

Supervised Dictionary Learning for Block Threshold Feature in Compressive Spectrum Sensing

Liyang Lu^{1b}, Graduate Student Member, IEEE, Wenbo Xu^{1b}, Member, IEEE,
Yue Wang^{1b}, Senior Member, IEEE, and Zhi Tian^{1b}, Fellow, IEEE

Abstract—With the development of communication systems towards the high-frequency band, the demand for spectrum resources is ever-increasing, where the research interest has changed from narrowband spectrum sensing to wideband spectrum sensing. The Nyquist-rate-based wideband spectrum sensing with high-rate sampling is being questioned whether it is suitable for real-time applications. On the contrary, the well-known compressive spectrum sensing (CSS) is more appealing due to the compressive sensing (CS) technology, resulting in lower signal acquisition costs. However, the CS algorithms for recovering sparse spectrum generally require multiple iterations, which presents a challenge to the low complexity implementation of spectrum sensing. To address this issue, this paper proposes a novel method for solving the block CSS (BCSS) problem, where the spectrum of primary users is modeled as a block structure signal. Specifically, the block threshold feature (BTF) is utilized to reconstruct the spectrum while bypassing any iterative operations. Furthermore, to improve the performance of the BTF-based BCSS, we develop a novel supervised dictionary learning (SDL) model, based on the theoretical analysis of mutual incoherence and restricted isometry properties. Simulation results not only verify the compatibility of BTF and the SDL model, but also demonstrate the effectiveness and robustness of SDL-BTF-based BCSS for practical implementation.

Index Terms—Block sparsity, compressive sensing, dictionary learning, threshold feature, wideband spectrum sensing.

I. INTRODUCTION

IN COGNITIVE radio network (CRN), spectrum sensing [1]–[4] can improve the utilization of scarce spectrum resources via providing dynamic spectrum access. The secondary users (SUs) attempt to sense the status of primary users (PUs) with the given measurements and dictionary as soon as possible to effectively avoid spectrum conflict. The

current literature has shown that a wider spectrum band provides more access opportunities for SUs [5]–[7], which attracts many researchers and scholars to study wideband spectrum sensing. Though the wider spectrum brings greater capacity to the CRN, the practical wideband spectrum sensing poses new challenges.

One challenge of wideband spectrum sensing in practical applications is the high sampling rate. The conventional Nyquist-rate sampling approach, which causes high hardware complexity and excessive delay in the sensing procedure, is thus unacceptable for practical and agile CRN applications. Compressive sensing (CS) is an effective technology to solve this challenge, leading to the research field of compressive spectrum sensing (CSS) [8]–[13]. CSS methods take advantage of the fact that the spectrum of received signals inherently exhibits sparsity in certain bases, which then allows compression during signal acquisition via sub-Nyquist-rate sampling and then lossless signal reconstruction by using CSS algorithms. Moreover, in practice, PUs' spectrum occupancy is usually endowed a block structure due to their multiple-band spectrum usage, which motivates the investigation of block CSS (BCSS) [14]. Greedy algorithms such as orthogonal matching pursuit (OMP) and block orthogonal matching pursuit (BOMP) are widely adopted for CSS and BCSS respectively due to their low computational complexity [6], [8], [14]. Although these greedy algorithms achieve certain tradeoffs between algorithm complexity and performance for practical implementation, the demand for real-time implementation of CSS in communication services is still continuously growing.

In response to this requirement, a remarkably effective choice for estimating spectrum $\mathbf{x} \in \mathcal{R}^{n \times 1}$ from the compressive measurements $\mathbf{y} \in \mathcal{R}^{m \times 1}$ is the threshold feature (TF) [15]–[18]:

$$\tilde{\mathbf{x}} = \max_K(\mathbf{D}^T \mathbf{y}), \quad (1)$$

where $\mathbf{y} = \mathbf{D}\mathbf{x}$, $\tilde{\mathbf{x}} \in \mathcal{R}^{n \times 1}$ is the sparse approximation of the original spectrum $\mathbf{x}$, $\mathbf{D} \in \mathcal{R}^{m \times n}$ is the dictionary, $\mathbf{y}$ is the measurement vector with sub-Nyquist-rate sampling and $\max_K(\cdot)$ is a quantitative operation that keeps the K elements with the largest absolute value unchanged and sets the others to zero. For BCSS, the corresponding version is called the block TF (BTF) by keeping the k spectrum blocks with the largest ℓ_2 norms unchanged, where $k = \frac{K}{d}$ and d is the block length of each spectrum block. Different from other optimization or iterative algorithms, the operations in

Manuscript received 18 October 2021; revised 11 May 2022; accepted 15 July 2022. Date of publication 25 July 2022; date of current version 9 December 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61871050; the U.S. National Science Foundation under Grant 1939553, Grant 2003211, Grant 2128596, and Grant 2136202; and in part by the Virginia Research Investment Fund under Grant CCI-223996. The associate editor coordinating the review of this article and approving it for publication was Y. Zeng. (Corresponding authors: Wenbo Xu; Yue Wang.)

Liyang Lu and Wenbo Xu are with the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: xuwb@bupt.edu.cn).

Yue Wang and Zhi Tian are with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030 USA (e-mail: ywang56@gmu.edu).

Digital Object Identifier 10.1109/TCCN.2022.3193403

TF only involve matrix multiplication and threshold truncation, which enjoy low computational complexity. However, the considerable low complexity of the TF is at the expense of accuracy [15], [16], which is the main technical challenge of applying the TF to wideband BCSS. Therefore, these observations motivate us to improve BTF to obtain satisfactory BCSS performance while inheriting its outstanding low complexity merits.

A. Related Work

Conventional TF, i.e., the method using the model of (1), and greedy algorithms, such as OMP, BOMP and BOLS, have been widely studied in many applications [6], [8]. In particular, TF is capable of highlighting the correct support elements of the signal by setting small elements to zero [16], where these small entries are usually caused by noise. In this sense, TF succeeds in feature enhancement such as effectively reducing the noise on feature transmission [19], providing powerful nonzero supports thereby achieving state-of-the-art accuracy for sparse weight learning [17], and so on. In [16], the authors apply TF to unsupervised sparse representation. They first provide the theoretical guarantees of TF for correct support recovery, and then develop an unsupervised dictionary learning method (DL-TF) correspondingly. Although the DL-TF method performs well in unsupervised sparse representation, the sparse signal recovery performance of DL-TF is worse than that using the original dictionary $\mathbf{D}$. This is because the unsupervised learning in DL-TF does not utilize the beneficial label information, causing the performance degradation. Moreover, DL-TF ignores the favorable gain brought by the inherent block structure of both the dictionary and the sparse signal. Due to these issues, DL-TF cannot be directly applied to wideband BCSS.

As an extended version of TF, BTF is a superior choice that considers the block structure of the signal. However, to the best of our knowledge, there is no research applying the pure BTF to sparse recovery. This is because although BTF inherits all the advantages of TF, it is limited by its common disadvantage, i.e., the inferior sparse recovery performance caused by its simple product and threshold operations. Fortunately, the supervised dictionary learning (SDL) has emerged as an effective performance improvement method in sparse recovery thanks to the availability of the label information, which shows superiority over the unsupervised dictionary learning [16]. The findings in [20] indicate that sparse recovery based on SDL obtains outstanding accuracy and strong stability. In [21], the authors assert that SDL generates superior dictionaries by introducing discriminative information. In SDL-based CSS, the occupied spectrum for training acts as the supervised information, i.e., the label information. Furthermore, the inherent block structure of the spectrum inspires the joint model of structure-based sparsity [22]–[24] and label-based sparsity [25]–[27]. Therefore, by inspecting the limitation of [16] and further investigating the spectrum sensing with supervised learning in [28]–[31], it is promising that SDL can bring extra benefits to BTF in sparse recovery, so as to promote the improvement of BCSS.

To the best of our knowledge, few researches have applied the low-complexity BTF to BCSS. Meanwhile, SDL has an enabling potential to improve the BCSS performance of BTF, which has not been studied in the current literature.

B. Our Contributions

To address all the above challenges, we propose a novel method, i.e., supervised dictionary learning for block threshold feature (SDL-BTF), to perform efficient and accurate wideband BCSS. The significance of this method is reflected from the following two perspectives. On one hand, it is still very challenging to obtain real-time sensing since conventional BCSS algorithms require multiple iterations for accurate spectrum recovery. For this issue, BTF works as a non-iterative algorithm with low complexity. It only requires a small number of multiplication and threshold operations, which significantly reduces the hardware and computation burdens. On the other hand, SDL is able to generate a discriminative dictionary matched with BTF, which effectively reduces the performance loss caused by the rough approximation of BTF. Therefore, a holistic integration of SDL and BTF in this paper not only addresses the challenge of the real-time requirement in wideband BCSS, but also solves the problem of the unsatisfactory performance of the BTF-based wideband BCSS. The main contributions of this work are summarized as follows.

- 1) *Theoretical analysis for BTF-based BCSS*: To better understand the fundamental limits of BTF in BCSS, its recovery conditions under both noiseless and noisy scenarios are derived. Specifically, we first derive the noiseless sufficient conditions of BTF based on mutual incoherence property (MIP) and restricted isometry property (RIP) to guarantee the exact support reconstruction. Thanks to the consideration of block structure, our theoretical results reveal superior upper bounds of reconstructible sparsity for BTF compared with the existing one for conventional TF [16]. Secondly, to make these analyses interpretable for practical wideband BCSS, we extend our analysis to the noisy scenario by modeling a new coherence metric between the dictionary atoms and the noise. Our results indicate that if this coherence is sufficiently small, then the product results of the atom vectors and the noise will also be small. In this way, BTF can perform effective threshold operation by setting the small product results that related to noise as zeros, so as to retain correct supports.
- 2) *SDL model integrated with BTF*: To effectively improve the sensing performance of BTF, a novel SDL model is proposed, which is elaborated based on the aforementioned theoretical analysis. To the best of our knowledge, this is the first work to develop SDL model for BTF-based BCSS. In this new model, a discriminative dictionary is generated by simultaneously using the block structure and the label information, which in turn facilitates the identification of the correct supports of the recovered spectrum. Meanwhile, a nuclear norm constraint is incorporated to enforce the low rankness

of the learned dictionary, while an orthogonal matrix constraint further contributes to its approximate orthogonality. In order to reduce the coherence between the dictionary atoms and the noise, we propose a reasonable $\ell_{2,\infty}$ norm minimization constraint to weaken their coherence block by block. Simulation results verify that the proposed SDL-BTF BCSS performs faster than the state-of-the-art CSS algorithms with satisfactory sensing performance.

- 3) *Optimization algorithm for the SDL model:* We present an iterative algorithm for optimizing the proposed SDL method via alternating direction method of multipliers (ADMM). To facilitate the utilization of ADMM, we prove that our proposed $\ell_{2,\infty}$ norm minimization is a proximal mapping problem, and thus can be solved with $\mathcal{O}(n \log n)$ complexity. Furthermore, we provide a closed-form solution to the sub-problem of sparse recovery by exploiting the positive semi-definite property of the dictionary product.
- 4) *Complexity and convergence analysis:* To explicitly present the low computational complexity of BTF, we give the complexity analysis for BTF, which reveals that the total complexity of BTF is approximately equal to that of the iterative algorithm with only one iteration. Furthermore, in order to illustrate the feasibility of our proposed optimization algorithm for the SDL method, we analyze its convergence behavior by explaining the convergence of the solution to each sub-problem one by one. The convergence of the entire algorithm is verified by simulations as well, which reveal that it can converge after several iterations.

