Structure-Aware Bayesian Compressive Sensing for Frequency-Hopping Spectrum Estimation with Missing Observations

Shengheng Liu, *Member, IEEE*, Yimin D. Zhang, *Senior Member, IEEE*, Tao Shan, *Member, IEEE*, and Ran Tao, *Senior Member, IEEE*

***Abstract*—In this paper, we address the problem of spectrum estimation of multiple frequency-hopping (FH) signals in the presence of random missing observations. The signals are analyzed within the bilinear time-frequency (TF) representation framework, where a TF kernel is designed by exploiting the inherent FH signal structures. The designed kernel permits effective suppression of cross-terms and artifacts due to missing observations while preserving the FH signal auto-terms. The kernelled results are represented in the instantaneous autocorrelation function domain, which are then processed using a re-designed structure-aware Bayesian compressive sensing algorithm to accurately estimate the FH signal TF spectrum. The proposed method achieves high-resolution FH signal spectrum estimation even when a large portion of data observations is missing. Simulation results verify the effectiveness of the proposed method and its superiority over existing techniques.**

***Index Terms*—Frequency hopping, spectrum estimation, missing observations, Bayesian compressive sensing, time-frequency distribution, kernel design.**

I. 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/from other users, resistance to jamming and multipath fading, and desirable ambiguity function property, FH signals have become a favorable choice in a wide range of communication and radar applications, particularly in the context of multiple-input multiple-output (MIMO) operations [3]–[7]. 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 challenging task when the hopping patterns of the constituent signals are unavailable. The problem becomes even more difficult when the hopping period is time-varying [9].

In this paper, we consider the spectrum estimation of multi-emitter FH signals with unknown and time-varying hopping periods in the context of Bayesian compressive sensing (BCS). In particular, we focus on the case where the received signal waveform is subject to missing observations. The specific FH signal structures are utilized to design time-frequency (TF) kernels and BCS structure priors to achieve reliable and high-resolution FH spectrum estimation.

The continuous-time noisy multi-emitter FH signal considered in this paper is expressed as [8]

$$s(t) = \sum_{h=1}^{H} \sum_{k=1}^{K_h} A_{h,k} \Pi_{T_h}(t - kT_h) \mathrm{e}^{\jmath 2\pi f_{h,k}(t-kT_h)} + v(t), \tag{1}$$

where $\jmath = \sqrt{-1}$, and $\Pi_{T_h}(t)$ represents a normalized boxcar function, which is equal to one for $t \in \left(-\frac{T_h}{2}, \frac{T_h}{2}\right]$ and 0 otherwise. In addition, T_h denotes the duration of each hop of the h-th individual FH emitter, and H is the number of FH emitters. Moreover, $A_{h,k}$ and $f_{h,k}$ are the complex amplitude and carrier frequency of the k-th tone in the h-th system-wise dwell, respectively. The number of tones, K_h, may vary with h because of emitter (de)activation or bandwidth mismatch [9]. $v(t)$ represents the additive circularly-symmetric complex white Gaussian noise. Let f_s and $\Delta t = 1/f_s$ respectively denote the sampling rate and the sampling interval. Then, the sampled discrete-time FH signal can be derived from (1) as

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

In practice, the measured data may experience missing samples due to channel distortion/fading, line-of-sight obstruction, removal of samples contaminated by impulsive noise, and collecting/storage equipment failures [10]. Denote $x[n]$ as signal $s[n]$ with missing data, and $N_m \subset \{1, 2, \ldots, N\}$ as the set of missing time instants with cardinality $|N_m| = M$,

The work of S.-H. Liu, T. Shan, and R. Tao was supported in part by the National Natural Science Foundation of China under Grants Nos. 61671060, 61421001, 61331021, and Natural Science Foundation of Beijing Municipality under Grant No. 4172052. The work of Y. D. Zhang was supported in part by the National Science Foundation under Grant No. AST-1547420. S.-H. Liu gratefully acknowledges the financial support from the China Scholarship Council for his stay at Temple University. Part of this work was presented at the 2016 IEEE Radar Conference, Philadelphia, PA, USA [1] and the 2016 SPIE Compressive Sensing Conference, Baltimore, MD, USA [2].

S.-H. Liu was with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China, and also with the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA (e-mail: henry@bit.edu.cn).

Y. D. Zhang is with the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA (e-mail: ydzhang@temple.edu).

T. Shan and R. Tao are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mails: shantao@bit.edu.cn, rantao@bit.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2018.2806351

where N is the total length of signal $x[n]$ and $s[n]$, and M/N represents the missing-sample ratio. Then, $x[n]$ can be interpreted as $s[n]$ modulated by a sum of Dirac delta functions (impulses) [11], i.e.,

$$x[n] = s[n]\left(1 - \sum_{n_m \in N_m} \delta\left[n - n_m\right]\right). \tag{3}$$

Random missing observations induce noise-like artifacts in the time-frequency distributions (TFD) [17], which makes the problem even more intractable.

A. Related Work

Time-varying spectrum signatures of non-stationary signals, such as FH signals, can be revealed in the joint TF domain representations. As FH signals generally exhibit sparsity in the joint TF domain, compressive sensing (CS) and sparse reconstruction techniques [12]–[14] enable effective FH spectrum representation and parameter estimation. In [9], [15], this problem is solved by formulating the problem as an underdetermined linear regression with a dual sparsity penalty, i.e., a penalty function that controls both the intrinsic sparsity and smoothness of the estimation. However, this approach requires appropriate tuning of the parameters, and obtaining the optimum solution still requires considerable effort. Another limitation of the approach proposed in [9], [15] is that they do not provide robust estimation performance due to the sensitivity of the differential operator used in fused least absolute shrinkage and selection operator (LASSO). To improve the parameter estimation performance, particularly in low signal-to-noise ratio (SNR) conditions, a BCS method was adopted in [16], where a logistic stick breaking process is employed to encourage the temporal clustering over each hopping interval. BCS algorithm enables, through the proper design of priors, the incorporation of the contiguity property of typical TF signatures and thus enhances sparse optimization solutions. However, all the aforementioned approaches are based on linear TF analyses, and do not account for the effect of missing observations. Actually, linear approaches fail in the case of missing samples, as we will show in Section IV.

As described in [10], [11], [17]–[19], the effect of artifacts induced by random missing samples can be substantially reduced by applying proper TF kernels, which involves developing FH spectrum estimation methods in the bilinear time-frequency representation (TFR) framework. Sparse reconstruction of TFRs using different CS methods can also be found therein. It is known that bilinear (quadratic) TFDs provide high-resolution time-varying spectrum representations. The Wigner-Ville distribution (WVD) is considered as a prototype of bilinear TFDs, which offers highest TF energy concentration for single-component linear frequency modulated signals. However, because of the bilinear nature, it causes cross-terms between different components that constitute false energy distributions. To resolve this problem, various reduced-interference distributions have been developed for cross-term reduction through the design of appropriate TF kernels in the general Cohen's class [20], [21]. Such TF kernels can be signal-independent or signal-dependent. The latter performs parameter tuning via optimization, and thus generally provides better performance in trading off the cross-term suppression and the auto-term preservation. In particular, the adaptive optimal kernel (AOK), which is based on the optimization of radial Gaussian functions in the ambiguity function (AF) domain [22], [23], is a commonly used signal-dependent kernel.

Recently, such approaches have been adopted to estimate FH spectrum from data with missing observations, and an orthogonal matching pursuit (OMP) algorithm based approach was developed to achieve both artifact mitigation and high-resolution FH signal spectrum estimation [1]. The filtering capability of TF kernels offers bilinear TFR unique advantages over its linear counterpart [9], [15], [16] to effectively suppress the artifacts induced by missing observations. In the underlying problem that deals with FH spectrum estimation, however, separately reconstructing the TFR in each time instant as in [1] does not utilize an important signal characteristic relevant to the contiguous structure of the FH signatures. In particular, the approach may likely generate isolated or sporadic entries in the reconstructed TFR in the presence of missing data and/or measurement noise. In [24], a novel continuous structure based BCS approach [25] is proposed for the sparse reconstruction of nonstationary signals with missing observations. On this basis, a re-designed BCS-based scheme that exploits the contiguous structure of the FH signal is applied in [2], [26] to provide additional robustness in the FH spectrum estimation. Compared with the FH spectrum estimation via OMP [1], the BCS-based approach is proved capable to achieve an improved sparse solution [25]. The BCS methods approach sparse solutions that are close to ℓ_0-norm optimization and support structure-aware sparse problems through the use of adequate priors.

B. Contributions

The main novelty of this paper lies in the development of a *Comprehensive Structure-Aware Spectrum Estimation* technique for FH signals which is more advantageous than existing techniques. In particular, it is the first time to investigate the spectrum estimation for FH signals *in the presence of missing observations*. The concept of structure-awareness contains two major components, namely, structure-aware TF kernel design and structure-aware TFR reconstruction. (a) A structure-aware TF kernel is first developed in the AF domain to perform effective suppression of cross-terms and artifacts due to missing observations while preserving the FH signal auto-terms. In particular, we propose a new waveform-adaptive TF kernel design which combines an automatically optimized pre-filtering window and the data-dependent AOK. The pre-filtering window function exploits the prior information of the FH waveform characteristics, whereas the AOK further optimizes the kernel for effective cross-term and artifact reduction while preserving the signal auto-terms. The kernelled AF is then transformed to the instantaneous autocorrelation function (IAF) domain through a one-dimensional (1-D) Fourier transform with respect to (w.r.t.) the frequency difference (Doppler) domain. The IAF results are then processed using sparse reconstruction methods for high-resolution reconstruction of the FH signal

TF spectrum. (b) In the sparse reconstruction process, a redesigned BCS approach is developed to estimate the TFD of the signal from the IAF. A novel structure prior for the TFD is imposed to enforce the unique horizontal continuity of the TFR, that characterize the underlying FH signals. Compared with [24], [27], in addition to designing new structure-aware patterns, we also propose nonlinear updating rules associating the hyper-parameters with the TF patterns, rather than simply select the hyper-parameters from fixed categories. As such, the proposed approach can robustly estimate the FH spectrum in the presence of a high number of missing samples and when the *a priori* information of the hopping patterns is unavailable. As we will show in Section IV, while existing methods coping with the FH parameter estimation problem developed in [9], [15], [16] deteriorate sharply when treating data with missing observations, the proposed approach achieves superior performance in such challenging scenarios.

