Structure-Aware Bayesian Compressive Sensing for Frequency-Hopping Spectrum Estimation

Shengheng Liu[a,b], Yimin D. Zhang[b], Tao Shan[a], Si Qin[c], and Moeness G. Amin[c]

[a]School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
[b]Department of Electrical and Computer Engineering, College of Engineering, Temple University, Philadelphia, PA 19122, USA
[c]Center for Advanced Communications, Villanova University, Villanova, PA 19085, USA

ABSTRACT

Frequency-hopping (FH) is one of the commonly used spread spectrum techniques that finds wide applications in communications and radar systems due to its capability of low probability of intercept, reduced interference, and desirable ambiguity property. In this paper, we consider the blind estimation of the instantaneous FH spectrum without the knowledge of hopping patterns. The FH signals are analyzed in the joint time-frequency domain, where FH signals manifest themselves as sparse entries, thus inviting compressive sensing and sparse reconstruction techniques for FH spectrum estimation. In particular, the signals' piecewise-constant frequency characteristics are exploited in the reconstruction of sparse quadratic time-frequency representations. The Bayesian compressive sensing methods are applied to provide high-resolution frequency estimation. The FH spectrum characteristics are used in the design of signal-dependent kernel within the framework of structure-aware sparse reconstruction.

Keywords: Frequency-hopping, spectrum estimation, Bayesian compressive sensing, quadratic time-frequency representation, signal-dependent kernel design

1. INTRODUCTION

Frequency-hopping (FH) signals are generated by varying the carrier frequencies according to a certain hopping pattern, which is typically pseudo-random. Due to their inherent capability of low probability of intercept, reduced interference to and from other users, resistance to jamming and multipath fading, and desirable ambiguity function property, FH signals have become a favorite choice in a wide range of communication and radar applications [1–9].

For a variety of tasks ranging from interception of non-cooperative emitters to exploitation of signals of opportunity for passive sensing, estimating and tracking the instantaneous spectrum of FH signals are an important yet a challenging task, since the hopping patterns of the constituent signals are unknown [10]. The time-varying spectrum signatures of non-stationary signals, such as FH signals, can be revealed in the joint time-frequency (TF) representations. As FH signals generally exhibit sparsity in the joint TF domain, compressive sensing (CS) and sparse reconstruction techniques [11,12] enable effective FH parameter estimation by formulating the problem as an underdetermined linear regression with a dual penalty [10,13], i.e., a penalty that controls both the intrinsic sparsity and smoothness of the estimation. To improve the performance, particularly in low signal-to-noise ratio (SNR) conditions, a sparse Bayesian learning-based approach was proposed in [14]. These approaches are based on linear time-frequency analysis, and do not consider the effect of missing samples.

In this paper, we develop FH spectrum estimation in the quadratic TF representation framework. It is known that quadratic (bilinear) TF distributions, such as the Wigner-Ville distribution (WVD), provide high-resolution time-varying spectrum representations. Because the bilinear nature causes cross-terms between different components that constitute false energy distributions, various reduced-interference distributions have been developed for cross-term reduction through the design of appropriate reduced interference distribution TF kernels [15,16]. TF kernels can be signal-independent or signal-dependent. The latter generally provides better performance in

*Contact information ydzhang@temple.edu

trading off the cross-term suppression and auto-term preservation. The adaptive optimal kernel (AOK), which is based on radially Gaussian functions in the ambiguity domain [17, 18], is a commonly used signal-dependent kernel.

In practice, data observations may experience missing samples due to channel distortion, line-of-sight obstruction, and removal of impulsive noise samples [19]. Missing data samples induce noise-like artifacts in the TF distribution [20]. As described in [20–23], the effect of artifacts induced by missing samples can be substantially reduced by using TF kernels. Sparse reconstruction of time-frequency representations (TFR) using different CS methods was also presented therein. Recently, this approach has been adopted to estimate the FH spectrum from data with missing samples using the orthogonal matching pursuit (OMP) algorithm. Both artifact mitigation and high-resolution FH signal spectrum estimation are achieved [24]. Since the artifacts induced by missing samples can be filtered with kernels in the ambiguity function (AF) domain, quadratic TFR based methods are advantageous, as compared to linear approaches [10, 13, 14], particularly when missing samples are present.

Based on our previous work in [24], in this paper, we develop a new approach for FH spectrum estimation via quadratic TFRs where a modified Bayesian CS (BCS) scheme that exploits the FH signal structure is applied. In the proposed sparse reconstruction process, a structure prior is imposed to enforce the unique piece-wise constant frequency signature of the FH signals. As such, the proposed approach provides additional robustness in the FH spectrum estimation where the hopping patterns are unavailable and the observed data contain missing samples. Compared with the FH spectrum estimation via OMP [24], the structure-aware BCS-based approach achieves an improved sparsity and provides a better approximation to the ℓ_0-norm sparsity measure [25].