C. Organization and Notations

The rest of this paper is organized as follows. In Section II, we give the signal and system model of wideband BCSS and study the recovery conditions for BTF. In Section III, we present our SDL model matched with BTF, and its optimization algorithm. In Section IV, simulation results of SDL-BTF-based BCSS are presented, followed by conclusions in Section V.

Throughout the paper, we denote vectors by boldface lowercase letters, e.g., $\mathbf{x}$ and matrices by boldface uppercase letters, e.g., $\mathbf{D}$. $\mathbf{D}_i$ is the i -th column of $\mathbf{D}$. For vector $\mathbf{x}$, the set of its nonzero indices is defined as Ω and $\mathbf{x}_\Omega$ is the vector consisting of the corresponding entries. x_i is the i -th element of the vector $\mathbf{x}$. $\mathbf{x}[i]$ and $\mathbf{D}[i]$ represent the i -th block of vector $\mathbf{x}$ and matrix $\mathbf{D}$. The ℓ_0 -norm, ℓ_1 -norm, ℓ_2 -norm and ℓ_∞ -norm of $\mathbf{x}$ are represented by $\|\mathbf{x}\|_0$, $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$ and $\|\mathbf{x}\|_\infty$ respectively. $\|\mathbf{D}\|_F$ represents the *Frobenius*-norm of $\mathbf{D}$. $\|\mathbf{x}\|_{2,\infty}$ is the mixed norm of $\mathbf{x}$, which obtains the maximum ℓ_2 -norm of the blocks in $\mathbf{x}$. Lowercase k represents the block sparsity, d is the block length and uppercase $K = kd$ stands for overall sparsity. The operation $|\cdot|$ represents the absolute value of its target and the operation $\text{supp}(\cdot)$ denotes the set that containing the indices of the nonzero elements of its objective. In this paper, the columns of the dictionary are normalized to have unit ℓ_2 -norm.

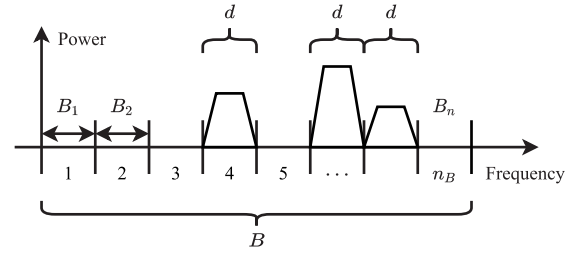


Fig. 1. Signal model of wideband BCSS.

II. MODELS AND RECOVERY GUARANTEES

In this section, we first formulate the signal and system models of the wideband BCSS. Then, we give the theoretical analysis that guarantees reliable recovery for BTF as a preparation for BTF-based BCSS.

A. Signal and System Models

Consider a practical wideband CRN, the spectrum signal of PUs that SUs collect is $\mathbf{s} \in \mathcal{R}^n$. Such spectrum signal can be sparsely represented through a certain basis $\Psi \in \mathcal{R}^{n \times n}$, e.g., the inverse discrete Fourier transform (IDFT) matrix. That is, $\mathbf{s} = \Psi \mathbf{x}$, where $\mathbf{x}$ is the spectrum signal represented in the frequency domain, which usually holds block sparsity structure due to the multiple-band spectrum occupancy by PU systems [14] as illustrated in Fig. 1. Suppose that $\mathbf{x}$ spans the frequency range $0 \sim B$ Hz, containing n_B nonoverlapping sub-channels with equal bandwidth $B_i = \frac{B}{n_B} = d$ ($1 \leq i \leq n_B$). The number of active PUs is called block sparsity. If the number of active PUs is k , then only k blocks of $\mathbf{x}$ have nonzero ℓ_2 -norms.

Then, let $\Phi \in \mathcal{R}^{m \times n}$ denote the compressed sampling matrix, where m and n are the numbers of sub-Nyquist-rate and Nyquist-rate samples, respectively, and the compression ratio is m/n . By denoting the additive noise as $\mathbf{n} \in \mathcal{R}^{m \times 1}$, the compressed measurement signal $\mathbf{y} \in \mathcal{R}^{m \times 1}$ is represented as

$$\mathbf{y} = \Phi \mathbf{s} + \mathbf{n} = \Phi \Psi \mathbf{x} + \mathbf{n} = \mathbf{D} \mathbf{x} + \mathbf{n}, \quad (2)$$

where $\mathbf{D} = \Phi \Psi \in \mathcal{R}^{m \times n}$ is the sensing matrix called dictionary in this paper. The matrix $\mathbf{D}$ in the form of a concatenation of n_B column blocks is given by

$$\mathbf{D} = \begin{bmatrix} \underbrace{\mathbf{D}_1 \cdots \mathbf{D}_d}_{\mathbf{D}[1]} \underbrace{\mathbf{D}_{d+1} \cdots \mathbf{D}_{2d}}_{\mathbf{D}[2]} \cdots \underbrace{\mathbf{D}_{n-d+1} \cdots \mathbf{D}_n}_{\mathbf{D}[n_B]} \end{bmatrix}, \quad (3)$$

where $\mathbf{D}[i] \in \mathcal{R}^{m \times d}$ is the i -th block of $\mathbf{D}$.

As shown in Fig. 2, the system model of our SDL-BTF contains two stages, i.e., the offline training stage and the online learning stage. In the first stage, the discriminative dictionary is learned from the input training data samples. Then, the learned dictionary is utilized to reconstruct the original spectrum signal in the second stage. For fast implementation, the BTF only involves the matrix multiplication and threshold operations [32]–[34], which is expressed as

$$\tilde{\mathbf{x}} = \max_k (\mathbf{D}^T \mathbf{y}), \quad (4)$$

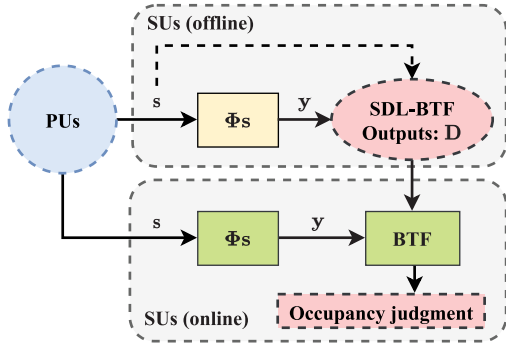


Fig. 2. System model of wideband BCSS based on SDL-BTF.

where $\max_k(\cdot)$ retains the k largest ℓ_2 -norm of the blocks in its object while setting the others to zeros.

Note that the system model considered in this paper exhibits more practical applicability than those based on conventional model-based wideband spectrum sensing solutions [4], [10], from the following two perspectives. On one hand, we utilize our proposed novel SDL method to learn an elaborate dictionary in the offline stage, which is more efficient than the conventional system model that directly uses random matrix as their dictionary. The learned dictionary effectively improves the discriminative ability and enhances the sensing performance. On the other hand, we not only take advantage of the block structure of the sparse spectrum but also employ BTF for fast sensing in the online stage. These new features allow agile sensing capabilities that are promising for the requirement of real-time BCSS.

B. The Recovery Guarantees for BTF

We begin with the basic definitions of MIP and block RIP.

Definition 1 (MIP): The coherence of a dictionary $\mathbf{D}$, which represents the similarity of its atoms, is defined as $\mu = \max_{i,j \neq i} |\mathbf{D}_i^T \mathbf{D}_j|$ [35]. The block-coherence and sub-coherence are defined as $\mu_B = \max_{i,j \neq i} \frac{\rho(\mathbf{M}[i,j])}{d}$ and $\nu = \max_l \max_{i,j \neq i} |\mathbf{D}_i^T \mathbf{D}_j|$ [36], respectively, where $\mathbf{M}[i,j] = \mathbf{D}^T[i]\mathbf{D}[j]$ and $\mathbf{D}_i, \mathbf{D}_j \in \mathbf{D}[l]$.

Definition 2 (Block RIP): For any k -block sparse signal $\mathbf{x} \in \mathcal{R}^n$ with block length d , the dictionary $\mathbf{D}$ satisfies the block RIP with order k [37], if it shows

$$(1 - \delta_{k|d}) \|\mathbf{x}\|_2^2 \leq \|\mathbf{D}\mathbf{x}\|_2^2 \leq (1 + \delta_{k|d}) \|\mathbf{x}\|_2^2. \quad (5)$$

The RIP constant (RIC) $\delta_{k|d}$ is defined as the smallest positive number that satisfies (5). In the following, we use δ instead of $\delta_{k|d}$ to represent the block RIP for convenience.

1) *Noiseless Case:* The following theorem gives the sufficient condition for correct support recovery.

Theorem 1: For the noiseless case $\mathbf{y} = \mathbf{D}\mathbf{x}$ and $\mathbf{x} = \max_k(\mathbf{D}^T \mathbf{y})$, if the sufficient condition

$$\frac{x_{\min}^2}{x_{\max}^2} \geq 2kd^2\mu_B + k^2d^4\mu_B^2 \quad (6)$$

holds, then $\text{supp}(\tilde{\mathbf{x}}) = \text{supp}(\mathbf{x})$, where $x_{\min}$ and $x_{\max}$ are the elements with the smallest and largest amplitudes in $\mathbf{x}$, and $\tilde{\mathbf{x}}$ is the estimation of $\mathbf{x}$.

Proof: See Appendix A. ■

Corollary 1: When $d = 1$, the condition in Theorem 1 corresponds to the conventional TF, which becomes:

$$\frac{x_{\min}^2}{x_{\max}^2} \geq 2K\mu + K^2\mu^2. \quad (7)$$

When $K \geq 1$, the condition (7) is more relaxed than that given in [16].

Proof: Since $d = 1$, the block structure disappears and μ_B converges to μ , and we get (7). The condition in (7) is more relaxed than the sufficient condition given in [16], i.e., $K\mu \leq \frac{|x_{\min}|}{2|x_{\max}|}$, if $\sqrt{2K\mu + K^2\mu^2} < 2K\mu$. This inequality can be simplified to $K > \frac{2}{3\mu} \geq 1$ since $\mu \in [0, 1]$. ■

Remark 1: Theorem 1 indicates that the smaller the block-coherence of the dictionary is, the easier the BTF is to exactly recover the nonzero supports of the given spectrum. This is consistent with the observations for many reconstruction algorithms [35], [36], [38]. That is, a lower coherence means the atoms of the dictionary are more diverse, so that more information can be used to identify the supports of the spectrum. Furthermore, Corollary 1 indicates that our sufficient condition for conventional TF is always more relaxed than that given in [16] since the sparsity of a sparse spectrum is invariably larger than 1.

In the following, we discuss the recovery conditions by further considering the block RIP.

Theorem 2: Define the RIC of $\mathbf{D}_\Omega^T[i]\mathbf{D}$ as $\delta_{\mathbf{D}_\Omega, i}$ and $\delta_{\max, \Omega} = \max_i \delta_{\mathbf{D}_\Omega, i}$ ($i \in \Omega$). Then, the condition for exactly recovery of BTF in the noiseless case, i.e., $\text{supp}(\tilde{\mathbf{x}}) = \text{supp}(\mathbf{x})$, is given as follows:

$$kd < \frac{\alpha}{1 + \alpha} \left(d + \frac{1 - (d-1)v}{\mu_B} \right), \quad (8)$$

where $\alpha = \sqrt{\frac{1 - \delta_{\max, \Omega}}{1 + \delta_{\max, \Omega}}}$.