Notations: Lower-case (upper-case) bold characters are used to denote vectors (matrices). $\mathrm{abs}(\cdot)$ returns the modulus of a given complex number. $\circ$ denotes Hadamard product. $\mathrm{diag}\{\cdot\}$ 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 1-D discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) matrices w.r.t. the d dimension, respectively, and $\mathbf{F}_{d_1,d_2}$ denotes a two-dimensional (2-D) DFT w.r.t. 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). $\mathcal{B}(\cdot)$, $\mathcal{CN}(\cdot)$, $\mathrm{Beta}(\cdot)$, and $\mathrm{Gamma}(\cdot)$ denote Bernoulli, complex Gaussian, Beta, and Gamma distributions, respectively.

II. Structure-Aware Adaptive Kernel Design

The main stages of the proposed structure-aware scheme are summarized in the flowchart depicted in Fig. 1. In this section, we first present a detailed description of the proposed signal-dependent kernel design. A joint-variable representation of the FH spectrum in the presence of missing samples is first described in Section II-A, and the adaptive kernel design is introduced in Section II-B. Section II-D discusses the optimization of the pre-filtering parameters.

A. Joint-Variable Representations of Missing-Sample FH Spectrum

The discrete-time IAF of signal $x[n]$ is defined as [21]

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

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 w.r.t. lag τ and Doppler frequency κ, can be obtained by performing 1-D IDFT on the IAF w.r.t. 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 w.r.t. variables τ and κ. Similarly, the WVD can be obtained by performing 1-D DFT on the IAF w.r.t. the lag index τ, i.e.,

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

Remarks: Note that we use $-\jmath 4\pi f\tau$ in the above expression to perform the DFT because integer lags are adopted. This is a common practice in computing the discrete WVD.

Substituting (3) and (4) into (5), the AF of the observed FH signal with missing samples can be obtained as follows

$$\begin{aligned}
A_{xx}[\tau, \kappa] &= \sum_n C_{ss}[\tau, n]\left(1 - \sum_{n_m \in N_m} \delta\left[n - n_m + \tau\right]\right) \\
&\quad \cdot \left(1 - \sum_{n_m \in N_m} \delta\left[n - n_m - \tau\right]\right)\mathrm{e}^{\jmath 2\pi\kappa n} \\
&= A_{ss}[\tau, \kappa] - \sum_{n_m \in N_m} s[n_m]s^*[n_m - 2\tau] \cdot \mathrm{e}^{\jmath 2\pi\kappa(n_m - \tau)} \\
&\quad - \sum_{n_m \in N_m} s[n_m + 2\tau]s^*[n_m] \cdot \mathrm{e}^{\jmath 2\pi\kappa(n_m + \tau)} \\
&\quad + \sum_n \sum_{n_m \in N_m} \Bigg(\left(\delta\left[n - n_m + \tau\right] s[n+\tau]\right. \\
&\quad \left. \cdot \sum_{n_l \neq n_m \in N_m} \delta\left[n - n_l - \tau\right] s^*[n-\tau]\right)\mathrm{e}^{\jmath 2\pi\kappa n} \\
&\quad + \sum_{n_m \in N_m} \delta\left[-2\tau\right]s[n_m]s^*[n_m - 2\tau]\mathrm{e}^{\jmath 2\pi\kappa(n_m - \tau)},
\end{aligned} \tag{7}$$

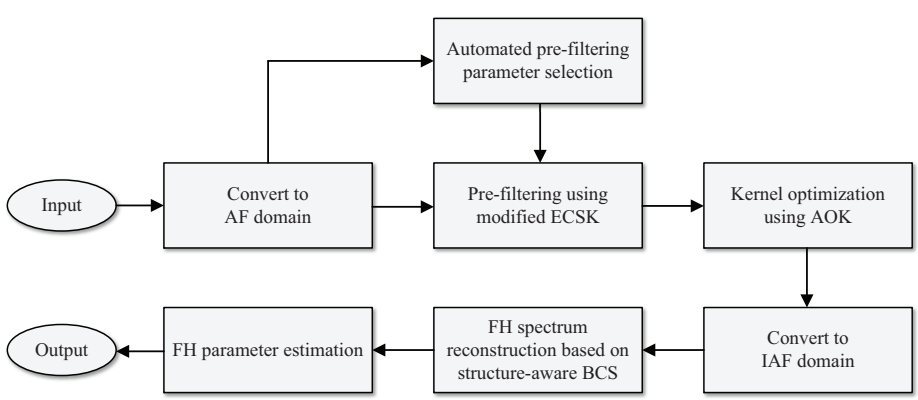
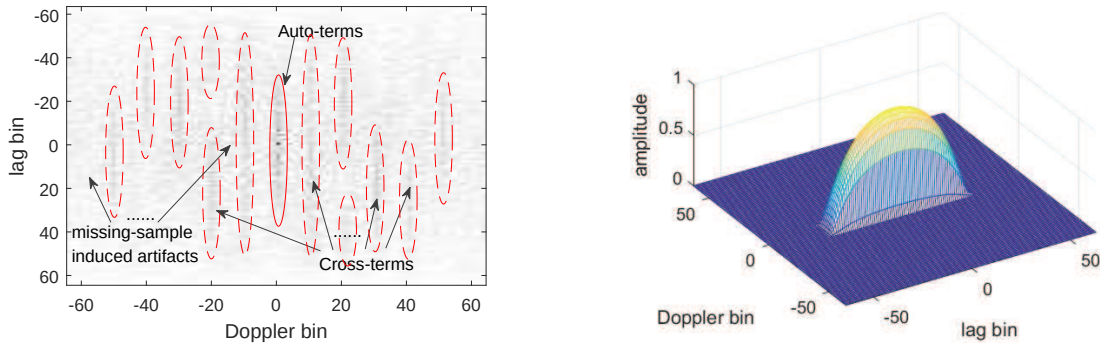


Fig. 1. Flowchart of the proposed signal processing scheme.



(a) (b)

Fig. 2. Illustration of undesired terms and kernel design. (a) A typical AF plane of an FH signal. (b) An illustrative example of revised ECSK.

where C_{ss} and A_{ss} respectively denote the IAF and AF of the original FH signal $s[n]$ without missing samples. The term A_{ss} in (7) contains $\sum_{h=1}^{H} hK_h$ auto-terms and $\left(\sum_{h=1}^{H} hK_h\right)^2 - \sum_{h=1}^{H} hK_h$ cross-terms.

It can be observed from (7) that, the missing-sample AF consists of two parts, i.e., the full-data AF A_{ss} of $s[t]$ and the artifacts due to missing samples. The latter contains the auto-terms of the missing samples and the cross-terms between the signal and the missing samples. The artifacts expressed in (7) resemble noise in the sense that they spread over the entire ambiguity domain. The noise pattern of the first two artifact terms in the ambiguity domain depends on the values of the missing observations and their positions, whereas the third artifact term is only affected by the missing-sample positions. As pointed out in references [10], [11], careful attention should be paid to the last artifact term, which is always located at $\tau = 0$, i.e., along the Doppler frequency axis. This discourages the use of conventional kernels which, due to the required marginal properties, capture all values along the $\tau = 0$ axis. A typical AF magnitude plot of an FH signal is depicted in Fig. 2(a).

B. Adaptive Kernel Design

With the use of the *a priori* information on the TF structure of the FH signal, i.e., its piecewise constant frequency TF signature, we can apply a proper pre-filtering window before optimizing the AOK so as to prevent the artifacts from being falsely identified as desired signal components and misguiding the AOK optimization process. Generally, for signals whose auto-terms are nearly parallel to either the lag or the Doppler axis, which are exactly the case with the FH signals considered in this paper, the extended compact support kernel (ECSK) outperforms the other kernels in terms of artifact suppression and auto-term preservation [21], [28], [29]. The ECSK also provides flexibility to independently adjust the shape and the size of the kernel. In this paper, we modify the ECSK such that different shape control parameters are used for the two branches, i.e., the lag and Doppler, to offer better flexibility. The modified ECSK is formulated as

$$\tilde{g}[\tau,\kappa] = \tilde{g}_1[\tau] \cdot \tilde{G}_2[\kappa], \tag{8}$$

where

$$\tilde{g}_1[\tau] = \begin{cases} \exp\left(\rho_1 + \dfrac{\rho_1 \Xi_1^2}{\tau^2 - \Xi_1^2}\right) & |\tau| < \Xi_1 N, \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

and

$$\tilde{G}_2[\kappa] = \begin{cases} \exp\left(\rho_2 + \dfrac{\rho_2 \Xi_2^2}{\kappa^2 - \Xi_2^2}\right) & |\kappa| < \Xi_2 N, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

respectively represent lag/Doppler window branches. In the above expressions, ρ_1 and ρ_2 denote the shape control parameters of the two branches, and Ξ_1 and Ξ_2 represent their respective sizes. Larger values of ρ_1 and ρ_2 result in a steeper kernel shape in the corresponding branch, whereas larger values of Ξ_1 and Ξ_2 imply a larger kernel size.