Notations: Lower-case (upper-case) bold characters are used to denote vectors (matrices). abs$(\cdot)$ returns the magnitude of a given complex number, and $\circ$ denotes the Hadamard product. diag$\{\cdot\}$ represents a diagonal matrix that uses the entries of a vector as its diagonal entries, and $\mathbf{I}_N$ denotes an $N \times N$ identity matrix. $\mathbf{F}_d$ and $\mathbf{F}_d^{-1}$ denote the one dimensional discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) matrices with respect to the d dimension, respectively, and $\mathbf{F}_{d_1,d_2}$ denotes a two dimensional discrete Fourier transform with respect to the d_1 and d_2 dimensions. $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ respectively denote complex conjugate, transpose and hermitian operations of a matrix. $\|\cdot\|_p$ represents the ℓ_p-norm of a vector, and $|\cdot|$ denotes the cardinality of a set. $p(\cdot)$ denotes the probability density function (PDF). $\mathcal{N}(\cdot)$, Beta$(\cdot)$, Gamma$(\cdot)$ and $\mathcal{B}(\cdot)$, denote Gaussian, Beta, Gamma and Bernoulli distributions, respectively.

2. SIGNAL MODEL

Let f_s and $\Delta t = 1/f_s$ respectively denote the sampling rate and sampling interval. Consider a discrete-time FH signal, corrupted by additive complex white Gaussian noise $v[n]$, expressed as [10, 13, 14]

$$s[n] = \sum_{k=1}^{K_h} A_{h,k} \mathrm{e}^{\jmath 2\pi f_{h,k} n \Delta t} + v[n], \quad n_{h,k-1} \leq n < n_{h,k}, \tag{1}$$

where $\jmath = \sqrt{-1}$, $n_{h,k}$ denotes the discrete instant of the k-th tone in the h-th system-wise dwell for $h = 1, 2, \ldots, H$, and $A_{h,k}$ and $f_{h,k}$, respectively, represent the complex amplitude and frequency corresponding to the time index $n_{h,k}$. The number of tones K_h can vary with h because of emitter (de)activation or bandwidth mismatch [10].

Consider a collection of N samples $\mathbf{x}[n] = [x[1], \ldots, x[n]]^{\mathrm{T}}$ as the observation data with $N - M$ missing samples, whose positions are assumed to be uniformly distributed. As such, $\mathbf{x}[n]$ can be regraded as the Hadamard product of original data vector $\mathbf{s}[n] = [s[1], \ldots, s[n]]^{\mathrm{T}}$ and a binary mask vector $\mathbf{b}[n] = [b[n], \ldots, b[n-N+1]]^{\mathrm{T}}$, i.e.,

$$\mathbf{x}[n] = \mathbf{s}[n] \circ \mathbf{b}[n], \tag{2}$$

where

$$b[n] := \begin{cases} 1, & \text{if } n \in \mathcal{I}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

$\mathcal{I} \subset \{1, 2, \ldots, N\}$ is the set of observed time instants with cardinality $|\mathcal{I}| = M$, and $1 - M/N$ represents the missing-sample ratio.

3. KERNEL DESIGN BASED ON FH SIGNAL STRUCTURES

In this section, a detailed description of the signal-dependent kernel design method is presented.

3.1 Joint-variable representations of FH spectrum

The discrete-time instantaneous autocorrelation function (IAF) of a signal $x[n]$ is defined as [16]

$$C_{xx}[\tau, n] = x[n+\tau]x^*[n-\tau], \tag{4}$$

where τ denotes the time lag index. Stacking $C_{xx}[\tau, n]$ corresponding to all values of τ and n results in an IAF matrix $\mathbf{C_{xx}}$. Then, the AF matrix of signal vector $\mathbf{x}[n]$, expressed with respect to lag τ and Doppler frequency κ, can be obtained by performing IDFT of the IAF with respect to the time index n, i.e.,

$$\mathbf{A_{xx}}\{\tau, \kappa\} = \mathbf{F}_n^{-1}\mathbf{C_{xx}}\{\tau, n\} = \sum_n \mathbf{C_{xx}}\{\tau, n\}\mathrm{e}^{\jmath 2\pi\kappa n}, \tag{5}$$

where the notation $\{\tau, \kappa\}$ is used to emphasize that the matrix $\mathbf{A_{xx}}$ is constructed with respect to variables τ and κ. The WVD can be obtained by performing DFT to its IAF with respect to the lag index τ, i.e.,

$$\mathbf{W_{xx}}\{f, n\} = \mathbf{F}_\tau\mathbf{C_{xx}}\{\tau, n\} = \sum_\tau \mathbf{C_{xx}}\{\tau, n\}\mathrm{e}^{-\jmath 4\pi f\tau}. \tag{6}$$

3.2 Data-dependent TF kernel

As described earlier, artifacts due to missing samples must be mitigated before reliable FH parameter estimations can be performed. This can be achieved by applying proper TF kernels, which are commonly used in bilinear TF analysis for cross-term reduction and signal auto-term enhancement [15]. These kernels can be classified as data-dependent and data-independent, with the former generally outperforming the latter due to their optimization with respect to the signal observed. The AOK is a well-known data-dependent kernel, which is designed by solving the following optimization problem [18]:

$$\begin{aligned} \Phi_{\mathrm{opt}}(r,\psi) = \underset{\Phi(r,\psi)}{\arg\max} & \int_0^{2\pi}\int_0^{\infty} |A(r,\psi)\Phi(r,\psi)|^2 r\mathrm{d}r\mathrm{d}\psi, \\ \text{s.t.} \quad & \begin{cases} \Phi(r,\psi) = \exp\left(-\dfrac{r^2}{2\sigma^2(\psi)}\right), \\ \dfrac{1}{4\pi^2}\displaystyle\int_0^{2\pi} \sigma^2(\psi)d\psi \le \alpha, \end{cases} \end{aligned} \tag{7}$$

where α denotes the kernel volume constraint, $A(r,\psi)$ represents the AF of the signal in polar coordinates, and r and ψ denote the radius and radial angle variables, respectively.