Proof: See Appendix B. ■

Remark 2: Considering that $\alpha < 1$ and thus $\frac{\alpha}{1 + \alpha} < \frac{1}{2}$, the condition required for the BTF is stricter than that of BOMP in [36]. The same phenomenon exists when comparing OMP and conventional TF [16]. Such performance gain of OMP-type algorithms is obtained with the compensation of much higher complexity. That is, the OMP-type algorithms iterates K or k times but BTF only iterates once. In Section V, we will give detailed experimental results to show the lightweight implementation of BTF.

Corollary 2: When $\mu_B > \nu(1 - \frac{1}{d}) - \frac{1}{d}$ and $\delta_{\max, \Omega} < 1 - \frac{2\sqrt{1+C^2}-2}{d^2\mu_B+d-d(d-1)\nu}$, the condition in Theorem 2 is more relaxed than that in Theorem 1, where $C = \frac{|x_{\min}|}{|x_{\max}|}$.

Proof: The condition (6) in Theorem 1 can be simplified to

$$kd \leq \frac{-1 + \sqrt{1 + C^2}}{d\mu_B}. \quad (9)$$

By solving the inequality $\frac{\alpha}{1 + \alpha} \left(d + \frac{1 - (d-1)v}{\mu_B} \right) > \frac{-1 + \sqrt{1 + C^2}}{d\mu_B}$, we complete the proof. ■

Remark 3: The precondition in Corollary 2, i.e., $\mu_B > \nu(1 - \frac{1}{d}) - \frac{1}{d}$, is easy to satisfy. For example, if the dictionary

is block orthogonal, i.e., $\nu = 0$, then the precondition always holds. Meanwhile, if the nonzero entries in $\mathbf{x}$ are equal to 1, i.e., $C = 1$, then we obtain $\delta_{\max, \Omega} < 1 - \frac{2\sqrt{2}-2}{d}$. Therefore, the aforementioned analysis indicates that Theorem 2 is more relaxed than Theorem 1. Furthermore, since $1 - \frac{2\sqrt{2}-2}{d}$ approaches 1 when d is large, Corollary 2 is easy to hold in this case.

2) *Noisy Case*: We first give a definition of the coherence between the dictionary $\mathbf{D}$ and the noise $\mathbf{n}$. Then, we extend Theorems 1 and 2 to the noisy case, i.e., Theorems 3 and 4, respectively.

Definition 3: The coherence between the dictionary $\mathbf{D}$ and the noise $\mathbf{n}$ is defined as $\mu_n = \max_i |\mathbf{D}_i^T \mathbf{n}|$.

Theorem 3: If the sufficient condition:

$$\frac{x_{\min}^2}{x_{\max}^2} \geq 2kd^2\mu_B + k^2d^4\mu_B^2 + \frac{\mu_n^2}{|x_{\max}|} + \frac{(2 + 2kd^2\mu_B)\mu_n}{|x_{\max}|} \quad (10)$$

holds, then $\text{supp}(\tilde{\mathbf{x}}) = \text{supp}(\mathbf{x})$, where $\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{n}$.

Proof: See Appendix C. ■

Theorem 3 reveals that it is necessary to minimize the mutual coherence μ_n to obtain high support recovery performance of BTF.

Theorem 4: The condition required for reliable recovery of BTF in noisy case is

$$kd < \frac{\beta}{1 + \beta} \left(d + \frac{1 - (d-1)\nu}{\mu_B} \right), \quad (11)$$

where

$$\frac{1}{\beta} = \left(\frac{\sqrt{1 + \delta_{\max, \Omega}}}{\sqrt{1 - \delta_{\max, \Omega}} - \frac{\|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2, \infty}}{\|\mathbf{x}\|_2}} + \frac{1}{\sqrt{1 - \delta_{\max, \Omega}} \frac{\|\mathbf{x}\|_2}{\|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2, \infty}} - 1} \right). \quad (12)$$

Proof: See Appendix D. ■

Theorem 4 has two crucial points different from the Theorem 2: 1) lower coherence between the support dictionary block and noise ($\|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2, \infty}$), and 2) higher amplitude of the sparse spectrum ($\|\mathbf{x}\|_2$). Theorems 2 and 4 reveal the upper bounds of reconstructible sparsity for reliable recovery. It is also noted that Theorem 4 is stricter than Theorem 2 because of the existing noise.

C. Complexity Analysis

It is self-evident that BTF-based BCSS has an absolute advantage in time complexity over the other CSS iterative algorithms. The time complexity of BTF-based BCSS is $\mathcal{O}(mn + n + k \log \frac{n}{d})$, which is roughly equivalent to the complexity of running an iterative block sparse reconstruction algorithm, such as BOMP, for only one iteration [16], [36]. Meanwhile, the time complexity of the conventional threshold feature (1) is $\mathcal{O}(mn + n + kd \log n)$. As we can see, since $(\frac{n}{d})^k < n^{kd}$, the complexity of BTF-based BCSS is lower than that of the conventional TF. However, because $mn \gg kd$, the

complexities of BTF-BCSS and the conventional TF are both approximated by $\mathcal{O}(mn)$. That is to say, the complexity of BTF-based BCSS is not greater than that of the conventional TF. In a word, compared with other algorithms, BTF-based BCSS is more suitable for real-time spectrum sensing.

III. SUPERVISED DICTIONARY LEARNING FOR BLOCK THRESHOLD FEATURE

In this section, we first present our proposed SDL model for BTF based on the aforementioned theoretical analysis. Then, we propose an iterative algorithm based on ADMM for optimizing this SDL model, followed by its convergence analysis.

A. Model Formulation for SDL-BTF

Our proposed SDL model not only considers the block structure which is ignored in DL-TF [16], but also introduces the supervised learning mechanism. The objective function of SDL-BTF is formulated as

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{X}} \quad & \frac{\lambda}{2} \sum_{i=1}^N \left(\|\mathbf{D}^T (\mathbf{Y}_i - \mathbf{D}\mathbf{X}_i)\|_{2, \infty}^2 - \rho \|\mathbf{X}_i\|_2^2 \right) + \|\mathbf{D}^T \mathbf{D} - \mathbf{I}\|_F^2 \\ & + \frac{\theta}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \frac{\gamma}{2} \|\mathbf{D}\|_* + \frac{\delta}{2} \|\mathbf{J} - \mathbf{X}\|_F^2 \\ \text{s.t.} \quad & \|\mathbf{X}_i\|_{2,0} \leq k, \quad i = 1, 2, \dots, N; \\ & \|\mathbf{D}_j\|_2 = 1, \quad j = 1, 2, \dots, n, \end{aligned} \quad (13)$$

where $0 \leq \lambda, \rho, \theta, \gamma, \delta \leq 1$ are scalars, $\mathbf{D}$ is the dictionary, $\mathbf{Y}$ represents the compressed signal matrix, $\mathbf{X}$ is the sparse signal matrix, $\mathbf{J}$ denotes the label matrix corresponding to $\mathbf{X}$, and $\mathbf{I}$ is the identity matrix.

1) *The dictionary $\mathbf{D}$ should have small block-coherence μ_B , which can be obtained from Theorems 1-4.* That is, the smaller the block-coherence μ_B , the easier these theorems are satisfied. Based on the existence of spectrum norm in the definition of μ_B , it is necessary to reduce the eigenvalues of the target dictionary blocks. In order to achieve this goal, we exploit an optimizable objective function that encourages small nuclear norm of each dictionary block, i.e., the term $\|\mathbf{D}[i]\|_*$, for any i . According to [39, Th. 7.3.9], it is equivalent to obtaining the minimum nuclear norm of the entire dictionary, i.e., $\|\mathbf{D}\|_*$, and this is the reason why it appears in (13). In fact, this condition indicates a low-rank structure of the learned dictionary since it has been proven [40] that the small nuclear norm property and low-rank property of a matrix are equivalent in optimization problem while the latter is highly non-convex.

2) *$\mathbf{D}$ should have small sub-coherence ν . This condition is concluded from Theorem 2 and Theorem 4.* Small sub-coherence ν improves the required sparsity level, which indicates a more relaxed condition for more sparse signals with more sparsity levels. This condition can be achieved by minimizing $\|\mathbf{D}[i]^T \mathbf{D}[i] - \mathbf{I}\|_F^2$ for any i . However, in order to simplify the variable complexity, we exploit the term $\|\mathbf{D}^T \mathbf{D} - \mathbf{I}\|_F^2$ in (13) to replace the minimum optimization problem $\|\mathbf{D}[i]^T \mathbf{D}[i] - \mathbf{I}\|_F^2$ for any i . It is reasonable because a smaller $\|\mathbf{D}^T \mathbf{D} - \mathbf{I}\|_F^2$ generally implies a smaller $\|\mathbf{D}[i]^T \mathbf{D}[i] - \mathbf{I}\|_F^2$.

3) *From Theorem 4, it is necessary to minimize $\frac{\|\mathbf{D}_\Omega^T \mathbf{n}\|_{2,\infty}}{\|\mathbf{x}\|_2}$ to restrain the adverse impact of noise.* This condition aims to improve the upper bound of reconstructible sparsity for reliable recovery with respect to the RIC in (11). It exploits mixed norm to describe the mutual incoherence property between the support blocks of dictionary and the additive noise. Moreover, this condition further takes the ℓ_2 -norm energy of sparse signal into account. Larger effective signal energy is more capable of reducing the negative effect of noise. Since it is hard to solve the optimization problems involving this term, we turn to optimizing the alternative cost function $\|\mathbf{D}_\Omega^T \mathbf{n}\|_{2,\infty} - \rho\|\mathbf{x}\|_2$, which can be well approximated by $\|\mathbf{D}^T \mathbf{n}\|_{2,\infty} - \rho\|\mathbf{x}\|_2$. The term $\sum_{i=1}^N (\|\mathbf{D}^T (\mathbf{Y}_i - \mathbf{D}\mathbf{X}_i)\|_{2,\infty} - \rho\|\mathbf{X}_i\|_2^2)$ in (13) is used to achieve this optimization purpose.

Now, it comes to discuss the label-based sparsity in (13). $\mathbf{J}$ is the corresponding label matrix, which is a predefined sparse label consistency indicator. A simple way is setting $\mathbf{J}$ to be a binary matrix where nonzero entries appear at the position of i -th row and j -th column if the j -th data sample is expected to be represented by the entries which are of the same category to the i -th entry [26]. The last term in (13) is the label consistency regularization function, which directly limits sparse coding by encouraging it to approximate the label consistency matrix sparsely.

Note that our proposed SDL-BTF-based BCSS is applicable to some of-the-shelf CS-based compressive sampling frameworks, such as modulated wideband converter (MWC) [41] and random demodulator (RD) [42], [43]. It is observed that MWC and RD are widely used sub-Nyquist sampling frameworks, and when generating training data set for effective dictionary learning, they are capable of providing compressed measurements, which reliably represent the original sparse spectrum. Therefore, our proposed SDL is able to generate discriminative dictionary for BTF in these systems. Moreover, the application of MWC and RD in wideband CSS aims to reduce sampling rate and thus achieve fast implementation, which coincides with the motivation of our proposed SDL-BTF method for real-time sensing. That is to say, our proposed SDL-BTF and these real CSS frameworks not only supplement to each other, but also facilitate each other.