In our proposed method, prior to the radial kernel optimization procedure, the short-time AF is pre-filtered by utilizing the modified ECSK *in a time-localized, short-time manner*, as illustrated in Fig. 2(b), where the preserved support for the auto-terms is a sufficiently small region where the Doppler frequency is nearly zero. In doing so, the vertical TFD stripes due to impulsive missing samples, whose AF components spread in the Doppler domain, and noise-like artifacts, whose AF components spread in the entire ambiguity domain, are effectively eliminated.

C. Pre-filtering Parameter Optimization

Enhanced TFD concentration generally yields sharp TF representations and reduced vicinal interference. To achieve an optimal pre-filtering performance that simultaneously maximizes the TFD concentration and minimizes the TFD artifacts, parameter pairs ρ_1 and ρ_2 as well as Ξ_1 and Ξ_2 should be tuned to their optima based on a proper criterion. Several optimization criteria are available in the literature for the evaluation of the concentration performance. Among these criteria, distribution norm-based measures [30], [31] and entropy-based measures [32], [33] are commonly used. However, norm-based measures tend to discriminate poorly concentrated components [34], whereas entropy-based measures are sensitive to amplitude and phase variations [35]. In this context, an efficient energy concentration measure is introduced in [34] that overcomes the aforementioned drawbacks, and has been

applied to automatic determination of the best window length in the computation of spectrogram. In this paper, an automatic parameter tuning method is proposed based on this energy concentration measure. The discrete-time expression of this energy concentration measure can be written as

$$\mathcal{M}\left(\mathbf{F}_{\tau,\kappa}\widetilde{\mathbf{A}}_{\mathbf{xx}}\{n;\tau,\kappa\}\right)=\left(\sum_{\tau}\sum_{\kappa}\left|\mathbf{F}_{\tau,\kappa}\widetilde{\mathbf{A}}_{\mathbf{xx}}\{n;\tau,\kappa\}\right|^{2}\right)^{2}. \tag{11}$$

We define the cost function in our pre-filtering parameter optimization process as

$$\begin{aligned}&\min_{\rho_1,\rho_2,\Xi_1,\Xi_2}\ \mathcal{M}\left(\mathbf{F}_{\tau,\kappa}\widetilde{\mathbf{A}}_{\mathbf{xx}}\{n;\tau,\kappa\};\rho_1,\rho_2,\Xi_1,\Xi_2\right)\\ &\text{s.t.}\quad 0.01\le\rho_1,\rho_2\le 10,\\ &\qquad\ \ 0.01\le\Xi_1,\Xi_2\le 0.5.\end{aligned} \tag{12}$$

The constraints in (12) are set according to the domain of definition and are applicable to different types of FH signals.

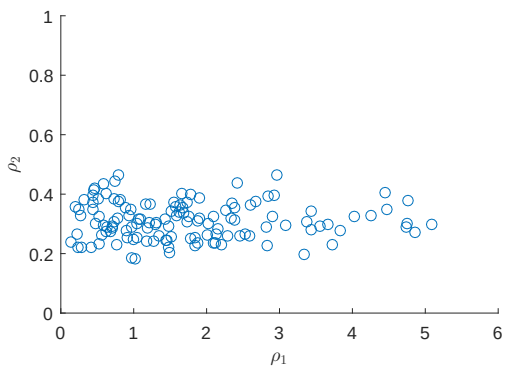


Fig. 3. Illustrative scatter plot of optimized parameters.

To achieve a fully automated optimization of the kernel parameters, an adaptive differential evolution algorithm [36] is adopted. When the *a priori* knowledge about the distribution of potential optima is available, we can further exploit it to arrange the initial population settings. In the simulation examples provided in Section IV, we assume that the potential optima follow a uniform distribution. To better illustrate the parameter optimization process, we provide a 128-point scatter plot of the optimized parameters ρ_1 and ρ_2 in our numerical trials. It can be observed that the optimal values of ρ_2 vary within a relatively narrower range than ρ_1.

D. AOK After Pre-filtering

After applying the modified ECSK as a pre-filtering window, AOK is then employed to further mitigate the effect of artifacts due to missing samples. As discussed in Section II-A, such artifacts spread over the entire ambiguity domain. The AOK is a well-known data-dependent kernel, which is designed by solving the following optimization problem [23]:

$$\begin{aligned}&\Phi_{\text{opt}}(r,\psi)=\underset{\Phi(r,\psi)}{\arg\max}\int_0^{2\pi}\int_0^{\infty}|A(r,\psi)\Phi(r,\psi)|^2 r\mathrm{d}r\mathrm{d}\psi\\ &\text{s.t.}\quad \Phi(r,\psi)=\exp\left(-\frac{r^2}{2\sigma^2(\psi)}\right),\\ &\qquad\frac{1}{4\pi^2}\int_0^{2\pi}\sigma^2(\psi)d\psi\le\alpha,\end{aligned} \tag{13}$$

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. Equation (13) is optimized in the sense that the signal auto-terms are preserved to the maximum extent within the low-pass Gaussian filter, while the pass-band area of the filter is limited to a total volume of α so as to filter out the cross-terms which are located away from the origin, and to reduce the artifacts and noise that spread over the entire ambiguity domain. The desired resolution and the cross-term attenuation are determined by a proper selection of α.

For signals with time-varying characteristics, AOK is usually implemented with a time-localized short-time AF. At time instant t, a time-adaptive kernel $\Phi_{\text{opt}}(t;r,\psi)$ is produced by substituting the short-time AF $A(t;r,\psi)$ for $A(r,\psi)$ in (13) and then following the polar-coordinate Gaussian kernel optimization procedure for each individual TFD slice [23]. Denoting the rectangular-coordinate short-time AF as $A[n;\tau,\kappa]$, the pre-filtered short-time AF can be expressed as

$$\begin{aligned}\tilde{A}_{xx}[n;\tau,\kappa]&=\tilde{g}[n;\tau,\kappa]\cdot A_{xx}[n;\tau,\kappa]\\ &=\tilde{g}[n;\tau,\kappa]\cdot\int x[u+\tau]w[u-n+\tau]\\ &\quad\cdot x^*[u-\tau]w^*[u-n-\tau]\mathrm{e}^{\jmath\kappa u}\mathrm{d}u,\end{aligned} \tag{14}$$

where $w[n]$ represents a rectangular short-time sliding window. Stacking $\tilde{A}_{xx}[n;\tau,\kappa]$ for all τ and κ results in the short-time AF matrix $\widetilde{\mathbf{A}}_{\mathbf{xx}}\{n;\tau,\kappa\}$. Then, the TFD corresponding to the kernelled AF is obtained as its 2-D DFT w.r.t. τ and κ, expressed as

$$\tilde{\mathbf{W}}_{\mathbf{xx}}\{f,n\}=\mathbf{F}_{\tau,\kappa}\widetilde{\mathbf{A}}_{\mathbf{xx}}\{n;\tau,\kappa\}\mathbf{\Phi}_{\text{opt}}\{n;\tau,\kappa\}, \tag{15}$$

where $\mathbf{\Phi}\{n;\tau,\kappa\}$ is the time-localized AOK matrix represented in the rectangular (τ,κ) coordinate system.

Remarks: It is worth emphasizing that, when compared to references [9], [15], [16], which consider FH spectrum estimation in the context of linear short-time Fourier transform (STFT), the utilization of the bilinear TFR in this paper enables us to better address the missing-sample problem because kernel design and its capability to filter out undesired signal components can be utilized *only* in bilinear TF analysis. This is a key novel contribution of this paper since so far only the linear TF analysis has been used in sparse FH spectrum estimation, and no missing samples have been considered in the literature.

III. BCS-Based FH Spectrum Estimation

A. 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 1-D IDFT of $\tilde{\mathbf{A}}_{\mathbf{xx}}\{\tau,\kappa\}$ w.r.t. κ, i.e.,

$$\tilde{\mathbf{C}}_{\mathbf{xx}}\{\tau,n\}=\mathbf{F}_{\kappa}^{-1}\tilde{\mathbf{A}}_{\mathbf{xx}}\{\tau,\kappa\}. \tag{16}$$

On the other hand, the bilinear TFR matrix is associated with the IAF matrix by the following 1-D Fourier relationship:

$$\tilde{\mathbf{C}}_{\mathbf{xx}}\{\tau,n\}=\mathbf{F}_{f}^{-1}\tilde{\mathbf{W}}_{\mathbf{xx}}\{f,n\}. \tag{17}$$

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 $\tilde{\mathbf{c}}_{\mathbf{xx}}[n]$ as the n-th column of the IAF matrix $\tilde{\mathbf{C}}_{\mathbf{xx}}\{\tau,n\}$, and $\tilde{\mathbf{w}}_{\mathbf{xx}}[n]$ as the n-th column of the bilinear TFR matrix $\tilde{\mathbf{W}}_{\mathbf{xx}}\{f,n\}$. Then, their relationship conforms to the following standard linear model commonly used in CS and sparse reconstruction:

$$\tilde{\mathbf{c}}_{\mathbf{xx}}[n] = \mathbf{F}_f^{-1}\tilde{\mathbf{w}}_{\mathbf{xx}}[n]. \tag{18}$$