3.3 Post-kernel window

Despite its general filtering capability, conventional AOK fails to filter out the artifacts induced by missing samples that are close to the origin of the AF domain. In particular, missing samples generate vertical strips in the TF domain, which are concentrated in the Doppler axis of the AF domain. On the other hand, it is known *a priori* that the auto-terms of the FH signal are concentrated in the lag axis, regardless of their hop frequencies. For the multi-component situations, the auto-terms remain in the same position (while the exact value is distorted due to the superposition of multiple component), whereas the cross-terms appear away from the origin depending on the signal lags and the hop frequency difference. The length of the auto-term along the lag axis depends on the maximum duration of the hops, and a larger maximum hopping duration results in a larger lag-wise length. On the other hand, the width of the auto-term along the Doppler axis depends on the observed data length, and a longer observed data length results in a narrower Doppler-wise width. Therefore,

we can design a post-AOK kernel by thresholding the *a prior* information based conjectural single-component auto-term AF $\hat{A}_{\mathbf{xx}}\{\tau,\kappa\}$, expressed as

$$\tilde{w}[\tau,\kappa] = \begin{cases} |\hat{A}_{\mathbf{xx}}[\tau,\kappa]|, & \text{if } |\hat{A}_{\mathbf{xx}}\{\tau,\kappa\}| > \xi, \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

where ξ is the threshold, which is chosen to only keep FH auto-terms with substantial signal concentration in the AF domain. The designed window will keep the auto-term ambiguity region of the FH signals while filtering out other ambiguity regions so as to reduce or eliminate cross-terms and artifacts due missing samples.

Note that brick-wall filters are often refrained in conventional TF analysis as they may cause ripples in the TF domain. This is not a concern in the underlying application because signal auto-terms are negligible outside the window. In addition, the effect of windows is further mitigated in the sparse reconstruction process.

After the AOK and the post-AOK window are applied, the AF becomes:

$$\tilde{\mathbf{A}} = \mathbf{A}_{\mathbf{xx}}\{\tau,\kappa\} \circ \mathbf{\Phi}\{\tau,\kappa\} \circ \tilde{\mathbf{w}}\{\tau,\kappa\}. \tag{9}$$

4. BCS-BASED FH SPECTRUM ESTIMATION

4.1 CS model for FH spectrum estimation

In this section, we consider a CS based approach which yields a high-resolution TFR. The IAF matrix corresponding to the kernelled AF is obtained as the Fourier transform of $\tilde{\mathbf{A}}_{\mathbf{xx}}\{\tau,\kappa\}$ with respect to κ, i.e.,

$$\tilde{\mathbf{C}}_{\mathbf{xx}}\{\tau,n\} = \mathbf{F}_{\kappa}\tilde{\mathbf{A}}_{\mathbf{xx}}\{\tau,\kappa\} = \mathbf{F}_{\kappa}\left(\mathbf{A}_{\mathbf{xx}}\{\tau,\kappa\} \circ \mathbf{\Phi}\{\tau,\kappa\} \circ \tilde{\mathbf{w}}\{\tau,\kappa\}\right). \tag{10}$$

Then, the bilinear TFR matrix is associated with the IAF matrix by the following Fourier relationship:

$$\tilde{\mathbf{C}}_{\mathbf{xx}}\{\tau,n\} = \mathbf{F}_f^{-1}\tilde{\mathbf{W}}_{\mathbf{xx}}\{f,n\}. \tag{11}$$

The CS approach obtains $\tilde{\mathbf{W}}_{\mathbf{xx}}\{f,n\}$ by exploiting the above Fourier transform relationship but through a sparse reconstruction operation. Denote the n-th column of the IAF matrix $\tilde{\mathbf{C}}_{\mathbf{xx}}\{\tau,n\}$ as $\tilde{\mathbf{c}}_{\mathbf{xx}}[n]$, and the n-th column of the bilinear TFR matrix $\tilde{\mathbf{W}}_{\mathbf{xx}}\{f,n\}$ as $\tilde{\mathbf{w}}_{\mathbf{xx}}[n]$. Then, their relationship conforms to the following standard linear model commonly used in CS and sparse reconstruction:

$$\tilde{\mathbf{c}}_{\mathbf{xx}}[n] = \mathbf{F}_f^{-1}\tilde{\mathbf{w}}_{\mathbf{xx}}[n]. \tag{12}$$

Therefore, the TFR can be obtained from sparse reconstruction, in lieu of conventional Fourier transform, by repeating the procedure for each time instant column. Various CS algorithms can be used for this purpose. In the following, we consider this problem from a BCS perspective, and the structure of the FH spectrum is utilized for improved spectrum estimation. For notational convenience, we simplify the notations $\tilde{\mathbf{c}}_{\mathbf{xx}}[n]$, $\mathbf{F}_f^{-1}$ and $\tilde{\mathbf{w}}_{\mathbf{xx}}[n]$ as $\mathbf{c}$, $\mathbf{F}$ and $\mathbf{w}$, respectively.