B. ADMM-Based Algorithm for SDL-BTF

First, two auxiliary variables $\mathbf{Q}$ and $\mathbf{P}$ are introduced to separate problem (13):

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{D}} \quad & \frac{\lambda}{2} \sum_{i=1}^N \left(\|\mathbf{Q}_i\|_{2,\infty}^2 - \rho\|\mathbf{X}_i\|_2^2 \right) + \|\mathbf{D}^T \mathbf{D} - \mathbf{I}\|_F^2 \\ & + \frac{\theta}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \frac{\gamma}{2} \|\mathbf{P}\|_* + \frac{\delta}{2} \|\mathbf{J} - \mathbf{X}\|_F^2 \\ \text{s.t.} \quad & \mathbf{Q} = \mathbf{D}^T (\mathbf{Y} - \mathbf{D}\mathbf{X}); \mathbf{P} = \mathbf{D}; \\ & \|\mathbf{X}_i\|_{2,0} \leq k, \quad i = 1, 2, \dots, N; \\ & \|\mathbf{D}_j\|_2 = 1, \quad j = 1, 2, \dots, n. \end{aligned} \quad (14)$$

Then, the augmented Lagrangian function of (14) is given by

$$\begin{aligned} \zeta(\mathbf{X}, \mathbf{Q}, \mathbf{P}, \mathbf{D}) &= \frac{\lambda}{2} \sum_{i=1}^N \left(\|\mathbf{Q}_i\|_{2,\infty}^2 - \rho\|\mathbf{X}_i\|_2^2 \right) + \|\mathbf{D}^T \mathbf{D} - \mathbf{I}\|_F^2 \\ &+ \frac{\theta}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \frac{\gamma}{2} \|\mathbf{P}\|_* + \frac{\delta}{2} \|\mathbf{J} - \mathbf{X}\|_F^2 \\ &+ \langle \mathbf{L}^1, \mathbf{Q} - \mathbf{D}^T (\mathbf{Y} - \mathbf{D}\mathbf{X}) \rangle + \langle \mathbf{L}^2, \mathbf{P} - \mathbf{D} \rangle \\ &+ \frac{\beta}{2} \left(\|\mathbf{Q} - \mathbf{D}^T (\mathbf{Y} - \mathbf{D}\mathbf{X})\|_F^2 + \|\mathbf{P} - \mathbf{D}\|_F^2 \right) \\ \text{s.t.} \quad & \|\mathbf{X}_i\|_{2,0} \leq k, \quad i = 1, 2, \dots, N; \\ & \|\mathbf{D}_j\|_2 = 1, \quad j = 1, 2, \dots, n, \end{aligned} \quad (15)$$

where $\mathbf{L}^1$ and $\mathbf{L}^2$ represent the Lagrange multipliers, and β is a regularization parameter. Then, we apply the ADMM to solve the optimization problem. The detailed sub-problems for $\mathbf{X}$, $\mathbf{Q}$, $\mathbf{P}$ and $\mathbf{D}$ at the t -th iteration ($t = 0, 1, \dots$) are introduced as follows.

1) *X Sub-Problem:* $\mathbf{X}$ can be updated by solving each $\mathbf{X}_i$ and we give the objective function as follows:

$$\begin{aligned} \mathbf{X}^{t+1} = \min_{\mathbf{X}_i} \quad & -\frac{\lambda\rho}{2\beta} \|\mathbf{X}_i\|_2^2 + \frac{\theta}{2\beta} \|\mathbf{Y}_i - \mathbf{D}\mathbf{X}_i\|_2^2 + \frac{\delta}{2\beta} \|\mathbf{J} - \mathbf{X}\|_2^2 \\ & + \frac{1}{2} \|\mathbf{Q}_i - \mathbf{D}^T (\mathbf{Y}_i - \mathbf{D}\mathbf{X}_i) + \frac{\mathbf{L}_i^1}{\beta}\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{X}_i\|_{2,0} \leq k, \quad i = 1, 2, \dots, N. \end{aligned} \quad (16)$$

There exists many techniques to solve this problem and an efficient method is the direct utilization of iterative algorithm, e.g., block IHT algorithm [33]. In order to reduce the complexity of iterative procedures, we give another solution here. The second order derivative of (16) is

$$\frac{\partial^2 \mathbf{X}^{t+1}}{\partial \mathbf{X}_i} = \frac{\delta - \lambda\rho}{\beta} \mathbf{I} + \frac{\theta}{\beta} \mathbf{D}^T \mathbf{D} + \mathbf{D}^T \mathbf{D} \mathbf{D}^T \mathbf{D}. \quad (17)$$

The matrix on the right side of the equation in (17) is positive semi-definite if $\lambda\rho$ is properly small and it easily holds in the actual optimization procedure. Then, the solution by using the block thresholding operator can be used as shown in [44]:

$$\begin{aligned} \mathbf{X}_i^{t+1} = \max_k \quad & \left(\left(\frac{\delta - \lambda\rho}{\beta} \mathbf{I} + \frac{\theta}{\beta} \mathbf{D}^T \mathbf{D} + \mathbf{D}^T \mathbf{D} \mathbf{D}^T \mathbf{D} \right)^{-1} \right. \\ & \times \left(\frac{\theta}{\beta} \mathbf{D}^T \mathbf{Y}_i + \frac{\delta}{\beta} \mathbf{J} - \mathbf{D}^T \mathbf{D} \mathbf{Q}_i \right. \\ & \left. \left. + \mathbf{D}^T \mathbf{D} \mathbf{D}^T \mathbf{Y}_i - \frac{1}{\beta} \mathbf{D}^T \mathbf{D} \mathbf{L}_i^1 \right) \right). \end{aligned} \quad (18)$$

2) *Q Sub-Problem:* $\mathbf{Q}$ can be solved for each $\mathbf{Q}_i$ and the separate objective function for $\mathbf{Q}_i$ is:

$$\begin{aligned} \mathbf{Q}^{t+1} = \min_{\mathbf{Q}} \quad & \frac{\lambda}{2} \sum_{i=1}^N \|\mathbf{Q}_i\|_{2,\infty}^2 \\ & + \frac{\beta}{\lambda} \left\| \mathbf{Q}_i - \left(\mathbf{D}^T (\mathbf{Y}_i - \mathbf{D}\mathbf{X}_i) - \frac{\mathbf{L}_i^1}{\beta} \right) \right\|_F^2. \end{aligned} \quad (19)$$

The proximal mapping of (19) for $\mathbf{Q}_i$ is

$$\text{prox}_{\tau}^{2,\infty}(\mathbf{h}) = \arg \min_{\mathbf{Q}_i} \tau \|\mathbf{Q}_i\|_{2,\infty}^2 + \|\mathbf{Q}_i - \mathbf{h}\|_2^2, \quad (20)$$

where $\tau = \frac{\lambda}{\beta}$ and $\mathbf{h} \in \mathcal{R}^{kd}$. This problem can be solved by the proximal mapping algorithm in [16], where the vector $\mathbf{h}$ is multi-peak. The detailed discussion is shown in Appendix E.

3) **P Sub-Problem:** Fix the other variables and update $\mathbf{P}$ by the following problem:

$$\mathbf{P}^{t+1} = \min_{\mathbf{P}} \|\mathbf{P}\|_* + \frac{\beta}{\gamma} \left\| \mathbf{P} - \left(\mathbf{D} - \frac{\mathbf{L}^2}{\beta} \right) \right\|_F^2. \quad (21)$$

This can be addressed via the singular-value thresholding operator [45], which is detailed as follows:

$$\mathbf{P}^{t+1} = \mathcal{S}_{\frac{\gamma}{2\beta}} \left(\mathbf{D} - \frac{\mathbf{L}^2}{\beta} \right) = \mathbf{U} \hat{\mathcal{S}}_{\frac{\gamma}{2\beta}}(\Sigma) \mathbf{V}^T, \quad (22)$$

where $\mathbf{U} \Sigma \mathbf{V}^T$ is the SVD of $\mathbf{D} - \frac{\mathbf{L}^2}{\beta}$, $\hat{\mathcal{S}}_{\frac{\gamma}{2\beta}}(\Sigma) = \text{diag}(\{|\Sigma_{ii}| - \frac{\gamma}{2\beta}\}_+)$ and $\{t\}_+ = \max(0, t)$.

4) **D Sub-Problem:** The dictionary $\mathbf{D}$ is updated by solving the following spherical constraint problem:

$$\begin{aligned} \mathbf{D}^{t+1} = \min_{\mathbf{D}} & \frac{1}{\beta} \|\mathbf{D}^T \mathbf{D} - \mathbf{I}\|_F^2 + \frac{\theta}{2\beta} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \\ & + \frac{1}{2} \left(\left\| \mathbf{Q} - \mathbf{D}^T (\mathbf{Y} - \mathbf{D}\mathbf{X}) + \frac{\mathbf{L}^1}{\beta} \right\|_F^2 \right. \\ & \left. + \left\| \mathbf{P} - \mathbf{D} + \frac{\mathbf{L}^2}{\beta} \right\|_F^2 \right) \\ \text{s.t. } & \|\mathbf{D}_j\|_2 = 1, j = 1, 2, \dots, n. \end{aligned} \quad (23)$$

We exploit the curvilinear search algorithm [46] to solve (23).

5) **L Sub-Problem:** The Lagrange multipliers are updated as follows:

$$\begin{cases} \mathbf{L}^{1(t+1)} = \mathbf{L}^{1(t)} + \beta \left(\mathbf{Q}^{t+1} - \mathbf{D}^{T(t+1)} (\mathbf{Y} - \mathbf{D}^{t+1} \mathbf{X}^{t+1}) \right), \\ \mathbf{L}^{2(t+1)} = \mathbf{L}^{2(t)} + \beta \left(\mathbf{P}^{t+1} - \mathbf{D}^{t+1} \right). \end{cases} \quad (24)$$

For complete presentation, we summarize the optimization procedures of our SDL-BTF in Algorithm 1. In Theorems 2 and 4, to obtain larger upper bounds of the reconstructible sparsity, the parameters α and β should be large enough. Since α and β are negatively correlated with the RIC, small RIC $\delta_{\max, \Omega}$ is expected. In other words, the smaller the RIC of the dictionary, the better recovery performance of the BTF. Meanwhile, it is known that MIP condition is stronger than RIP [47], which reveals that small coherence of the dictionary leads to small RIC. Therefore, the orthogonal and nuclear norm constraints in (13), i.e., $\|\mathbf{D}^T \mathbf{D} - \mathbf{I}\|_F^2$ and $\|\mathbf{D}\|_*$, not only guarantee the small coherence of the dictionary but also ensure its small RIC. In Algorithm 1, the sub-problems of these two constraints are addressed by curvilinear search algorithm and singular-value thresholding operator, respectively. This means Algorithm 1 is able to generate a satisfactory dictionary with superior RIP.

Algorithm 1 Dictionary Learning for Block Threshold Feature

Input: Training set $\mathbf{Y}$ and its corresponding label matrix $\mathbf{J}$, $\mathbf{Q}$, $\mathbf{P}$, $\mathbf{L}^1$, $\mathbf{L}^2$, number of data samples N , block sparsity k , ϵ , $t_{\max}$, regularization parameters λ , ρ , θ , γ , δ and β .

Output: $\mathbf{D}$.