Therefore, the TFR can be obtained from sparse reconstruction, in lieu of conventional Fourier transform, by repeating the procedure for each time instant. Various CS algorithms can be used for this purpose. In the following, we consider this problem from a BCS perspective [25], and the structure of the FH spectrum is utilized for improved spectrum estimation. BCS methods are known for their capability to flexibly model sparse signals that not only promote the sparsity of its solution, but also exploit additionally known structures of the sparse signal [37]. 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 $\mathbf{c}$, $\mathbf{\Lambda}$ and $\mathbf{w}$, respectively, i.e.,

$$\mathbf{c} = \mathbf{\Lambda}\mathbf{w}. \tag{19}$$

B. Sparsity Prior

The BCS is a nonparametric solver of sparse linear inverse problems imposing a conditional Gaussian prior with its precision (reciprocal of the variance) guided by a hyperprior of Gamma distribution, i.e., $\alpha_0 \sim \mathrm{Gamma}(c, 1/d)$. The BCS assumes the following likelihood model [38]

$$p(\mathbf{c};\mathbf{w},\gamma_0) = \mathcal{CN}(\mathbf{c};\mathbf{\Lambda}\mathbf{w},\gamma_0\mathbf{I}), \tag{20}$$

where $\gamma_0 = \alpha_0^{-1}$ is the variance. To encourage sparsity of the FH signal TFR, a Dirichlet process prior with a spike-and-slab centering distribution [24], [39] is employed to w_i, i.e., the i-th entry of $\mathbf{w}$, which allows different predictors to have identical coefficients while performing variable selection. That is,

$$p(w_i;\gamma_i,\pi_i) = (1-\pi_i)\delta_0 + \pi_i\mathcal{CN}(w_i;0,\gamma_i), \tag{21}$$

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

To make the inference analytical, a product of two latent variables z_i and θ_i, i.e., $w_i = z_i \cdot \theta_i$, is introduced to follow the PDF in (21), where $\theta_i \sim \mathcal{CN}(\theta_i;0,\gamma_i)$, and z_i is a binary variable that follows the 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. Denote $\mathbf{z} = [z_1,\ldots,z_N]^{\mathrm{T}}$ and $\boldsymbol{\theta} = [\theta_1,\ldots,\theta_N]^{\mathrm{T}}$. The overall prior on $\boldsymbol{\theta}$ w.r.t. a and b can be evaluated analytically through the integration over $\boldsymbol{\alpha}$, and it corresponds to the Student-t distribution [40]. 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 [39]. In practice, the hyper-parameters a, b, c, d are usually assigned to small values to make the corresponding priors flat.

C. Structure Prior

The FH spectrum shows sparse piecewise constant frequencies. This structure characteristic can be exploited to improve the accuracy and robustness of the sparse learning performance. For the underlying FH signals, a continuous structure prior that encourages the FH spectrum to have a longer horizontally linear structure in the TFR is desired. With a slightly increased computational complexity, we extend the model to size 3×5, i.e., the neighborhood entries that are within a Euclidean distance of 2, and the vertical pixels from the proximate frequency rows are taken into consideration when decision is made to the TF entry under test. It is evident from Fig. 4 that the utilization of the structure model with a higher dimension enables more comprehensive characterization and treatment of pixel patterns. In this case, simply dividing various patterns into a fixed number of categories does not adequately characterize the relationship between the neighboring pixels. In addition, in the situations with a low SNR, there will be more artifact residue and a higher level of spectrum distortion. As a result, simply rejecting all entries with vertical non-zero neighbors will degrade the robustness of the algorithm.

In this context, we propose a new structure prior which is related to each individual pattern with a proper nonlinear

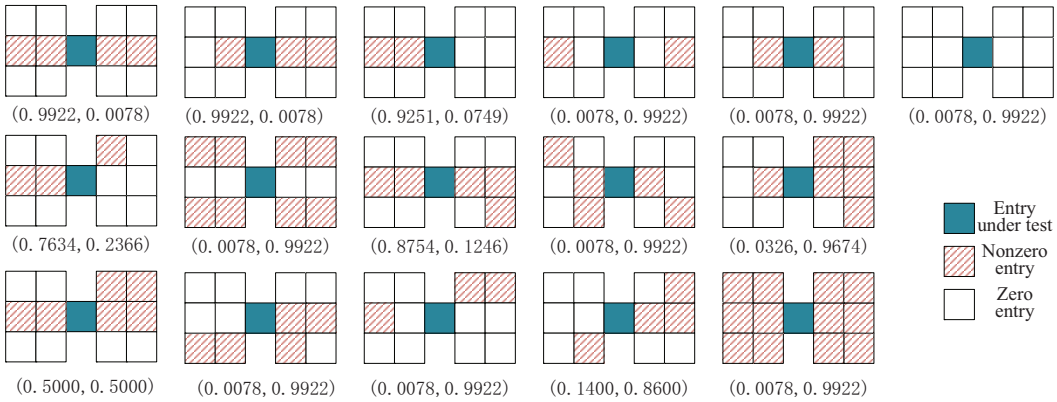


Fig. 4. Examples of different 3×5 TF patterns.

relationship. We first define the neighborhood of index i as

$$\mathcal{J}_{\odot i} \triangleq \{j \,|d\,(i,j) \leq 2, j \in [1,N]\}\,, \tag{22}$$

where $d\,(i,j)$ is the Euclidean distance between i and j. We then define the deleted neighborhood of index i, i.e., the neighborhood of index i with i itself excluded, as

$$\mathcal{J}_{\otimes i} \triangleq \{j \,|d\,(i,j) \leq 2, j \in [1,N]\,, j \neq i\}\,. \tag{23}$$

(a) (b)

Fig. 5. Two patterns of vertical adjacent structure. (a) Indirectly adjacent. (b) Directly adjacent.

The number of nonzero entries at a location i and its neighborhood is denoted as $z_{\mathcal{J}_{\odot i}}$. Note that during the pattern classification process, three rows of $\mathbf{z}$ are investigated, and we denote the location i in the $[n+1]$-th and $[n-1]$-th rows of $\mathbf{z}$ as $i+$ and $i-$, respectively. In Bayesian probability theory, if the posterior distributions belong to the same family as the prior probability distribution, then the prior and posterior are termed conjugate distributions. The Beta distribution is conjugate to Bernoulli likelihood, so π_i is assumed to follow the Beta distribution. For a certain structure prior, the posterior distribution of π_i is derived as

$$p(\pi_i; e, f) = \mathrm{Beta}\left(e + z_{\mathcal{J}_{\odot i}}, f + |\mathcal{J}_{\odot i}| - z_{\mathcal{J}_{\odot i}}\right). \tag{24}$$

The $\mathrm{Beta}(e,f)$ distribution tends to draw small values of π_i when $e < f$, and a large value when $e > f$, while it has no tendency when $e = f$. By choosing proper values of e and f, therefore, we can encourage or discourage the sparsity of the pixel under test, depending on the sparsity support in the neighboring pixels. The value of hyper-parameters e and f should be decimal fractions between 0 and 1. In the previous three-decision-category based method [24], [27], these hyper-parameters are multiples of $1/N$, and N is chosen to be integer power of 2 for computation efficiency. Also, as a rational nonlinear relationship associating the hyper-parameters with the TF structure should encourage longer horizontal structures while discourage high vertical-to-horizontal non-zero neighborhood ratios, we choose a straightforward formula $\left(1 - \left(\frac{1}{2}\right)^{\mathrm{func}(z_{\mathrm{ver}}/z_{\mathrm{hor}})}\right) + \left(\frac{1}{2}\right)^{\mathrm{func}(z_{\mathrm{hor}})}$ to express the nonlinear relationship between hyper-parameter f and the TF structure patterns, and let $e = 1 - f$. On the other hand, some modifications should be made to the formula to ensure that the value of f corresponds to the boundary values of z_{hor}, whereas z_{ver} is constrained to a reasonable range and to avoid zero denominator. As a result, the hyper-parameter f can be derived as

$$f = \begin{cases} 1/N, & \varpi < 1/N, \\ \varpi, & 1/N \leq \varpi < 1, \\ 1 - 1/N, & \varpi \geq 1, \end{cases} \tag{25}$$

where the value of ϖ is determined in (26)-(28).

Remarks: The statistical properties of Beta distribution, such as mode, mean, and variance are closely related to the weight of each shape parameters in their summation. We set these parameters e and f in order to encode the structure prior beliefs. For example, the mean of the Beta distribution in this paper is set to $(1 - \varpi + z_{\mathcal{J}_{\odot i}})/\,|J_{\odot i}|$. A similar parameter setting has also been adopted in several existing references (c.f., e.g., [24], [27]). We set $e = 1 - f$ and keep both hyperparameters for better interpretation and consistency with the existing references.

In the above expressions, the vertical structure factor z_{ver} is assigned different weights $1/2$ and 1 respectively to indirectly and directly adjacent structures as shown in Fig. 5. The reason we discriminate between indirectly and directly adjacent structures is that directly adjacent structures tend to broaden the signal bandwidth. This is contradictory to the fact that the underlying FH signals are instantaneously narrowband. In contrast, indirectly adjacent structures may be formed by the distortion of the desired signal component, noise, or artifact residue, so the weight should be relatively smaller. On the other hand, the horizontal structure factor z_{hor} is obtained by counting the number of continuous adjacent entries. Note that, if the entry under test is located in an isolated line, i.e.,

$$z_{\mathrm{ver}}, z_{\mathrm{hor}} \to \varpi: \ \varpi \triangleq \left(1 - \left(\frac{1}{2}\right)^{\left(\frac{1}{2}\left(\sqrt{(1+4\times 2)\frac{1+2z_{\mathrm{ver}}}{1+z_{\mathrm{hor}}}} - 1\right)\right)^4}\right) + \left(\frac{1}{2}\right)^{z_{\mathrm{hor}}^2}, \tag{26}$$

with

$$z_{\mathrm{ver}} \triangleq \frac{1}{2}\left(z_{\mathcal{J}_{\otimes i+}} + z_{\mathcal{J}_{\otimes i-}} + \sum_{j=1}^{2} z_{i+j}\left(z_{(i+j)+} + z_{(i+j)-}\right)\right), \tag{27}$$

$$z_{\mathrm{hor}} \triangleq \begin{cases} 0, & \prod_{j=1}^{2} z_{i+j} + \prod_{j=1}^{2} z_{i-j} = 0, \\ \sum_{j=1}^{2}\Big((z_{i+1} + z_{i-1}) + (z_{i+1}z_{i+2} + z_{i-1}z_{i-2}) + \dots \\ \qquad + (z_{i+1}z_{i+2}\dots z_{i+j} + z_{i-1}z_{i-2}\dots z_{i-j})\Big), & \text{otherwise.} \end{cases} \tag{28}$$

both the left and right edge pixels are 0, the value of $z_{\rm hor}$ in (28) will be set to 0. Because the codomain of $z_{\rm ver}$ and $z_{\rm hor}$ can be derived as $[0,4]$ and $[0,8]$, respectively, according to (27) and (28), we can further obtain the domain of ϖ as $\varpi \in \left[(1/2)^{16}, 2-(1/2)^{256}\right]$.