4.2 Sparsity prior

The BCS is a nonparametric sparse linear inverse solver imposing a conditional Gaussian prior with its precision guided by a Gamma distributed hyperprior. The BCS assumes the likelihood model as [26]

$$p\left(\mathbf{c};\mathbf{w},\gamma_0\right) = \mathcal{N}(\mathbf{c};\mathbf{F}_f^{-1}\mathbf{w},\gamma_0\mathbf{I}), \tag{13}$$

where $\gamma_0 = \alpha_0^{-1}$ is the variance, and α_0 follows the Gamma distribution, i.e., $\alpha_0 \sim \text{Gamma}(c,d)$. To encourage sparsity of the FH signal TFR, a spike-and-slab prior model [27,28] is employed to w_i, i.e., the i-th entry of $\mathbf{w}$, as,

$$p(w_i;\gamma_i,\pi_i) = (1-\pi_i)\delta_0 + \pi_i\mathcal{N}(w_i;0,\gamma_i), \tag{14}$$

where π_i is a mixing weight standing for the prior probability of a nonzero entry, δ_0 represents the delta function with a unit point measure concentrated at zero, and we also assign a Gamma prior to the inverse variance as $\gamma_i^{-1} = \alpha_i \sim \text{Gamma}(a,b)$.

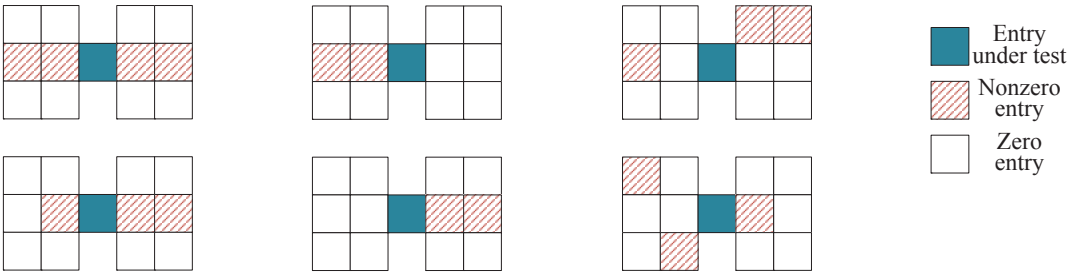

Figure 1. Example patterns of three different categories for TFR structure patterns for $D = 2$: (a) Category 1: acceptance. (b) Category 2: neutral. (c) Category 3: rejection.

To make the inference analytical, a product of two latent variables z and θ, i.e., $w_i = z_i \cdot \theta_i$ is introduced to follow the PDF in Eq. (14), where $\theta_i \sim \mathcal{N}(\theta_i; 0, \gamma_i)$. In addition, z_i is a binary variable with a Bernoulli distribution $\mathcal{B}(\pi_i)$. $z_i = 1$ implies that the i-th entry is nonzero, whereas $z_i = 0$ implies a zero entry. The overall prior on $\boldsymbol{\theta}$ with respect to a and b can be evaluated analytically through the integration over $\boldsymbol{\alpha}$, and it corresponds to the Student-t distribution [29]. With an appropriate choice of a and b, the Student-t distribution is strongly peaked about $\boldsymbol{\theta} = 0$, and thus the overall prior on $\boldsymbol{\theta}$ favors sparseness [27].

4.3 Structure prior

The FH spectrum shows a piecewise constant frequency spectrum. Fig. 1 shows several TFR structure patterns for a neighborhood within an Euclidean distance of 2, where the horizonal direction shows the discretized time entries, whereas the vertical direction represents the frequency entries. The possible TFR structure patterns are divided into three categories, and priors can be designed to encourage the FH spectrum to have a horizontally linear structure. On the other hand, entries that are on adjacent rows are discouraged because they tend to broaden the signal bandwidth, thus is contradictory to the fact that the underlying FH signals are instantaneously narrowband. Fig. 1(a) includes the cases where the entry under test is more likely to be nonzero since it has a high number of horizontal neighboring nonzero pixels, i.e., it is likely to belong to the spectrum of an FH component. While for Fig. 1(b), we take a neutral position because it cannot be determined whether the entry under test tends to be nonzero or not. For the third case depicted in Fig. 1(c), we assert a rejection so as to prevent the broadening of nonzero entries in the frequency domain.

Mathematically, we define the neighborhood of entry index i within the same row as

$$\mathcal{J}_{\odot i} = \{j \,|d\,(i,j) \leq D, j \in [1, N]\}\,, \tag{15}$$

where $d\,(i,j)$ is the Euclidean distance between i and j, and D is a natural number that represents the maximum distance to be included in the evaluation. We then define the deleted neighborhood of index i as

$$\mathcal{J}_{\otimes i} = \{j \,|d\,(i,j) \leq D, j \in [1, N]\ , j \neq i\}\,. \tag{16}$$

Comparing Eqs. (15) and (16), we can see that the relationship between neighborhood and deleted neighborhood is $\mathcal{J}_{\odot i} = \mathcal{J}_{\otimes i} \cup \{i\}$. The number of nonzero entries at a location i and its neighborhood is denoted as $z_{\mathcal{J}_{\odot i}}$. In this paper, we consider $D = 2$ case.