- 1: **Initialization:** Random Gaussian matrix $\mathbf{D}^0$. $\mathbf{0}$ matrices $\mathbf{Q}^0$, $\mathbf{P}^0$, $(\mathbf{L}^1)^0$, $(\mathbf{L}^2)^0$ and $\mathbf{X}^1$. $\mathbf{X}^0 = \mathbf{J}$. Regularization parameters are initialized in detail in Section IV. Let $t = 0$.
 - 2: **while** $\|\mathbf{X}^{t+1} - \mathbf{X}^t\|_F > \epsilon$ and $t \leq t_{\max}$ **do**
 - 3: Update the sparse codes $\mathbf{X}_i^t$, $i = 1, 2, \dots, N$ by calculating (18);
 - 4: Update $\mathbf{Q}_i^t$, $i = 1, 2, \dots, N$ by solving (20);
 - 5: Update $\mathbf{P}^t$ by calculating (22);
 - 6: Update $\mathbf{D}^t$ by solving (23);
 - 7: Update $(\mathbf{L}^1)^t$ and $(\mathbf{L}^2)^t$ by calculating (24);
 - 8: $t = t + 1$;
 - 9: **end while**
 - 10: **return** $\mathbf{D}$.
-

C. Convergence Analysis

For the ADMM optimization process of the SDL-BTF given in Algorithm 1, the optimal solution for the $\mathbf{X}$, $\mathbf{Q}$, $\mathbf{P}$ and $\mathbf{D}$ can be addressed by optimizing corresponding functions. First, the closed-form solution of $\mathbf{X}$ can be efficiently processed with the block thresholding operator that guarantees convergence. Then, the convergence of $\mathbf{Q}$ can be guaranteed by using the algorithm in [16] with $\mathcal{O}(n \log n)$ complexity. Moreover, the convergence of $\mathbf{D}$ can be guaranteed by introducing the corresponding auxiliary variable $\mathbf{P}$, which are solved by curvilinear search algorithm [46] and singular-value thresholding [45] respectively. Therefore, the convergence of the SDL-BTF can be guaranteed.

IV. SIMULATION RESULTS

A. Determination for Hyper-Parameters

In this subsection, we first determine the appropriate hyper-parameters by evaluating the support recovery performance of the SDL-BTF in recovering binary signals. Then, We compare the SDL-BTF with four methods to verify the effectiveness of the obtained hyper-parameters. Note that the reconstruction of binary signals can be considered as a generalization of the problem of spectrum estimation, that is, it only focuses on obtaining the occupied positions of the spectrum. We regard that the parameters obtained in this framework are more general and not limited to a certain distribution of spectral amplitude. The data is generated as follows. The sparse signal $\mathbf{X} \in \mathcal{R}^{n \times N}$ to be constructed has k nonzero blocks, containing d entries with value 1 in the randomly selected nonzero blocks. It is worth noting that the size of the sparsity set in this experiment does not exceed 10% of the signal length. Then, the samples in training set are generated as $\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{N}$, where $\mathbf{D} \in \mathcal{R}^{m \times n}$ is a random Gaussian matrix and $\mathbf{N} \in \mathcal{R}^{m \times N}$ is the additive Gaussian noise, whose elements are generated from Gaussian distribution $\mathcal{N}(0, 1)$ and $\mathcal{N}(0, 0.01)$ respectively. The label matrix $\mathbf{J} = \mathbf{X}$. By default,

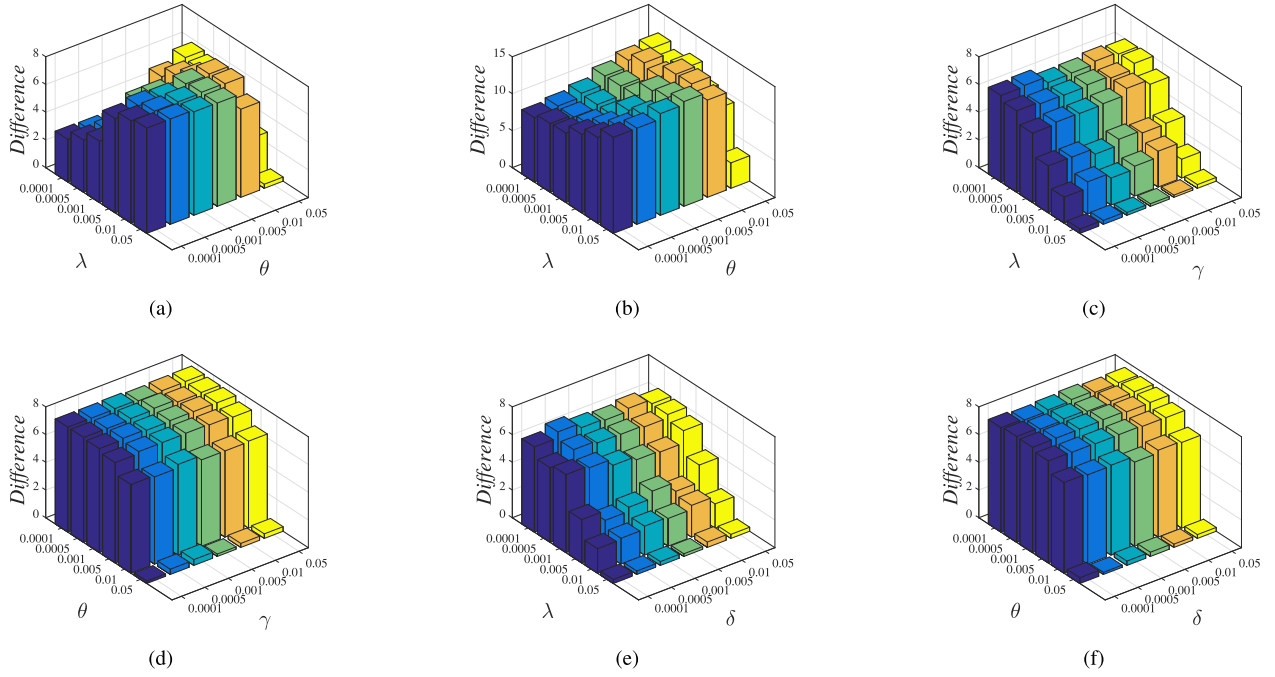


Fig. 3. Results of SDL-BTF on the synthetic data by varying hyper-parameters with $m = 64$ and $n = 128$. (a), (c), (d), (e), (f) are with $k = 4$ and $d = 2$; (b) is with $k = 8$ and $d = 2$.

the hyper-parameters are first set as $\lambda = 0.05$, $\rho = 0.0002$, $\theta = 0.05$, $\beta = 0.001$, $\gamma = 0.05$ and $\delta = 0.01$.

The aforementioned four methods for comparison include the methods using original matrix, random matrix, K-SVD and DL-TF. The detailed settings are given as follows:

- **Original method:** The dictionary $\mathbf{D}$ is set the same as the matrix originally used to generate the measurement signals. The recovery methods are $\tilde{\mathbf{x}} = \max_k(\mathbf{D}^T \mathbf{y})$ and $\tilde{\mathbf{x}} = \max_K(\mathbf{D}^T \mathbf{y})$, where $k = \frac{K}{d}$ is the block sparsity. It is noted that, in unsupervised dictionary learning, the baseline using $\tilde{\mathbf{x}} = \max_K(\mathbf{D}^T \mathbf{y})$ is regarded as having the best performance in [16].

- **Random method:** The dictionary is generated randomly and has nothing to do with the training set. Therefore, the performance of this method should be the worst, which can be considered as the lower bound of the performance of the BTF.

- **K-SVD method:** We exploit the standard K-SVD algorithm to generate the dictionary and apply the learned dictionary to the non-block threshold feature, i.e., $\tilde{\mathbf{x}} = \max_K(\mathbf{D}^T \mathbf{y})$, where $K = kd$ is the total sparsity level.

- **DL-TF method:** The dictionary is learned from the DL-TF method and we apply the conventional threshold feature to obtain the solution.

- **SDL-BTF method:** This is our proposed method. We generate the dictionary by solving (13), where $\mathbf{D}$ is initialized randomly. Then, we exploit the block threshold feature.

We use the support difference [16] to evaluate the accuracy of the reconstructed support as:

$$\text{Difference} = \frac{1}{N} \sum_{i=1}^N \frac{\|\tilde{\mathbf{x}}_{\Omega} \oplus \mathbf{x}_{\Omega}\|_1}{2}, \quad (25)$$

where $\oplus$ represents the element-wise xor operation. The lower the difference is, the better the performance of the method achieves.

The support difference results with varying hyper-parameters are given in Fig. 3. It is observed that the performance of SDL-BTF varies more obviously for different λ and θ , compared with that of other hyper-parameters, which indicates that SDL-BTF is more sensitive to λ and θ . In Fig. 3-(b), the block sparsity is twice as that in Fig. 3-(a). It can be seen that the performance trend in these two figures are similar, but the performance in Fig. 3-(a) is generally better than that in Fig. 3-(b). This result reveals that the stronger structure of the block sparsity, the better the recovery performance. Then, from the results in Figs. 3-(c), (d), (e), (f), the performance of SDL-BTF keeps stable when γ and δ vary, which means its insensitivity of γ and δ . Therefore, γ and δ can be selected in a relatively wide range. Overall, based on these discussions, we choose $\lambda = 0.05$, $\theta = 0.05$, $\gamma = 0.05$ and $\delta = 0.005$ as our default settings in the following experiments.

The second experiment compares the recovered sparse supports using the SDL-BTF-trained dictionary with the actual sparse supports, and shows the convergence of the SDL-BTF method. The ground truth of sparse supports is a random block-diagonal matrix with eight blocks, and $k = 2$, $d = 2$. The synthetic measurement data is shown in Fig. 4(i) (a), and Fig. 4(i) (b) and (c) represent the ground truth of sparse supports and the solution of SDL-BTF. It can be seen that SDL-BTF obtains similar sparse supports corresponding to the ground truth sparse code. The last four sub figures in Fig. 4(i) are the convergence curves with respect to $\|\mathbf{X}^{t+1} - \mathbf{X}^t\|_F$ and the objective function in (13). It is observed that under different parameter settings, SDL-BTF converges within about six

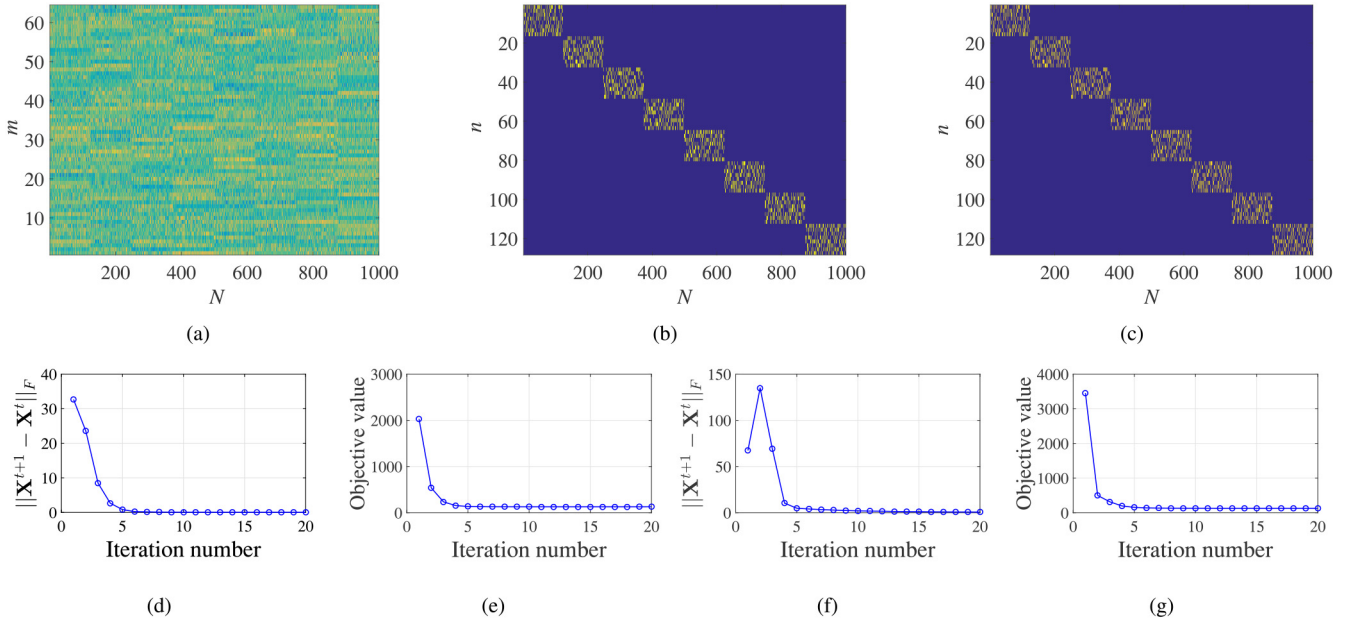


Fig. 4. Results of SDL-BTF on the synthetic data. (a) Input synthetic data; (b) the ground truth of sparse codes; (c) sparse coding results of SDL-BTF from the synthetic data; (d) increment of Frobenius norm of $\mathbf{X}^{t+1}$ and $\mathbf{X}^t$ with $k = 2$ and $d = 2$; (e) the value of objective function of SDL-BTF with $k = 2$ and $d = 2$; (f) increment of Frobenius norm of $\mathbf{X}^{t+1}$ and $\mathbf{X}^t$ with $k = 4$ and $d = 4$; (g) the value of objective function of SDL-BTF with $k = 4$ and $d = 4$.