Assume that $N=128$. According to the proposed structure prior formation method, the hyper-parameter pairs (e,f) for all the patterns are listed under each case in Fig. 4. These hyper-parameters better reflect the corresponding cluttering situation, by automatically assigning a moderate value to the pattern where a long straight line is present whereas the vertical pixels in the nearby rows take a small value. For those cases where nonzero entries extend in the frequency domain or occur isolatedly, a discouraging value will be asserted to prevent or restrain the structure.

D. Bayesian Inference

Since no closed-form expressions of the Bayesian estimators can be derived, Markov-chain Monte Carlo sampling is used to implement the inference. The maximum likelihood estimation of w_i and γ_i from (21) will generally lead to severe overfitting. To obviate the overfitting problem, a smoother inference model is formulated by defining an automatic relevance determination Gaussian prior over the weights [40]:

$$p\left(\mathbf{w};\boldsymbol{\gamma},\boldsymbol{\pi}\right)=\prod_{i=1}^{N}\left[(1-\pi_i)\delta_0+\pi_i\mathcal{CN}\left(w_i;0,\gamma_i\right)\right], \tag{29}$$

where $\boldsymbol{\gamma}=\left[\gamma_1,\ldots,\gamma_N\right]^{\rm T}$ is a vector consisting of N hyper-parameters that independently control the prior variance of each weight. We can then acquire the posterior distribution of $\mathbf{w}$ by combining (29) with the observation likelihood $p\left(\mathbf{c};\mathbf{w},\gamma_0\right)$ in (20), i.e.,

$$p\left(\mathbf{w}\left|\boldsymbol{\gamma},\boldsymbol{\pi},\mathbf{c},\alpha_0\right.\right)\propto p\left(\mathbf{w};\boldsymbol{\gamma},\boldsymbol{\pi}\right)p\left(\mathbf{c};\mathbf{w},\alpha_0\right). \tag{30}$$

A Gibbs sampler is adopted to implement the Bayesian inference as following. Let $\boldsymbol{\lambda}_i$ be the i-th column of $\boldsymbol{\Lambda}$. Then, the paired Gibbs sampler iteratively samples the observations from the following conditional PDF [24], [39]

$$\begin{aligned}p\left(w_i\left|\mathbf{w}_{\backslash i},\mathbf{c}\right.\right)&=p\left(\theta_i,z_i\left|\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right)\\&=p\left(\theta_i\left|z_i,\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right)p\left(z_i\left|\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right),\end{aligned} \tag{31}$$

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

$$p\left(z_i=1\left|\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right)=\frac{\alpha_i}{1-\alpha_i}\frac{\mathcal{CN}\left(0,\gamma_i\right)}{\mathcal{CN}\left(\tilde{\mu}_i,\gamma_i\right)}, \tag{32}$$

where $\tilde{\mu}_i$ and $\tilde{\gamma}_i$ are respectively updated as

$$\tilde{\mu}_i=\alpha_i^{-1}\alpha_0\boldsymbol{\lambda}_i^H\mathbf{c}_{\backslash i}, \tag{33}$$

$$\tilde{\gamma}_i^{-1}=\tilde{\alpha}_i=\left(\alpha_0\boldsymbol{\lambda}_i^H\boldsymbol{\lambda}_i+\alpha_i\right). \tag{34}$$

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

$$p\left(\theta_i\left|z_i=1,\boldsymbol{\theta}_{\backslash i},\mathbf{z}_{\backslash i},\mathbf{c}\right.\right)=\mathcal{CN}\left(w_i;\tilde{\mu}_i,\gamma_i\right). \tag{35}$$

For the $z_i=0$ 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 (24).

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

$$p\left(\alpha_i;a,b,z_{\mathcal{J}_{\odot i}}\right)=\text{Gamma}\left(a+\frac{z_{\mathcal{J}_{\odot i}}}{2},\frac{1}{b+\frac{\left\|z_{\mathcal{J}_{\odot i}}\right\|_2^2}{2}}\right). \tag{36}$$

After completing all the i iterations, the posterior distribution of the noise precision α_0 is updated as

$$\begin{aligned}&p\left(\alpha_0;c,d,\mathbf{c},\boldsymbol{\Lambda},\theta,\mathbf{z}\right)\\&\quad=\text{Gamma}\left(c+\frac{\text{rank}\{\boldsymbol{\Lambda}\}}{2},\frac{1}{d+\frac{\|\mathbf{c}-\boldsymbol{\Lambda}(\theta\circ\mathbf{z})\|_2^2}{2}}\right).\end{aligned} \tag{37}$$

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

$$\hat{\mathbf{w}}=\arg\max_{\mathbf{w}}p\left(\mathbf{w}\left|\mathbf{c}\right.\right), \tag{38}$$

where marginal distribution $p\left(\mathbf{w}\left|\mathbf{c}\right.\right)$ can be obtained by integrating out the hyper-parameters $\boldsymbol{\gamma}$, $\boldsymbol{\pi}$, and α_0 in (30), as expressed in (39) [39], where $\Gamma\left(u\right)\triangleq\int_0^\infty t^{u-1}{\rm e}^{-t}{\rm d}t$ denotes a Gamma function.

This completes the sparse reconstruction result of (18) for one time instant. The estimation of the entire FH spectrum is rendered by repeating the BCS-based estimation for each column of $\tilde{\mathbf{W}}_{\mathbf{xx}}\{f,n\}$.

Remarks: The proposed method in this paper differs from that of [2] in two aspects: (a) In [2] a threshold-based post-AOK window was adopted, whereas in this paper we pre-filter the running AF with a new ECSK. ECSK is known as the best TF kernel for signals with axially distributed auto-terms [21], [28], [29], and it facilitates independent controlling of the shape and size according to the *a priori* knowledge on the signal structure. An automatic parameter optimization approach for the pre-filtering ECSK kernel is also proposed in this paper.

$$\begin{aligned}p\left(\mathbf{w}\left|\mathbf{c}\right.\right)&\propto\int p\left(\mathbf{w}\left|\boldsymbol{\gamma},\boldsymbol{\pi},\mathbf{c},\alpha_0\right.\right){\rm d}\boldsymbol{\gamma}{\rm d}\boldsymbol{\pi}{\rm d}\alpha_0\\&\propto\left(d+\frac{1}{2}\left\|\mathbf{c}-\boldsymbol{\Lambda}\mathbf{w}\right\|_2^2\right)^{-c-\frac{N}{2}}\prod_{i=1}^{N}\left(\frac{\Gamma\left(a+\frac{z_{\mathcal{J}_{\odot i}}}{2}\right)}{\left(b+\frac{\left\|z_{\mathcal{J}_{\odot i}}\right\|_2^2}{2}\right)^{a+\frac{z_{\mathcal{J}_{\odot i}}}{2}}}\cdot\text{Beta}\left(e+z_{\mathcal{J}_{\odot i}},f+\left|\mathcal{J}_{\odot i}\right|-z_{\mathcal{J}_{\odot i}}\right)\right).\end{aligned} \tag{39}$$

As a result, the cleanest possible running AF is delivered to the AOK optimization process, so that the resultant adaptive kernel design significantly improves the desired TF filtering performance. (b) Unlike in [2] where the structure priors for BCS-based TF reconstruction were designed based on a fixed three-category pattern, in this paper we associate the hyper-parameters with a nonlinear relationship of the TF structure. The modified structure prior is designed to more adequately model the diversified relationship with neighboring TF entries.

E. Computational Complexity

In this subsection, we analyze the computational complexity of the proposed scheme and compare it with the existing approaches for FH spectrum estimation. Three methods are compared, namely, the STFT, sparse linear regression (SLR) [9], [15], and sparse Bayesian learning (SBL) [16] based approaches. Note that, although the terms SBL and BCS are used interchangeably for the same algorithm in the literature, we use SBL and BCS in the sequel to respectively refer to the algorithms developed in the linear and bilinear TF frameworks for convenience.

Let ζ be the length of the short-time slide window. The computational complexity of the STFT-based method is $O(N\log_2\zeta)$, which is the least among all the existing approaches. In comparison, the complexity of the SLR-based method is $O(N^2L^2)$ [9], where L is the number of frequency bins. Similar to the STFT-based method, the SBL approach [16] also partitions the input signals into P overlapped segments through a sliding window. The computational complexity of the SBL approach is then $O(P\zeta^3+Kg^3))$ [16], where K is the number of latent parameters, which is normally truncated to a value close to P for a tractable Bayesian inference, and g denotes the cardinality of the sampled time set in the temporal kernel basis vector, which is typically smaller than P. As stated in [16], the computational complexities of both linear TF based methods [9], [15], [16] are actually in a very similar order. In our proposed scheme, the complexities of the pre-filtering parameter optimization, pre-filtering plus AOK processing, as well as BCS reconstruction stages are $O(GQ^2)$ [41], $O(NL\log_2 L)$ [23], and $O(N^3)$ [40], respectively, where G is the total number of generations, and Q is the dimension of the problem, i.e., the number of the parameters to be optimized. When considering the overall computational complexity, which includes multiple terms, its order is determined by that of the fastest growing term (with the highest order of N). As such, the overall asymptotic computational complexity of the proposed scheme is $O(N^3)$. As such, the computational complexity of the proposed method is much higher than the STFT-based method, but is only slightly higher than the SLR and SBL approaches. This is the price we pay in order to achieve robust and accurate spectrum estimation with missing observations, as we will demonstrate in the next Section.