Note that during the pattern classification process, three rows of $\tilde{\mathbf{c}}_{\mathbf{xx}}$ are investigated. The i-th entry in the $[n+1]$ and $[n-1]$-th row of $\tilde{\mathbf{c}}_{\mathbf{xx}}$ is denoted as $i+$ and $i-$, respectively.

Since the Beta distribution is conjugate to Bernoulli likelihood, π_i is assumed to follow the Beta distribution. For a certain structure pattern q, where $q \in \{1, 2, 3\}$ is one of the three categories, the posterior distribution of $\pi_i^{<q>}$ is derived as

$$p(\pi_i^{<q>}; e^{<q>}, f^{<q>}) = \text{Beta}\left(e^{<q>} + z_{\mathcal{J}_{\odot i}}, f^{<q>} + |\mathcal{J}_{\odot i}| - z_{\mathcal{J}_{\odot i}}\right). \tag{17}$$

The Beta(e, f) distribution tends to draw a small sample when $e < f$ and a big sample when $e > f$, while it has no tendency when $e = f$. Therefore, by choosing proper values of e and f, we can encourage or discourage the sparsity of the pixel under test, depending on the sparsity support in the neighboring pixels.

Category 1 (acceptance): When 3 or more nonzero pixels exist in the same row, i.e., $z_{J_{\otimes i}} \geq 3$, and no nonzero pixels exist in the neighboring rows, the pixel under test is encouraged to take a nonzero support by setting $e^{<1>} > f^{<1>}$;

Category 2 (neutral): When 2 nonzero pixels exists in the same row, $z_{J_{\otimes i}} = 2$, and no nonzero pixels in the neighboring rows, we choose $e^{<2>} = f^{<2>}$ to take a neutral position;

Category 3 (rejection): For other situations, we choose $\rightarrow e^{<3>} < f^{<3>}$ to discourage the pixel under test to have a nonzero support.

4.4 Bayesian inference

The BCS algorithm assumes that each element of $\mathbf{w}$ depicted in Eq. (14) is independently distributed. The correlation between neighboring pixels is treated through the proper design of priors as described above. Therefore, the joint PDF of $\mathbf{w}$ becomes:

$$p\left(\mathbf{w};\boldsymbol{\gamma},\boldsymbol{\pi}\right) = \prod_{i=1}^{N}\left[(1-\pi_i)\delta_0 + \pi_i \mathcal{N}\left(w_i;0,\gamma_i\right)\right], \tag{18}$$

where $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_N]^{\mathrm{T}}$ is a vector of N hyperparameters that controls the prior variance of each weight. We can then acquire the posterior distribution of $\mathbf{w}$ by combing (18) with the observation likelihood $p\left(\mathbf{c};\mathbf{w},\gamma_0\right)$ in (13), i.e.,

$$p\left(\mathbf{w}\left|\boldsymbol{\gamma},\boldsymbol{\pi},\mathbf{c},\alpha_0\right.\right) \propto p\left(\mathbf{w};\boldsymbol{\gamma},\boldsymbol{\pi}\right) p\left(\mathbf{c};\mathbf{w},\alpha_0\right) \tag{19}$$

A Gibbs sampler is adopted to implement the Bayesian inference as following. Denote $\boldsymbol{\Lambda}$ as the measurement matrix of the CS model described in Eq. (12), i.e., $\boldsymbol{\Lambda} = \mathbf{F}_f^{-1}$, and let $\boldsymbol{\lambda}_i$ be the i-th column of $\boldsymbol{\Lambda}$. Then, the paired Gibbs sampler iteratively samples from the following conditional PDF [27, 28]

$$p\left(w_i\left|\mathbf{w}_{\backslash i},\mathbf{c}\right.\right) = p\left(\theta_i, z_i\left|\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right) = p\left(\theta_i\left|z_i,\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right) p\left(z_i\left|\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right), \tag{20}$$

where the notation $(\cdot)_{\backslash i}$ denotes the subvector excluding the i-th entry. The probability $p\left(z_i = 1\left|\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right)$ is acquired as

$$p\left(z_i = 1\left|\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right) = \frac{\alpha_i}{1-\alpha_i}\frac{\mathcal{N}\left(0,\gamma_i\right)}{\mathcal{N}\left(\tilde{\mu}_i,\gamma_i\right)}, \tag{21}$$

where μ_i and γ_i are updated as

$$\tilde{\mu}_i = \alpha_i^{-1}\alpha_0 \boldsymbol{\lambda}_i^H \mathbf{c}_{\backslash i}, \tag{22}$$

$$\tilde{\gamma}_i^{-1} = \tilde{\alpha}_i = \left(\alpha_0 \boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i + \alpha_i\right). \tag{23}$$

The conditional distribution of $p\left(\theta_i\left|z_i = 1,\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right)$ can be expressed as

$$p\left(\theta_i\left|z_i = 1,\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right) = \mathcal{N}\left(w_i;\tilde{\mu}_i,\gamma_i\right). \tag{24}$$

For $z_i = 0$ case, as the value of θ_i does not affect the result of w_i, we directly draw the value of θ_i from its prior. Subsequently, the Gibbs sampler updates the mixing weight π_i according to (17).