TABLE I
THE SUPPORT RECOVERY PERFORMANCE COMPARISON WITH $m = 64$, $n = 128$, $d = 2$ AND $d = 4$

	$d = 2$					$d = 4$		
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 1$	$k = 2$	$k = 3$
Original block	0	0.001	0.166	0.647	1.517	0	0.0184	0.5280
Original non-block	0	0.235	1.144	2.125	3.303	0.166	2.074	4.383
Random	1.964	3.874	5.721	7.488	9.218	3.888	7.495	10.651
K-SVD [49]	0.480	0.925	1.792	2.659	3.526	0.520	2.314	10.785
DL-TF [16]	0.296	0.457	1.370	2.306	6.640	0.381	2.178	10.651
SDL-BTF	0	0.001	0.066	0.347	0.977	0	0	0.288

TABLE II
THE SUPPORT RECOVERY PERFORMANCE COMPARISON WITH $m = 96$, $n = 256$, $d = 2$ AND $d = 4$

	$d = 2$					$d = 4$				
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Original block	0	0	0.014	0.228	0.812	0	0	0.116	0.632	1.924
Original non-block	0	0.042	0.600	1.547	2.626	0.091	1.470	3.640	6.302	8.839
Random	1.972	3.942	5.876	7.742	9.584	3.964	7.744	11.368	15.068	18.396
K-SVD [49]	0.233	0.481	0.910	1.608	2.849	0.550	1.818	11.237	14.936	18.414
DL-TF [16]	0.234	0.256	0.555	1.294	3.646	0.333	1.2879	4.918	14.396	18.461
SDL-BTF	0	0	0.008	0.136	0.582	0	0	0.080	0.576	1.612

iterations, which illustrates the effectiveness and correctness of the proposed dictionary learning method..

In the third experiment, we compare the support recovery performance between different methods. The original block and non-block methods are utilized to determine whether SDL-BTF outperforms the previous best performance [16]. The other dictionary learning methods are exploited to see if the dictionary obtained by SDL-BTF possesses more discriminative property than the other trained dictionaries. The results are given in Table I and Table II, where we vary the block

sparsity k and the block length $d \in [2, 4]$, and the dimensions of the dictionaries are $m = 64$, $n = 128$ and $m = 96$, $n = 256$ respectively. As shown in the tables, the support recovery performance of original block and SDL-BTF methods outperform the other methods and SDL-BTF obtains the best performance in general. As shown in Table I, not surprisingly, since the random matrix is not optimized, the method using random matrix has the worst performance. Among the methods of dictionary learning, K-SVD performs the worst when the block sparsity is low. In Table II, we change the

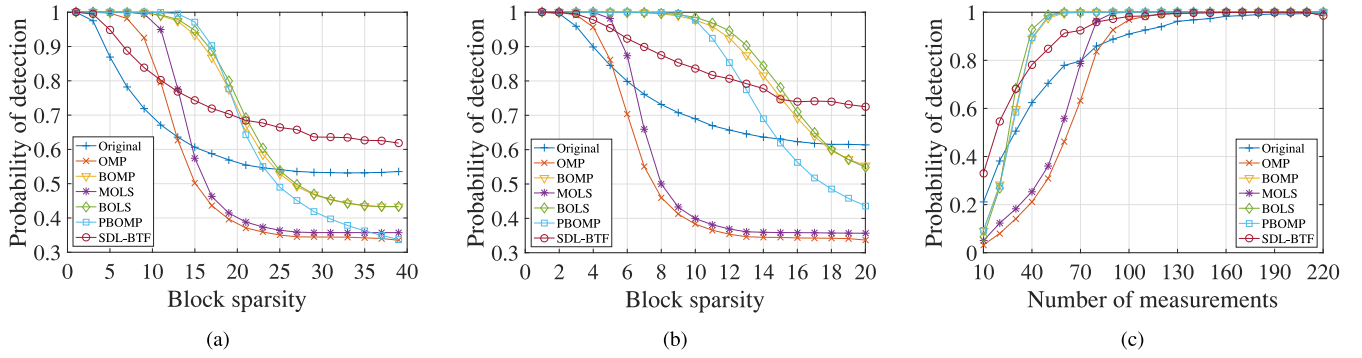


Fig. 5. Spectrum sensing performance versus (a) k with $m = 96$, $n = 256$ and $d = 2$; (b) k with $m = 96$, $n = 256$ and $d = 4$; (c) m with $k = 4$, $d = 4$ and $n = 256$.

dimensions of the dictionary from 64×128 to 96×256 and the compression ratio changes from $1/2$ to $3/8$ accordingly. The overall performance is similar to that in Table I. Due to the increase of the number of measurements, the performance of all the methods is improved when compared with that in Table I. Meanwhile, we know that the larger the block length d , the better the support recovery performance of SDL-BTF when the total sparsity $K = kd$ is fixed. Overall, SDL-BTF exhibits competitive performance among all the methods in support recovery.

B. BCSS on Synthetic Data

The goal of the simulations in this subsection is to verify the low computational complexity and competitive accuracy of our SDL-BTF-based wideband BCSS, compared with other benchmark CSS algorithms. Consider a wideband scenario with $n = 256$. The positions of the nonzero blocks are randomly selected, and the corresponding amplitudes are denoted as $c_i \mathcal{A}$, where c_i is generated from the uniform distribution $\mathcal{U}(1, 2)$ ($1 \leq i \leq k$), and $\mathcal{A}$ is distributed as $\mathcal{N}(1, 0.01)$ [9]. Then, the compressively collected measurement signals can be expressed in a matrix form as $\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{N}$ where $\mathbf{N}$ is the additive Gaussian noise. If not specified, the entries of $\mathbf{N}$ are distributed as $\mathcal{N}(0, 0.01)$. The number of training set for SDL-BTF is set as $N = 10,000$, and we generate a testing set of 10,000 samples separately. The parameters ϵ and $t_{\max}$ are set as 10^{-3} and 20, respectively. The probability of detection [49] used to measure the spectrum sensing performance is given as follows:

$$P_d = \mathbb{E} \left(1 - \frac{\mathbf{d}^T (\mathbf{d} \neq \hat{\mathbf{d}})}{\mathbf{1}^T \mathbf{d}} \right), \quad (26)$$

where $\mathbb{E}(\cdot)$ represents expectation, $\mathbf{d} \in \{0, 1\}^n$ is the ground-truth state vector and $\mathbf{1}$ denotes an all-one vector. The running time of each method is counted to evaluate their computational efficiency. Our simulation environment is MATLAB R2016a on a desktop computer with 2.90GHz Intel Core i7-10700 CPU and 32.0 GB random access memory (RAM).

For comparative purposes, our simulation includes the following CSS algorithms: Original block; OMP [35]; BOMP [36]; MOLS [50]; BOLS [38]; PBOMP [32]; SDL-BTF. Among them, MOLS and BOLS are two variants of

OLS, which respectively choose multiple entries and a block in each iteration. PBOMP is a perturbed BOMP, which iteratively updates the perturbed dictionary and reconstructs the sparse signals. Note that in our test, all the iterative algorithms run k or K iterations at most before stopping. Meanwhile, if the residual obtained at the intermediate iteration is smaller than a given constant, the iterative algorithms stop immediately. For MOLS, we set the maximum iteration number as K and the number of entries that the algorithm selects in each iteration as 3.

In Fig. 5-(a), we fix the length of each spectrum block d as 2 and the number of compressed measurements as $m = 96$. With the increase of the block sparsity value k of the spectrum, the probabilities of correct detection for all the methods degrade, while our proposed SDL-BTF obtains the least performance degradation. This result indicates that SDL-BTF always presents better recovery performance than the other benchmarks even when the spectrum becomes less sparse. Meanwhile, when the sparsity is strong, the performance of SDL-BTF is competitive with that of the iterative algorithms. Furthermore, the SDL-BTF is always better than the original block method, which reveals that our proposed SDL method does provide a discriminative dictionary for improving the sensing performance.

In Fig. 5-(b), we set $d = 4$. It is shown that the performance of the block algorithms is better than that in Fig. 5-(a), which reveals that a larger d provides stronger structure information on improving the sensing accuracy. When compared with Fig. 5-(a), the similar trend can be obtained in Fig. 5-(b).

Furthermore, we depict the probabilities of correct detection as a function of the number of measurements in Fig. 5-(c), where $k = 4$ and $d = 4$. In general, the performance of all the methods improves with the increasing number of measurements. Meanwhile, the performance of the iterative algorithms exploiting block structure is better compared with the algorithms ignoring the block structure. Since our proposed SDL-BTF effectively improves the performance of the BTF, it performs better than the iterative algorithms that do not use block structure for a wide range of the number of measurements. Such range corresponding to the range of performance gain is wider than the one obtained for the original block method. Moreover, when the number of measurements is small, the probabilities of correct detection of SDL-BTF are

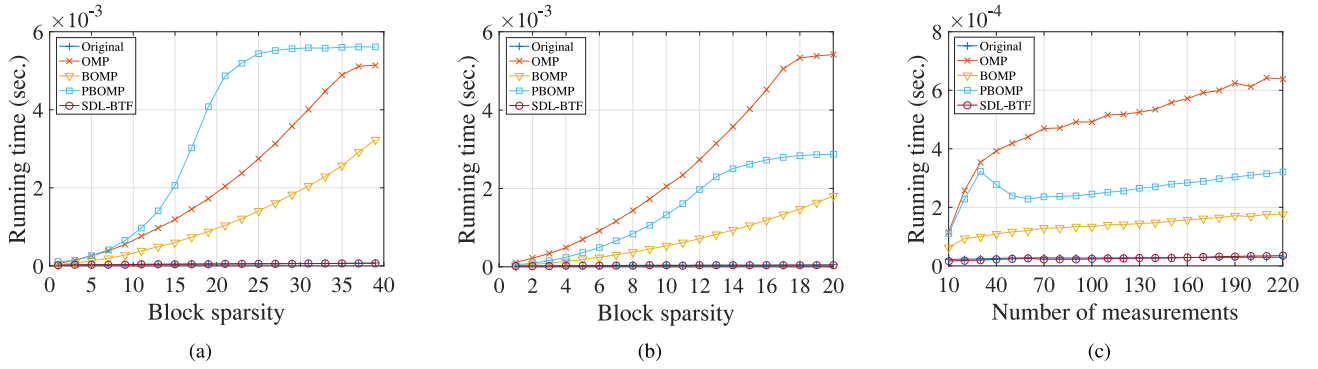


Fig. 6. Running time for different methods. (a) Running time for methods in Fig. 5-(a); (b) Running time for methods in Fig. 5-(b); (c) Running time for methods in Fig. 5-(c).

higher than those of the iterative algorithms. Combining the results in Figs. 5-(a) and 5-(b), it concludes that our proposed SDL-BTF is more capable in providing high accuracy in more challenging scenarios, where the spectrum becomes crowd or when the measurements are parsimoniously collected.