IV. Numerical Experiments and Analysis

In this section, numerical experiments are conducted to evaluate the performance of the proposed algorithm in comparison with those reported in the literature. In this section, the input SNR is defined as [9], [16]

$$\mathrm{SNR} \triangleq 10\log_{10}\left(\frac{\|\mathbf{x}\|_2^2}{N\sigma^2}\right), \tag{40}$$

where $\mathbf{x}$ is the signal vector, and σ^2 denotes the power of additive white Gaussian noise.

In particular, two performance measures are defined for the evaluation of the hopping time and the instantaneous frequency (IF) detection performance, respectively. The ratio of correct hopping time detection is defined as [16]

$$P_t \triangleq \frac{1}{M_c}\sum_{i=1}^{M_c} D_t(i), \tag{41}$$

where M_c is the number of Monte Carlo trails and $D_t(i)$ is the ratio of correct detections in the i-th Monte Carlo trial. A correct hopping time detection is declared if the estimated hopping instant is less than 3 observations away from the associated true hopping instant. The hopping time statistic is defined as $\Delta_n \triangleq \|x_{n+1}-x_n\|_2^2$. The same definition is used in references [9], [16]. The ratio of incorrect IF detection is defined as [16]

$$E_f \triangleq 1-\frac{1}{M_c}\sum_{i=1}^{M_c} D_f(i), \tag{42}$$

where $D_f(i)$ is the ratio of correct frequency detections in the i-th Monte Carlo trial.

Simulation results are provided to demonstrate the effectiveness of the proposed approach. First, an illustrative example is given in Fig. 6(a), where the FH signals are identical to those used in reference [16]. The signals are generated as follows:

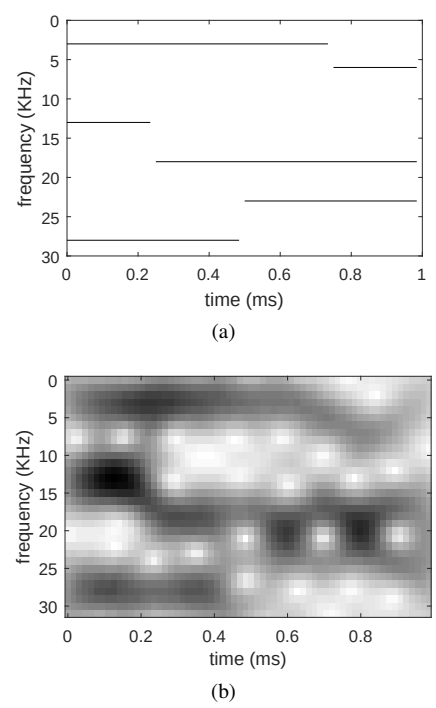


Fig. 6. Simulated FH signals. (a) True TF trajectories of the FH signals; (b) Spectrogram of FH signals with missing samples.

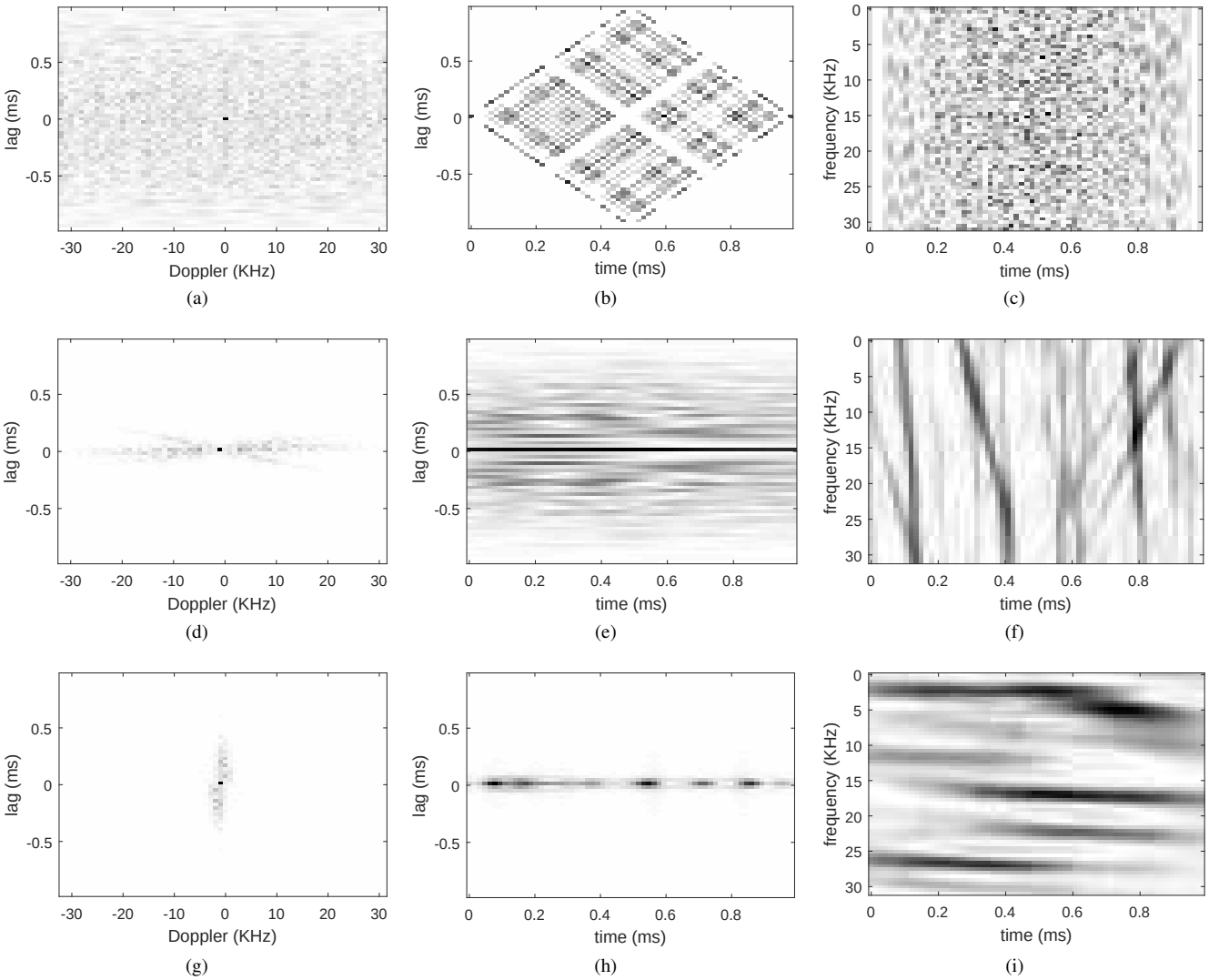


Fig. 7. Joint-variable representations of FH signal with missing observations and kernel processed results. (a) AF; (b) IAF; (c) WVD; (d) AF after applying AOK; (e) IAF after applying AOK; (f) TFR after applying AOK; (g) AF after applying the proposed kernels; (h) IAF after applying the proposed kernels. (i) TFR after applying the proposed kernels.

The first FH component is active with a carrier frequency of 13 KHz within the range of time index $[0:15]$ and the carrier frequency hops to 18 KHz within the range of time index $[16:63]$. The second hopping component is active with a carrier frequency of 28 KHz within the range of time index $[0:31]$ and the carrier frequency hops to 23 KHz within the range of time index $[32:63]$. The third hopping component is active with a carrier frequency of 35 KHz within the range of time index $[0:47]$ and the carrier frequency hops to 6 KHz within the range of time index $[48:63]$. The sampling frequency f_s is 64 KHz. The model hyper-parameters for the priors are set as follows: $a = b = c = d = 10^{-6}$, the value of f is assigned as in (25), and $e = 1 - f$. The initial conditions are set as $\alpha_i(0) = 1, \pi_i(0) = 0$, and $\alpha_0(0) = 10^2/\mathrm{var}(\mathbf{c})$, where $\mathrm{var}(\cdot)$ yields the scalar variance of a vector. Fig. 6(a) shows the true TF trajectories of the generated FH signals. The TF analysis of such multi-component FH signals, particularly at a low input SNR, is a challenging problem. Fig. 6(b) shows the spectrogram with 10% missing samples and input SNR of 30 dB. It is evident that the TF signatures can be hardly recognized with linear approach even in the case where the missing-sample rate is low and the input SNR is high.

In the following, we show the superior performance achieved by the proposed method for the situation where the SNR is set to 0 dB, and the missing-sample rate is 25%. The joint-variable representations of the missing-sample FH spectrum and their kernelled versions are presented in Fig. 7. In Figs. 7(a) through 7(c), no kernel is adopted. The impact of missing samples can be clearly observed from the IAF showing in Fig. 7(b), and the auto-terms can hardly be identified from both AF and WVD in Figs. 7(a) and 7(c). Figs. 7(d) through 7(f) show the corresponding joint-variable representations when the AOK is applied.

In this case, because of the low SNR and the missing samples as well as the required marginal properties, the optimization process in the AOK is severely distorted. As the result, although the AF plane is much cleaner compared with Fig. 7(a), a satisfactory kernelled result cannot be achieved.

Rather, the estimated TFR in Fig. 7(f) shows strong vertical strips. In Figs. 7(g) to 7(i), the proposed revised ECSK plus AOK scheme is adopted. It can be observed from Fig. 7(g) that the auto-term energy in the AF is integrally preserved, while nearly all the undesired terms are suppressed. Nevertheless, direct estimation of the instantaneous frequencies from this plot is still difficult because of the low TF resolution. Therefore, we use the structure-aware BCS to obtain an improved FH spectrum estimation with a finer resolution. The yielding result and the comparison between true and estimated hopping time statistics are respectively depicted in Figs. 8 (a) and (b), which showcase a significant improvement as compared to all the above results depicted in Fig. 7.