Next, we update the precision variable α_i. By utilizing the conjugate property of the Gaussian and Gamma distributions, we analytically acquire the posterior distribution of α_i as

$$p\left(\alpha_i;a,b,z_{\mathcal{J}_{\odot i}}\right) = \mathrm{Gamma}\left(a + \frac{z_{\mathcal{J}_{\odot i}}}{2}, b + \frac{\left\|z_{\mathcal{J}_{\odot i}}\right\|_2^2}{2}\right). \tag{25}$$

After all the i iteration loops are completed, the posterior distribution of noise precision α_0 is updated as

$$p\left(\alpha_0; c, d, \mathbf{c}, \mathbf{\Lambda}, \theta, \mathbf{z}\right) = \text{Gamma}\left(c+\frac{\text{rank}\{\mathbf{\Lambda}\}}{2}, d+\frac{\|\mathbf{c}-\mathbf{\Lambda}\left(\theta \circ \mathbf{z}\right)\|_2^2}{2}\right). \tag{26}$$

The maximum a posteriori (MAP) estimator is adopted to infer the estimation for $\mathbf{w}$ as

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} p\left(\mathbf{w}\,|\mathbf{c}\right), \tag{27}$$

where marginal distribution $p\left(\mathbf{w}\,|\mathbf{c}\right)$ can be obtained by integrating out the hyperparameters $\boldsymbol{\gamma}$, $\boldsymbol{\pi}$, and α_0 in (19) as

$$p\left(\mathbf{w}\,|\mathbf{c}\right) \propto \int p\left(\mathbf{w}\,|\boldsymbol{\gamma}, \boldsymbol{\pi}, \mathbf{c}, \alpha_0\right) \mathrm{d}\boldsymbol{\gamma}\mathrm{d}\boldsymbol{\pi}\mathrm{d}\alpha_0 \tag{28}$$

Hereby we obtain the sparse reconstruction result of Eq. (12). The estimation of the entire FH spectrum is rendered by repeating the BCS-based estimation for each columns of $\tilde{\mathbf{W}}_{\mathbf{xx}}\{f, n\}$.

5. SIMULATION AND ANALYSIS

Simulation results are provided to demonstrate the effectiveness of the proposed approach. The parameter settings are listed in Table 1. The model hyperparameters for the priors are set as follows: $a = b = c = d = 10^{-6}$, $\left(e^{<0>}, f^{<0>}\right) = (1/N, 1-1/N) \times |\mathcal{J}_{\odot i}|$, $\left(e^{<1>}, f^{<1>}\right) = (1/N, 1/N) \times |\mathcal{J}_{\odot i}|$, $\left(e^{<2>}, f^{<2>}\right) = (1-1/N, 1/N) \times |\mathcal{J}_{\odot i}|$, the initial conditions are set to $\alpha_i(0) = 1, \pi_i(0) = 0$, and $\alpha_0(0) = 10^2/\text{var}(\mathbf{c})$.

Fig. 2 shows the simulation results. From Fig. 2(a) we can see two hopping components in the FH signal: The first component is active within the range of $[0 : 0.25]$ μs and the carrier frequency hops from 27.75 MHz to 13.50 MHz within the range of $[0.27 : 1.05]$ μs, whereas the second component hops from 2.25 MHz to 20.40 MHz at the 31-st sample point, i.e., at the time instant 0.52 μs. Cross-terms can be observed in the WVD and AF planes depicted in Figs. 2(b) and 2(c), and the impact of missing samples are clearly exhibited in the IAF plane depicted in Fig. 2(d).

Figs. 2(e) and 2(f) show the AF and TFR results using AOK, where artifacts induced by missing samples are observed along the Doppler axis in the AF domain and as vertical strips in the TFR. A vertical post-AOK window is then multiplied in the AF domain to filter out the artifacts, and the corresponding AF and TFR results are depicted in Figs. 2(g) and 2(h). It is clear that the FH auto-terms are preserved whereas the vertical strips are suppressed in the TFR. Fig. 2(i) shows the BCS estimation of the FH spectrum. This represents a significant improvement as compared to the WVD depicted in Fig. 2(b).

Table 1. Simulation Parameters

f_s (MHz)	N	M	H	SNR (dB)
60	64	32	2	10

6. CONCLUSION

In this paper, a novel structure-aware FH spectrum estimation approach with the consideration of missing samples was proposed within a sparse reconstruction framework. The kerneled joint-variable representation over time and lag was used to provide the TF signal representation through sparse reconstruction. In the sparsity-based spectrum estimation process, the structure of the entry under test and its neighborhood is exploited to impose a structure-aware prior on the Bayesian inference. Simulation results clearly show that the proposed approach outperforms existing approaches devised for the same problem.

