17] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused LASSO," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 1, pp. 91–108, 2005.
- [18] X. Liu, J. Li, and X. Ma, "An EM algorithm for blind hop timing estimation of multiple FH signals using an array system with bandwidth mismatch," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2545–2554, 2007.
- [19] L. Stankovic, I. Orovic, S. Stankovic, and M. Amin, "Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4562–4572, 2013.
- [20] D. Wipf and B.D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions Signal Processing*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
- [21] L. Ren, L. Du, L. Carin, and D. Dunson, "Logistic stick-breaking process," *The Journal of Machine Learning Research*, vol. 12, pp. 203–239, 2011.
- [22] Y. Qi, D. Liu, D. Dunson, and L. Carin, "Multi-task compressive sensing with dirichlet process priors," in *Proceedings of the 25th international conference on Machine learning*, ACM, 2008, pp. 768–775.
- [23] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, Sep. 2001.

- [24] L. Zhao, G. Bi, L. Wang, and H. Zhang, "An improved auto-calibration algorithm based on sparse Bayesian learning framework," *IEEE Signal Processing Letters*, vol. 20, no. 9, pp. 889–892, 2013.
- [25] D. M. Blei and M. I. Jordan, "Variational inference for dirichlet process mixtures," *Bayesian analysis*, vol. 1, no. 1, pp. 121–143, 2006.
- [26] C. M. Bishop et al., *Pattern recognition and machine learning*, vol. 1, Springer New York, 2006.
- [27] L. Wang, L. Zhao, G. Bi, and C. Wan, "Hierarchical sparse signal recovery by variational Bayesian inference," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 110–113, Jan. 2014.
- [28] L. Zhao, L. Wang, G. Bi, and L. Yang, "An autofocus technique for high-resolution inverse synthetic aperture Radar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6392–6403, Oct. 2014.
- [29] D. Wipf and B.D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions Signal Processing*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.
- [30] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.
- [31] D. G. Tzikas, C. L. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, 2008.
- [32] Y. Xue, X. Liao, L. Carin, and B. Krishnapuram, "Multi-task learning for classification with dirichlet process priors," *The Journal of Machine Learning Research*, vol. 8, pp. 35–63, 2007.
- [33] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.
- [34] Z. Liu, Z. Huang, and Y. Zhou, "An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 1–11, Oct. 2012.
- [35] Z. Liu, Z. Huang, and Y. Zhou, "Sparsity-inducing direction finding for narrowband and wideband signals based on array covariance vectors," *IEEE Transactions on Wireless Communications*, vol. 12, no. 8, pp. 1–12, August 2013.
- [36] S. J. Gershman and D. M. Blei, "A tutorial on Bayesian nonparametric models," *Journal of Mathematical Psychology*, vol. 56, no. 1, pp. 1–12, 2012.
- [37] S. N. MacEachern, "Dependent nonparametric processes," in *ASA proceedings of the section on Bayesian statistical science*. American Statistical Association, pp. 50–55, Alexandria, VA, 1999, pp. 50–55.

sonar, radar, and communications.

Guoan Bi received the B.Sc. degree in radio communications from Dalian University of Technology, P. R. China, 1982, and the M.Sc. degree in telecommunication systems, and the Ph.D. degree in electronics systems, both from Essex University, U.K., in 1985 and 1988, respectively. Since 1991, he has been with the school of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His current research interests include DSP algorithms and hardware structures, and signal processing for various applications including

Dr. Zhang is currently an adjunct professor with University of South Australia. Dr. Zhang has vast experience as an engineer, academia and researcher in the field of network modeling and performance analysis, teletraffic engineering and network characterization, optical communications and networking, network resource management and quality of services (QoS) control, video communications over next generation Internet, wireless/mobile communications and networking, switching and routing algorithm, ad-hoc networks, radio frequency and electro-optical sensor networks, and information assurance and network security. He has published more than 100 research papers in international journals. Dr. Zhang is a Senior Member of IEEE and also he is the associate editor of the Journal of Computer Communications and the editor for Journal of Networks and Computer Applications.

Liren Zhang received his M.Eng. (1988) from the University of South Australia and Ph.D (1993) from the University of Adelaide, Australia, all in electronics and computer systems engineering. He joined the College of Information Technology, UAE University as Chair Professor in Network Engineering in 2009. Dr Zhang was a Senior Lecturer in Monash University, Australia (1990-1995), an Associate Professor in Nanyang Technological University, Singapore (1995-2007) and a Research Professor of Systems Performance Modeling in Defense and System Institute and Professor in Network Engineering at University of South Australia (2007-2009).

Lifan Zhao received the B.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2010. He is currently working towards the Ph.D. degree in the School of Electrical and Electronic Engineering of Nanyang Technological University, Singapore.

His research interests include compressive sensing, sparse signal recovery techniques, machine learning and their applications in source localization and wireless communications.

Lu Wang received the B.S. in electrical and electronic engineering from Xidian University in 2007 and M.Eng. degree in signal processing from National Key Lab of Radar Signal Processing (RSP), Xidian University, China, in 2010. And now she is currently pursuing the Ph.D. degree in the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Her major research interests include sparse signal processing, radar imaging (SAR/ISAR), sonar signal processing.

Haijian Zhang received the B.Sc. degree in electronic information engineering from Wuhan University, Wuhan, China, in 2006 and the joint Ph.D. degrees from Conservatoire National des Arts et Métiers (CNAM), Paris, France, and Wuhan University, in 2011. He is currently a Research Fellow with the School of Electrical Electronic Engineering, Nanyang Technological University, Singapore. His main research interests include time-frequency analysis, array signal processing, and cognitive radio.