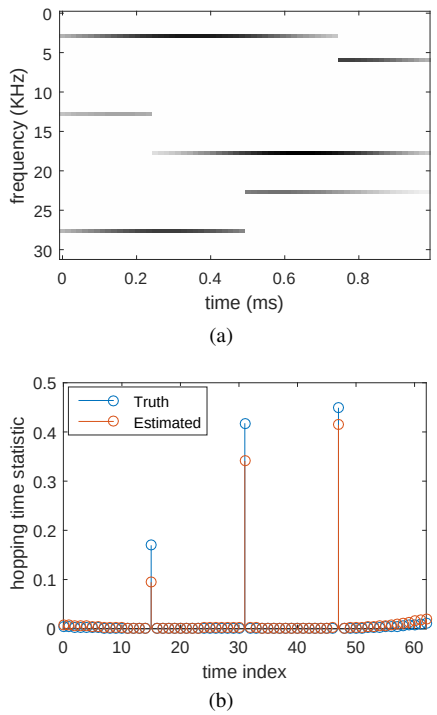


Fig. 8. Estimation results: (a) Estimated TFR; (b) Hopping time statistics.

To better demonstrate the effectiveness of the proposed method with statistical results, 1000 Monte Carlo trials are conducted with the input SNRs varying from 0 dB to 15 dB. In Fig. 9, comparisons are drawn among different existing approaches. It can be summarized from Fig. 9 that an improved performance is obtained by using the proposed method, and the advantage is more remarkable in the low SNR cases. Regarding the influence of missing samples and SNR on the algorithm performance, the statistical results are provided in Fig. 10. Note that the detection performance of the other methods are very poor and thus are not included in Fig. 10 when we compare the performance in the presence of missing observations. As stated above, existing linear TF analysis based approaches cannot robustly perform spectrum estimation with missing observations. Hence, in the presence of missing samples as studied in this paper, these methods yield a detection ratio which is very close to 0 for all SNR values being investigated.

To explore the possibility to skip the pre-filtering parameter optimization process by adopting the average values after collecting sufficient estimations, we conduct numerical trials using the above simulation settings. As the simulation results shown in Fig. 11 indicate, this inevitably affect the pre-filtering performance and consequently slightly degrade the reconstruction accuracy.

Remarks: Unlike the method proposed in [16] which considers FH signal recovery using linear TF analysis (i.e., STFT), the proposed work utilizes the bilinear TF methods. As the bilinear TF methods can use kernel designs to filter out undesired signal components, the proposed method can better utilize the known properties of FH signatures to design the kernels, thus enhancing the FH signal before applying BCS-based sparse reconstruction. This is particularly important in the presence of strong artifacts and noise. Note that the design of such kernels in the structure-aware context is a core contribution of this paper. Such kernel design is not offered in the linear STFT-based approaches. As such, the proposed work is very different to that in [16] and its advantages can be easily understood in concept and are clearly demonstrated through the above simulation results.

V. Conclusion

In this paper, a novel structure-aware FH spectrum estimation approach with the consideration of missing observations was proposed in the sparse reconstruction framework. In particular, a TF kernel was designed to effectively utilize the inherent FH signal structure. The kernelled 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 prior on the Bayesian inference. It was shown that this approach significantly outperforms existing approaches devised for the same problem.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments, which have helped improve the quality and clarity of this paper.

References

[1] S.-H. Liu, Y. D. Zhang, and T. Shan, "Sparsity-based frequency-hopping spectrum estimation with missing samples," in *Proc. 2016 IEEE Radar Conference*, Philadelphia, PA, USA, May 2016.

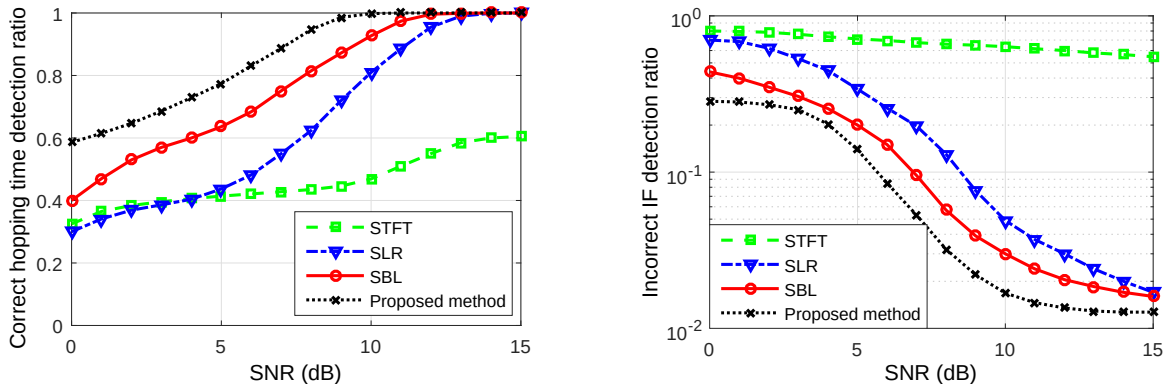
[2] S.-H. Liu, Y. D. Zhang, and T. Shan, "Structure-aware Bayesian compressive sensing for frequency-hopping spectrum estimation," in *Proc. SPIE 9857, Compressive Sensing V: From Diverse Modalities to Big Data Analytics*, Baltimore, MD, USA, pp. 98570N, May 2016.

[3] S. V. Maric and E. L. Titlebaum, "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.*, vol. 40, no. 9, pp. 1442–1447, Sept. 1992.

[4] Y. Zhang and M. G. Amin, "MIMO radar exploiting narrowband frequency-hopping waveforms," in *Proc. European Signal Process. Conf.*, Lausanne, Switzerland, pp. 1–5, Aug. 2008.

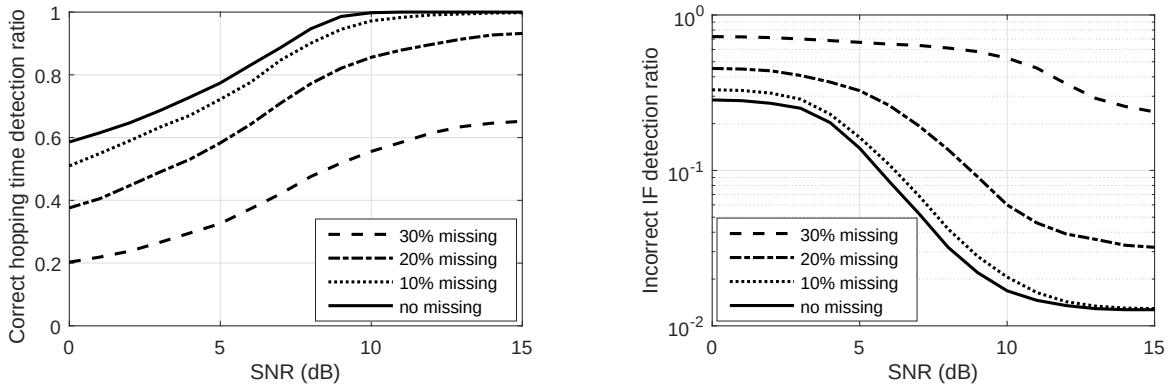
[5] C. Y. Chen and P. P. Vaidyanathan, "MIMO radar ambiguity properties and optimization using frequency-hopping waveforms," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5926–5936, Dec. 2008.

[6] A. R. Hunt, "Use of a frequency-hopping radar for imaging and motion detection through walls," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1402–1408, May 2009.



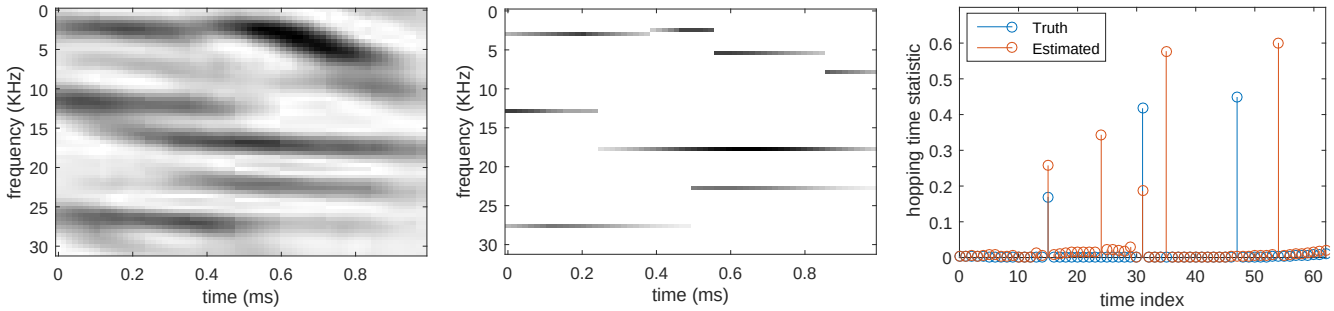
(a) (b)

Fig. 9. Statistic comparisons among different methods without missing observations. (a) Comparison of the correct hopping time detection ratio; (b) Comparison of the incorrect IF detection ratio.



(a) (b)

Fig. 10. Statistic comparisons of the proposed method with different missing-sample rates. (a) Comparison of the correct hopping time detection ratio; (b) Comparison of the incorrect IF detection ratio.



(a) (b) (c)

Fig. 11. Impact of adopting average optimized parameters in the pre-filtering process: (a) TFR after applying the proposed kernels; (b) Estimated TFR; (c) Hopping time statistics.

[7] S. Gogineni, A. Nehorai, "Frequency-hopping code design for MIMO radar estimation using sparse modeling," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3022–3035, June 2012.