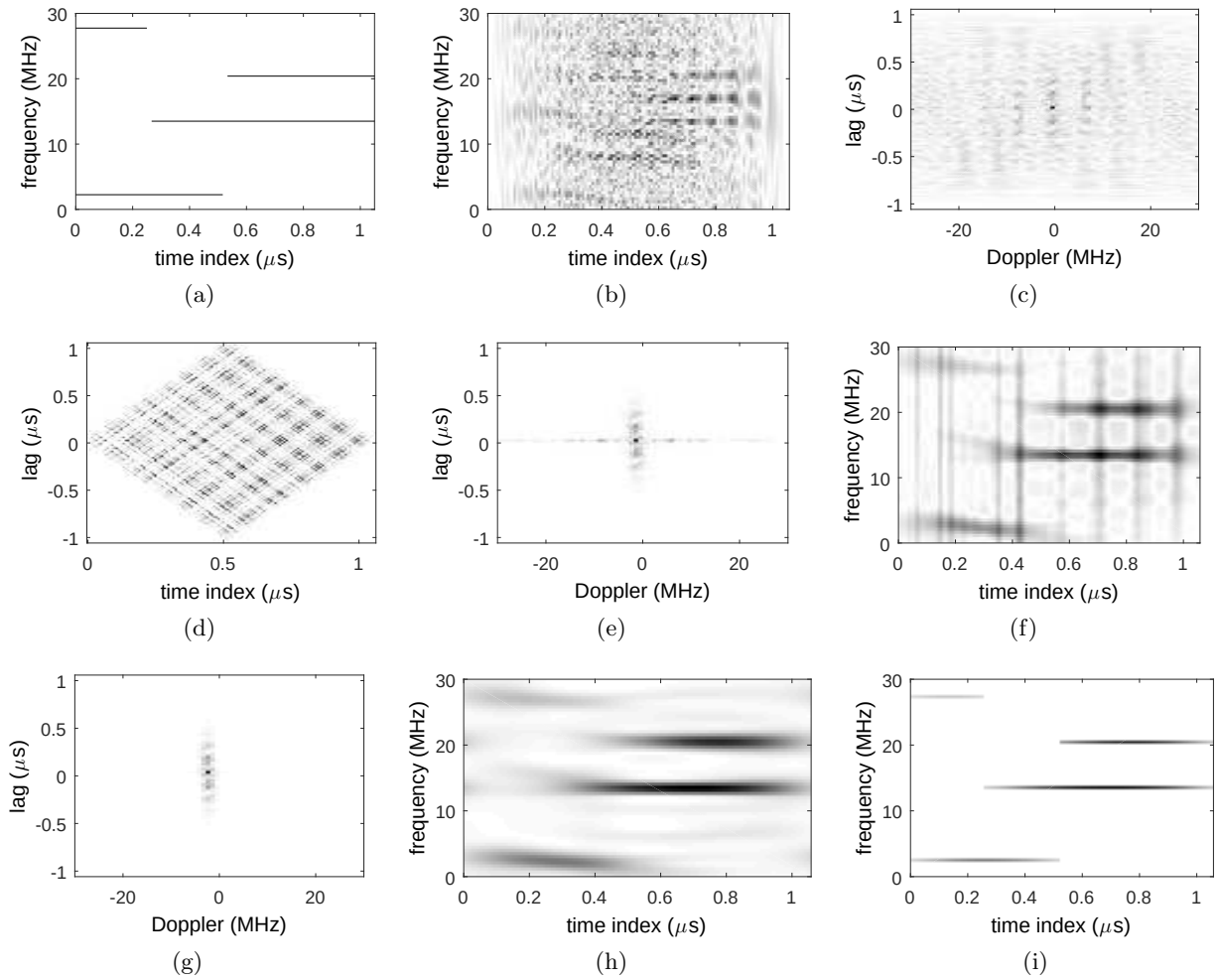

Figure 2. Simulation results for the proposed structure-aware BCS based FH spectrum estimation algorithm: (a) Truth; (b) WVD; (c) AF; (d) IAF; (e) Kernelled AF with AOK; (f) Kernelled TFR with AOK; (g) AF of filtered-AOK processed signal; (h) TFR of filtered-AOK processed signal; (i) Structure-aware BCS result of filtered-AOK processed signal.

ACKNOWLEDGMENTS

The work of S.-H. Liu and T. Shan is supported in part by the National Natural Science Foundation of China under Grants No. 61421001 and 61331021. The work of Y. D. Zhang, S. Qin, and M. G. Amin is supported in part by the National Science Foundation under Grant No. AST-1547420. S.-H. Liu acknowledges the financial support from the China Scholarship Council for his stay at Temple University.

REFERENCES

[1] Maric, S. V. and Titlebaum, E. L., "A class of frequency hop codes with nearly ideal characteristics for use in multiple-access spread-spectrum communications and radar and sonar systems," *IEEE Trans. Commun.* **40**(9), 1442–1447 (1992).

[2] Chen, C.-Y. and Vaidyanathan, P. P., "MIMO radar ambiguity properties and optimization using frequency-hopping waveforms," *IEEE Trans. Signal Process.* **56**(12), 5926–5936 (2008).