The running time of the aforementioned methods is given in Fig. 6. Since the running time of MOLS and BOLS is much longer than that of the other methods, we do not plot their running time curves. It is noted that the original block and SDL-BTF methods have a significant advantage in running speed over the other methods. For all the cases of the block sparsity of the spectrum, the length of each spectrum block and the number of measurements, the running time of the original block and the SDL-BTF methods are always shorter than that of the others. Furthermore, compared with the iterative algorithms, the increasing speed of the running time of the original block and the SDL-BTF methods is particularly slow. Since the performance of the SDL-BTF is always better than that of the original block method, the SDL-BTF is indeed a desirable choice for real-time BCSS with limited computing resources.

C. BCSS on TV White Space Signal

In this subsection, we consider the TV white space (TVWS) signal [6], [51], whose frequency ranges from 470 ~ 790MHz. The active PUs' signal in time domain is described by:

$$x(t) = \sum_{i=1}^k \sqrt{E_i B_i} \text{sinc}(B_i(t - \Delta t)) \times \cos(2\pi f_i(t - \Delta t)) + n(t), \quad (27)$$

where $\text{sinc}(b) = \frac{\sin(\pi b)}{\pi b}$ and $n(t)$ is the additive noise. The other parameters and their value settings are presented in Table III. The spectrum in frequency domain is obtained by using $N_S = 512$ points DFT. We only consider half of the spectrum, i.e., 256 points, in our simulations due to spectrum symmetry.

In Fig. 7-(a), we plot the probabilities of correct detection as a function of SNR (dB). In general, the performance of the original block method and our proposed SDL-BTF is better than that of the iterative algorithms without using the block structure, but worse than the iterative algorithms using block structure with a moderate loss. The SDL-BTF exhibits better sensing performance than that of the original block method,

TABLE III
PARAMETER SETTINGS

Symbols	Descriptions	Settings
k	Number of active PUs	4
E_i	Energy coefficient	6 ~ 20
B_i	Bandwidth of each sub-channel	2MHz
Δt	Time offset	0 ~ $2\mu s$
f_i	Carrier frequency	469 ~ 789MHz
N_S	Nyquist-rate samples	512

revealing the effectiveness of the SDL method on improving the performance of BTF.

In Fig. 7-(b), the number of active PUs is set to be 8. The probabilities of correct detection of all the methods decrease. Since our proposed SDL-BTF is more stable than the original block method, its performance degrades less than that of the latter. The SDL-BTF method provides a promising sensing performance with remarkably low computing complexity as exhibited by the running time results in Section IV-B. When compared with Fig. 7-(a), the similar conclusions can be obtained in Fig. 7-(b).

V. CONCLUSION

This paper studies the SDL-BTF-based BCSS, which is demonstrated to enjoy low computational complexity. Meanwhile, to illustrate the improvement of the SDL-BTF-based BCSS, we not only give the theoretical guarantees based on MIP for BTF in noiseless and noisy scenarios, but also present extended theoretical analysis based on block RIP theory. Motivated by the theoretical results, we propose an innovative dictionary learning method to learn a more discriminative dictionary for BTF to further improve its spectrum sensing performance.

Numerical experiments on SDL-BTF-based support recovery and wideband BCSS have verified the effectiveness and robustness of the proposed method. It consistently and visibly outperforms K-SVD and DL-TF methods in support recovery, which demonstrates the discriminative ability of the learned dictionary. Furthermore, the SDL-BTF-based wideband BCSS has competitive performance, equivalent to and sometimes even better than the recovery alternatives, e.g., BOMP and BOLS, which prove its remarkable effectiveness. In summary,

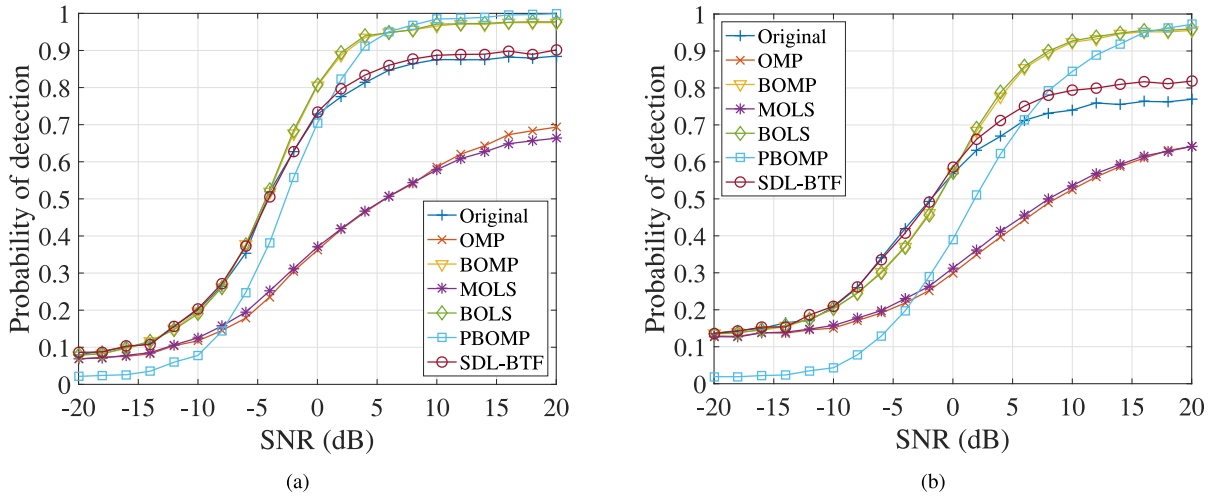


Fig. 7. Spectrum sensing performance versus SNR (dB) with $m = 128$, $n = 256$, $d = 4$ and (a) $k = 4$; (b) $k = 8$.

our proposed SDL-BTF coincides with our motivation. That is, it not only addresses the urgent demand for real-time BCSS by using BTF with fairly low complexity but also significantly improves the sensing performance of BTF via a novel SDL model.

Note that when compared with BOMP and BOLS, our proposed SDL-BTF method experiences slight degradation for some block sparsity levels. Therefore, one future work on SDL-BTF-based BCSS is to seek alternative formulations for generating more discriminative dictionaries, so as to provide further improvement of the sensing performance.

APPENDIX A PROOF OF THEOREM 1

Proof: Without loss of generality, we assume the first k blocks are the support for $\mathbf{x}$. Denote $\mathbf{t} = \mathbf{D}^T \mathbf{y} = \mathbf{D}^T \mathbf{D} \mathbf{x}$. For $\mathbf{t}[l]$ with $1 \leq l \leq k$, we obtain

$$\begin{aligned}
 \|\mathbf{t}[l]\|_2^2 &= \sum_{i=(l-1)d+1}^{ld} \left| x_i + \sum_{j=1, j \neq i}^{kd} \langle \mathbf{D}_i, \mathbf{D}_j \rangle x_j \right|^2 \\
 &\geq \sum_{i=(l-1)d+1}^{ld} \left\| x_i - \left| \sum_{j=1, j \neq i}^{kd} \langle \mathbf{D}_i, \mathbf{D}_j \rangle x_j \right| \right\|^2 \\
 &\geq \sum_{i=(l-1)d+1}^{ld} \left(|x_{\min}|^2 - 2kd\mu |x_{\max}|^2 \right) \\
 &= d \left(|x_{\min}|^2 - 2kd\mu |x_{\max}|^2 \right) \\
 &\geq^{(a)} d \left(|x_{\min}|^2 - 2kd^2\mu_B |x_{\max}|^2 \right), \quad (28)
 \end{aligned}$$

where (a) is obtained from the relationship between μ and μ_B , i.e., $\mu \leq d\mu_B$. Then, for $l > k$:

$$\begin{aligned}
 \max_{l > k} \|\mathbf{t}[l]\|_2^2 &= \max_{l > k} \left(\sum_{i=(l-1)d+1}^{ld} \left(\sum_{j=1}^{kd} \langle \mathbf{D}_i, \mathbf{D}_j \rangle x_j \right)^2 \right) \\
 &\leq d(kd\mu |x_j|)^2 \leq d(kd^2\mu_B |x_{\max}|)^2. \quad (29)
 \end{aligned}$$

Let the smallest ℓ_2 -norm of the block in the support set greater than the largest ℓ_2 -norm of the block in non-support set, and we get

$$\left(|x_{\min}|^2 - 2kd^2\mu_B |x_{\max}|^2 \right) \geq \left(kd^2\mu_B |x_{\max}| \right)^2. \quad (30)$$

Finally, we complete the proof. $\blacksquare$

APPENDIX B PROOF OF THEOREM 2

Proof: Let $\bar{\Omega}$ represent the complementary set of Ω . Since $\mathbf{y} = \mathbf{D} \mathbf{x}$, thus $\mathbf{y} = \mathbf{D}_{\Omega} \mathbf{x}_{\Omega}$. Therefore, $\mathbf{y}$ lies in the space spanned by $\mathbf{D}_{\Omega}$. Then, the following condition should be satisfied to guarantee the correction of the support of $\tilde{\mathbf{x}}$ for $\forall i \in \Omega$:

$$\frac{\|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty}}{\|\mathbf{D}_{\Omega}^T [i] \mathbf{y}\|_2} < 1.$$

It holds that $\mathbf{D}_{\Omega} \mathbf{D}_{\Omega}^{\dagger} \mathbf{y} = \mathbf{y}$. Because $\mathbf{D}_{\Omega} \mathbf{D}_{\Omega}^{\dagger}$ is Hermitian, we obtain $(\mathbf{D}_{\Omega}^{\dagger})^T \mathbf{D}_{\Omega}^T \mathbf{y} = \mathbf{y}$. As a result, $\|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty} = \|\mathbf{D}_{\Omega}^T (\mathbf{D}_{\Omega}^{\dagger})^T \mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty}$. As shown in [36], $\|\mathbf{A} \mathbf{x}\|_{2,\infty} \leq \rho_r(\mathbf{A}) \|\mathbf{x}\|_{2,\infty}$, the equation becomes

$$\begin{aligned}
 \|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty} &\leq \rho_r \left(\mathbf{D}_{\Omega}^T (\mathbf{D}_{\Omega}^{\dagger})^T \right) \|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty} \\
 &= \rho_c \left(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega} \right) \|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty}. \quad (31)
 \end{aligned}$$

where $\rho_r(\cdot)$ and $\rho_c(\cdot)$ are defined in [36] corresponding to spectrum norm of their objects. Therefore,

$$\begin{aligned}
 \frac{\|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty}}{\|\mathbf{D}_{\Omega}^T [i] \mathbf{y}\|_2} &\leq \frac{\rho_c(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega}) \|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty}}{\|\mathbf{D}_{\Omega}^T [i] \mathbf{y}\|_2} \\
 &= \frac{\rho_c(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega}) \|\mathbf{D}_{\Omega}^T \mathbf{D} \mathbf{x}\|_{2,\infty}}{\|\mathbf{D}_{\Omega}^T [i] \mathbf{D} \mathbf{x}\|_2}, \quad (32)
 \end{aligned}$$

for $\forall i \in \Omega$, where the last equality follows from $\mathbf{y} = \mathbf{D} \mathbf{x}$.