[8] 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 Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1144–1154, Feb./Mar./April 1995.

[9] D. Angelosante , G. B. Giannakis, and N. D. Sidiropoulos, "Estimating multiple frequency-hopping signal parameters via sparse linear regression," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5044–5056, Oct. 2010.

[10] M. G. Amin, B. Jokanović, Y. D. Zhang, and F. Ahmad, "A sparsity-perspective to quadratic time-frequency distributions," *Digital Signal Process.*, vol. 46, pp. 175–190, Nov. 2015.

[11] B. Jokanović and M. G. Amin, "Reduced interference sparse time-frequency distributions for compressed observations," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6698–6709, Dec. 2015.

[12] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[13] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pur. Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[14] P. Flandrin and P. Borgnat, "Time-frequency energy distributions meet compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp.

2974-2982, June 2010.
[15] D. Angelosante, G. Giannakis, and N. Sidiropoulos, "Sparse parametric models for robust nonstationary signal analysis: Leveraging the power of sparse regression," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 64–73, Nov. 2013.
[16] L. Zhao, L. Wang, G. Bi, L. Zhang, and H. Zhang, "Robust frequency-hopping spectrum estimation based on sparse Bayesian method," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 781–793, Feb. 2015.
[17] Y. D. Zhang, M. G. Amin, and B. Himed, "Reduced interference time-frequency representations and sparse reconstruction of undersampled data," in *Proc. 21st European Signal Process. Conf. (EUSIPCO)*, Marrakech, Morocco, pp. 1–5, Sep. 2013.
[18] B. Jokanović, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Time-frequency kernel design for sparse joint-variable signal representations," in *Proc. 22nd European Signal Process. Conf. (EUSIPCO)*, Lisbon, Portugal, pp. 1–5, Sep. 2014.
[19] L. Stanković, S. Stanković, I. Orović, and Y. D. Zhang, "Time-frequency analysis of micro-Doppler signals based on compressive sensing," in M. Amin (ed.), *Compressive Sensing for Urban Radars*, CRC Press, 2014.
[20] L. Cohen, *Time-Frequency Analysis*. Prentice Hall, 1995.
[21] B. Boashash, *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, Academic Press, 2015.
[22] D. L. Jones and R. G. Baraniuk, "Signal-dependent time-frequency analysis using a radially Gaussian kernel," *Signal Process.*, vol. 32, no. 3, pp. 263–284, June 1993.
[23] R. G. Baraniuk and D. L. Jones, "An adaptive optimal-kernel time-frequency representation," *IEEE Trans. Signal Process.*, vol. 43, no. 10, pp. 2361–2371, Oct. 1995.
[24] Q. Wu, Y. D. Zhang, and M. G. Amin, "Continuous structure based Bayesian compressive sensing for sparse reconstruction of time-frequency distributions," in *Proc. Int. Conf. Digital Signal Process.*, Hong Kong, China, Aug. 2014, pp. 831–836.
[25] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, June 2008.
[26] S.-H. Liu, Y. D. Zhang, and T. Shan, "Detection of weak astronomical signals with frequency-hopping interference suppression," *Digital Signal Process.*, vol. 72, pp. 1–8, Jan. 2018.
[27] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Process.*, vol. 92, no. 1, pp. 259–269, Jan. 2012.
[28] B. Boashash, N. A. Khan, and T. Ben-Jabeur, "Time-frequency features for pattern recognition using high-resolution TFDs: A tutorial review," *Digital Signal Process.*, vol. 40, pp. 1–30, May 2015.
[29] M. Abed, A. Belouchrani, M. Cheriet, and B. Boashash, "Time-frequency distributions based on compact support kernels: Properties and performance evaluation," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2814–2827, June 2012.
[30] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 12, pp. 2127–2135, Dec. 1990.
[31] D. L. Jones and R. Baraniuk, "A simple scheme for adapting time-frequency representations," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3530–3535, Dec. 1994.
[32] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, and O. J. J. Michel, "Measuring time-frequency information content using the Rényi entropies," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1391–1409, May 2001.
[33] S. Aviyente and W. J. Williams, "Minimum entropy time-frequency distributions," *IEEE Signal Process. Lett.*, vol. 12, no. 1, pp. 37–40, Jan. 2005.
[34] L. J. Stanković, "A measure of some time-frequency distributions concentration," *Signal Process.*, vol. 81, no. 3, pp. 621–631, Mar. 2001.
[35] O. Michel, R. G. Baraniuk, and P. Flandrin, "Time-frequency based distance and divergence measures," in *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, Philadelphia, PA, USA, pp. 64–67, Oct. 1994.
[36] J. Zhang and A. C. Sanderson, "JADE: adaptive differential evolution with optional external archive," *IEEE Trans. Evol. Comput.*, vol. 13, no. 5, pp. 945–958, Oct. 2009.
[37] Y. Zai, L. Xie, and C. Zhang, "Variational Bayesian algorithm for quantized compressed sensing," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2815–2824, June 2013.
[38] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
[39] L. Yu, J. P. Barbot, G. Zheng, and H. Sun, "Compressive sensing for cluster structured sparse signals: Variational Bayes approach," Technical Report, 2011. Available at http://hal.archives-ouvertes.fr/docs/00/57/39/53/PDF/clussvb.pdf.
[40] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sept. 2001.
[41] J.-Q. Zhang and A. C. Sanderson, *Adaptive Differential Evolution: A Robust Approach to Multimodal Problem Optimization*. Springer Berlin Heidelberg, 2009.

Shengheng Liu (S'14-M'17) is currently a Postdoctoral Fellow at the Institute for Digital Communications, School of Engineering, The University of Edinburgh, UK. Prior to joining UoE, he received the B.Eng. and Ph.D. degrees in Electronics Engineering from the School of Information and Electronics, Beijing Institute of Technology, China, in 2010 and 2017 respectively. He also worked as a Visiting Research Associate from 2015 to 2016 at the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA, USA, under the support of the China Scholarship Council. His research interests include compressive sensing and time-frequency analyses for non-stationary signals, interference cancellation and coherent integration for passive bistatic radars, as well as image reconstruction in electrical impedance tomography. He is a frequent reviewer for several top-tier journals, including IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, and IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.

Yimin D. Zhang (SM'01) received his Ph.D. degree from the University of Tsukuba, Tsukuba, Japan, in 1988.

He joined the faculty of the Department of Radio Engineering, Southeast University, Nanjing, China, in 1988. He served as a Director and Technical Manager at the Oriental Science Laboratory, Yokohama, Japan, from 1989 to 1995, and a Senior Technical Manager at the Communication Laboratory Japan, Kawasaki, Japan, from 1995 to 1997. He was a Visiting Researcher at the ATR Adaptive Communications Research Laboratories, Kyoto, Japan, from 1997 to 1998. From 1998 to 2015, he was with the Villanova University, Villanova, PA, where he was a Research Professor at the Center for Advanced Communications, and was the Director of the Wireless Communications and Positioning Laboratory and the Director of the Radio Frequency Identification (RFID) Laboratory. Since August 2015, he has been with the Department of Electrical and Computer Engineering, College of Engineering, Temple University, Philadelphia, PA, where is an Associate Professor. His general research interests lie in the areas of statistical signal and array processing for radar, communications, and satellite navigation applications, including compressive sensing, convex optimization, nonstationary signal and time-frequency analysis, MIMO systems, radar imaging, target localization and tracking, wireless and cooperative networks, and jammer suppression.

He has 12 book chapters and more than 300 journal articles and peer-reviewed conference papers. Dr. Zhang is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and an Editor for the *Signal Processing* journal. He was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS during 2006–2010, and an Associate Editor for the *Journal of the Franklin Institute* during 2007–2013. Dr. Zhang is a member of the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society, and a Technical Co-chair of the 2018 IEEE Sensor Array and Multichannel Signal Processing Workshop.

Tao Shan (M'15) received his B.S. degree from Xidian University, Xi'an, in 1991 and Ph.D. degree from Beijing Institute of Technology in 2004. Currently, he is an Associate Professor with the School of Information and Electronics, Beijing Institute of Technology. From 2014 to 2015, he was a Senior Visiting Scholar at the Center for Advanced Communications, Villanova University, PA. He was a recipient of the first prize of science and technology progress awarded by the Ministry of Education in 2006 and 2007 respectively. His research interests include radar signal processing and time-frequency analysis for non-stationary signals.

Ran Tao (M'00-SM'04) received the B.S. degree from Electronic Engineering Institute of PLA, Hefei, in 1985 and the M.S. and Ph.D. degrees from Harbin Institute of Technology, Harbin, in 1990 and 1993, respectively. He has been a senior visiting scholar at the University of Michigan, Ann Arbor, MI, and the University of Delaware, DE, in 2001 and 2016, respectively. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. He is a Fellow of the Institute of Engineering and Technology (IET), and a Fellow of the Chinese Institute of Electronics (CIE).

Dr. Tao was a recipient of National Science Foundation of China for Distinguished Young Scholars in 2006, and a Distinguished Professor of Changjiang Scholars Program in 2009. He has been a Chief Professor of the Creative Research Groups of the National Natural Science Foundation of China since 2014, and he was a Chief Professor of the Program for Changjiang Scholars and Innovative Research Team in University during 2010 to 2012. He is currently the Vice Chair of IEEE China Council. He is also the Vice Chair of the International Union of Radio Science (URSI) China Council and a Member of Wireless Communication and Signal Processing Commission of URSI. He was a recipient of the first prize of science and technology progress in 2006, 2007, respectively, and the first prize of natural science in 2013, both awarded by the Ministry of Education. His current research interests include fractional Fourier transform and its applications, theory and technology for radar and communication systems. He has 3 books and more than 100 peer-reviewed journal articles.