[3] Zhang, Y. and Amin, M. G., "MIMO radar exploiting narrowband frequency-hopping waveforms," *Proc. European Signal Process. Conf. (EUSIPCO)*, 1–5 (2008).

[4] Hunt, A. R., "Use of a frequency-hopping radar for imaging and motion detection through walls," *IEEE Trans. Geosci. Remote Sens.* **47**(5), 1402–1408 (2009).

[5] Torrieri, D., [*Principles of Spread-Spectrum Communication Systems: 3rd Edition*], Springer (2013).
[6] Lourey, S. J., "Frequency hopping waveforms for continuous active sonar," *Proc. 2015 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 1832–1835 (2015).
[7] Schaefer, F. M. and Kays, R., "Frequency hopping for indoor fading channels with varying level of environmental mobility," *IEEE Wireless Commun. Lett.* **4**(1), 42–45 (2015).
[8] Fischman, M., Chan, S., Huang, N., and Pak, K., "Frequency-agile radar electronics for the soil moisture active/passive (SMAP) mission," *IEEE Geosci. Remote Sens. Mag.* **3**(1), 10–19 (2015).
[9] Bae, S., Kim, S., and Kim, J., "Efficient frequency-hopping synchronization for satellite communications using dehop-rehop transponders," *IEEE Trans. Aerosp. Electron. Syst.* **52**(1), 261–274 (2016).
[10] Angelosante, D., Giannakis, G. B., and Sidiropoulos, N. D., "Estimating multiple frequency-hopping signal parameters via sparse linear regression," *IEEE Trans. Signal Process.* **58**(10), 5044–5056 (2010).
[11] Donoho, D., "Compressed sensing," *IEEE Trans. Inform. Theory* **52**(4), 1289–1306 (2006).
[12] Candès, E., Romberg, J., and Tao, T., "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pur. Appl. Math.* **59**(8), 1207–1223 (2006).
[13] Angelosante, D., Giannakis, G., and Sidiropoulos, N., "Sparse parametric models for robust nonstationary signal analysis: Leveraging the power of sparse regression," *IEEE Signal Process. Mag.* **30**(6), 64–73 (2013).
[14] Zhao, L., Wang, L., Bi, G., Zhang, L., and Zhang, H., "Robust frequency-hopping spectrum estimation based on sparse Bayesian method," *IEEE Trans. Wireless Commun.* **14**(2), 781–793 (2015).
[15] Cohen, L., [*Time-Frequency Analysis*], Prentice Hall (1995).
[16] Boashash, B., [*Time-Frequency Signal Analysis and processing: A Comprehensive Reference*], Academic Press (2015).
[17] Baraniuk, R. G. and Jones, D. L., "Signal-dependent time-frequency analysis using a radially Gaussian kernel," *Signal processing* **32**(3), 263–284 (1993).
[18] Jones, D. L. and Baraniuk, R. G., "An adaptive optimal-kernel time-frequency representation," *IEEE Trans. Signal Process.* **43**(10), 2361–2371 (1995).
[19] Amin, M. G., [*Compressive Sensing for Urban Radars*], CRC Press (2014).
[20] Zhang, Y. D., Amin, M. G., and Himed, B., "Reduced interference time-frequency representations and sparse reconstruction of undersampled data," *Proc. European Signal Process. Conf. (EUSIPCO)*, 1–5 (2013).
[21] Jokanovic, B., Amin, M. G., Zhang, Y. D., and Ahmad, F., "Time-frequency kernel design for sparse joint-variable signal representations," *Proc. Proc. European Signal Process. Conf. (EUSIPCO)*, 1–5 (2014).
[22] Stankovic, L., Stankovic, S., Orovic, I., and Zhang, Y. D., "Time-frequency analysis of micro-Doppler signals based on compressive sensing," in M. Amin (ed.), [*Compressive Sensing for Urban Radars*], CRC Press (2014).
[23] Amin, M. G., Jakonovic, B., Zhang, Y. D., and Ahmad, F., "A sparsity-perspective to quadratic time-frequency distributions," *Digital Signal Process.* **46**, 175–190 (2015).
[24] Liu, S.-H., Zhang, Y. D., and Shan, T., "Sparsity-based frequency-hopping spectrum estimation with missing samples," *Proc. 2016 IEEE Radar Conf.* (2016).
[25] Ji, S., Xue, Y., and Carin, L., "Bayesian compressive sensing," *IEEE Trans. Signal Process.* **56**(6), 2346–2356 (2008).
[26] Wipf, D. P. and Rao, B. D., "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.* **52**(8), 2153–2164 (2004).
[27] Yu, L., Barbot, J. P., Zheng, G., and Sun, H., "Compressive sensing for cluster structured sparse signals: Variational Bayes approach." Technical Report, 2011 http://hal.archives-ouvertes.fr/docs/00/57/39/53/PDF/cluss_vb.pdf. (Accessed: 5 April 2016).
[28] Wu, Q., Zhang, Y. D., and Amin, M. G., "Continuous structure based Bayesian compressive sensing for sparse reconstruction of time-frequency distributions," *Proc. Int. Conf. Digital Signal Process.*, 831–836 (2014).
[29] Tipping, M. E., "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.* **1**, 211–244 (2001).