As stated in [37], the block RIP is a less stringent requirement when compared with the standard RIP. It has been shown that some random matrices, such as Gaussian matrices and sub-Gaussian matrices, satisfy the block RIP with overwhelming probability. For $i \in \Omega$, since the RIC of $\mathbf{D}_{\Omega}^T [i] \mathbf{D}$ is defined

as $\delta_{\mathbf{D}_{\Omega,i}}$, and $\delta_{\max,\Omega} = \max_i \delta_{\mathbf{D}_{\Omega,i}}$. There exists $i \in \Omega$, the following equality holds:

$$\|\mathbf{D}_{\Omega}^T \mathbf{D} \mathbf{x}\|_{2,\infty} = \|\mathbf{D}_{\Omega}^T [i] \mathbf{D} \mathbf{x}\|_2 \leq \sqrt{1 + \delta_{\max,\Omega}} \|\mathbf{x}\|_2. \quad (33)$$

Meanwhile,

$$\|\mathbf{D}_{\Omega}^T [i] \mathbf{D} \mathbf{x}\|_2 \geq \sqrt{1 - \delta_{\max,\Omega}} \|\mathbf{x}\|_2. \quad (34)$$

As a result, cooperated with (33) and (34), (32) becomes

$$\frac{\|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty}}{\|\mathbf{D}_{\Omega}^T [i] \mathbf{y}\|_2} \leq \frac{\rho_c(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega}) \sqrt{1 + \delta_{\max,\Omega}}}{\sqrt{1 - \delta_{\max,\Omega}}} < 1. \quad (35)$$

According to [36], $\rho_c(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega}) \leq \frac{kd\mu_B}{1 - (d-1)\nu - (k-1)d\mu_B}$, thus (35) will be

$$\frac{kd\mu_B}{1 - (d-1)\nu - (k-1)d\mu_B} \sqrt{\frac{1 + \delta_{\max,\Omega}}{1 - \delta_{\max,\Omega}}} < 1. \quad (36)$$

Finally, by setting $\alpha = \sqrt{\frac{1 - \delta_{\max,\Omega}}{1 + \delta_{\max,\Omega}}}$, we obtain (8). ■

APPENDIX C

PROOF OF THEOREM 3

Proof: Denote $\mathbf{t} = \mathbf{D}^T \mathbf{y} = \mathbf{D}^T (\mathbf{D} \mathbf{x} + \mathbf{n})$. For $\mathbf{t}[l]$ with $1 \leq l \leq k$, (28) in the proof of Theorem 1 is rewritten as:

$$\begin{aligned} \|\mathbf{t}[l]\|_2^2 &= \sum_{i=(l-1)d+1}^{ld} |x_i + \sum_{j=1, j \neq i}^{kd} \langle \mathbf{D}_i, \mathbf{D}_j \rangle x_j \\ &\quad + \langle \mathbf{D}_j, \mathbf{n} \rangle|^2 \geq d \left(|x_{\min}|^2 - 2|x_{\max}| \right. \\ &\quad \times \left. (kd\mu|x_{\max}| + \mu_n) \right). \end{aligned} \quad (37)$$

Then, for $l > k$:

$$\begin{aligned} \max_{l>k} \|\mathbf{t}[l]\|_2^2 &= \max_{l>k} \left(\sum_{i=(l-1)d+1}^{ld} \left(\sum_{j=1}^{kd} \langle \mathbf{D}_i, \mathbf{D}_j \rangle x_j \right. \right. \\ &\quad \left. \left. + \langle \mathbf{D}_j, \mathbf{n} \rangle \right)^2 \right) \\ &\leq d \left(kd^2 \mu_B |x_{\max}| + \mu_n \right)^2. \end{aligned} \quad (38)$$

Finally, we complete the proof in a similar way to the proof of Theorem 1. ■

APPENDIX D

PROOF OF THEOREM 4

Proof: The objective of this proof is similar to that of Theorem 2, i.e., for $\forall i \in \Omega$, $\frac{\|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty}}{\|\mathbf{D}_{\Omega}^T [i] \mathbf{y}\|_2} < 1$. According to the properties of matrix norm, we obtain

$$\begin{aligned} \|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty} &= \|\mathbf{D}_{\Omega}^T \mathbf{y} - \mathbf{D}_{\Omega}^T \mathbf{n} + \mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty} \\ &= \|\mathbf{D}_{\Omega}^T \mathbf{y} - \mathbf{D}_{\Omega}^T (\mathbf{y} - \mathbf{D} \mathbf{x}) + \mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty} \\ &= \|\mathbf{D}_{\Omega}^T \mathbf{D} \mathbf{x} + \mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty} \\ &\leq \|\mathbf{D}_{\Omega}^T \mathbf{D} \mathbf{x}\|_{2,\infty} + \|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty}. \end{aligned} \quad (39)$$

Recall that $\|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty} \leq \rho_c(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega}) \|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty}$. By using (39), we have

$$\begin{aligned} \|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty} &\leq \rho_c(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega}) \left(\|\mathbf{D}_{\Omega}^T \mathbf{D} \mathbf{x}\|_{2,\infty} + \|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty} \right) \\ &\leq \rho_c(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega}) \left(\sqrt{1 + \delta_{\max,\Omega}} \|\mathbf{x}\|_2 + \|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty} \right), \end{aligned} \quad (40)$$

where the last inequality can be obtained from (33). Furthermore, due to $\mathbf{y} = \mathbf{D} \mathbf{x} + \mathbf{n}$, we get

$$\begin{aligned} \|\mathbf{D}_{\Omega}^T [i] \mathbf{y}\|_2 &\geq \|\mathbf{D}_{\Omega}^T [i] (\mathbf{D} \mathbf{x} + \mathbf{n})\|_2 \\ &\geq \|\mathbf{D}_{\Omega}^T [i] \mathbf{D} \mathbf{x}\|_2 - \|\mathbf{D}_{\Omega}^T [i] \mathbf{n}\|_2 \\ &\geq \sqrt{1 - \delta_{\max,\Omega}} \|\mathbf{x}\|_2 - \|\mathbf{D}_{\Omega}^T [i] \mathbf{n}\|_2. \end{aligned} \quad (41)$$

Therefore, by using (40) and (41), we obtain

$$\begin{aligned} &\frac{\|\mathbf{D}_{\Omega}^T \mathbf{y}\|_{2,\infty}}{\|\mathbf{D}_{\Omega}^T [i] \mathbf{y}\|_2} \\ &\leq \rho_c(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega}) \left(\frac{\sqrt{1 + \delta_{\max,\Omega}} \|\mathbf{x}\|_2}{\sqrt{1 - \delta_{\max,\Omega}} \|\mathbf{x}\|_2 - \|\mathbf{D}_{\Omega}^T [i] \mathbf{n}\|_2} \right. \\ &\quad \left. + \frac{\|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty}}{\sqrt{1 - \delta_{\max,\Omega}} \|\mathbf{x}\|_2 - \|\mathbf{D}_{\Omega}^T [i] \mathbf{n}\|_2} \right) \\ &\leq \rho_c(\mathbf{D}_{\Omega}^{\dagger} \mathbf{D}_{\Omega}) \left(\frac{\sqrt{1 + \delta_{\max,\Omega}}}{\sqrt{1 - \delta_{\max,\Omega}} - \frac{\|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty}}{\|\mathbf{x}\|_2}} \right. \\ &\quad \left. + \frac{1}{\sqrt{1 - \delta_{\max,\Omega}} \frac{\|\mathbf{x}\|_2}{\|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty}} - 1} \right), \end{aligned} \quad (42)$$

where the last inequality follows from $\frac{\|\mathbf{D}_{\Omega}^T [i] \mathbf{n}\|_2}{\|\mathbf{D}_{\Omega}^T \mathbf{n}\|_{2,\infty}} \leq 1$. By defining $\frac{1}{\beta}$ as (12), we complete the proof. ■

APPENDIX E

DISCUSSION FOR SOLVING PROBLEM (20) USING ALGORITHM 1 IN [16]

It is noted that for the problem (20) with $\mathbf{h} > 0$, the order of coordinates in the optimal solution $\mathbf{Q}_i$ is the same as the order of the corresponding coordinates in $\mathbf{h}$. The reasons are given in many researches [16], [52], [53]. Based on this truth, we can only focus on the magnitudes of $\mathbf{h}$ since we can take absolute operation for any $\mathbf{h}$.

For the subsequent analysis, we first define a multi-peak signal form. Given an absolute block signal $\mathbf{c}$ with block d , we sort the elements in each block separately from small to large. Then we further sort the signal blocks corresponding to the ℓ_2 -norm of each block and we obtain the signal $\mathbf{c}^m$. We call this processed signal the multi-peak signal and the constraints of multi-peak signal are expressed as:

$$\begin{aligned} \mathbf{c} \text{ s.t. } &\|\mathbf{c}[1]\|_2 \leq \|\mathbf{c}[2]\|_2 \leq \dots \leq \|\mathbf{c}[n_B]\|_2, \\ &\mathbf{c}_1 \leq \mathbf{c}_2 \leq \dots \leq \mathbf{c}_d, \mathbf{c}_{d+1} \leq \dots \leq \mathbf{c}_{2d}, \dots, \\ &\mathbf{c}_{n_B-d} \leq \dots \leq \mathbf{c}_{n_B}. \end{aligned} \quad (43)$$

For simplicity, we use $\mathbf{c} \sim \mathbf{c}^m$ to describe (43) in the following. Then, the problem (20) is converted to

$$\begin{aligned} \arg \min_{\mathbf{q}} \quad & \tau \sum_{i=(k-1)d+1}^{kd} \mathbf{q}_i^2 + \|\mathbf{q} - \mathbf{h}\|_2^2, \\ \text{s.t.} \quad & \mathbf{q} \sim \mathbf{q}^m, \end{aligned} \quad (44)$$

where $\mathbf{q}$ represents $\mathbf{Q}_i$. Further, (44) is equivalent to

$$\begin{aligned} \arg \min_{\mathbf{q}} \quad & \sum_{i=(k-1)d+1}^{kd} (1+\tau) \left(\mathbf{q}_i - \frac{1}{1+\tau} \mathbf{h}_i \right)^2 + \sum_{i=1}^{(k-1)d} (\mathbf{q}_i - \mathbf{h}_i)^2, \\ \text{s.t.} \quad & \mathbf{q} \sim \mathbf{q}^m. \end{aligned} \quad (45)$$

This problem can be solved by the algorithm in [16] if the input vector $\mathbf{h}$ is the multi-peak signal.

ACKNOWLEDGMENT

The authors would like to appreciate the first author of [16] for sharing the codes of the benchmark algorithm.

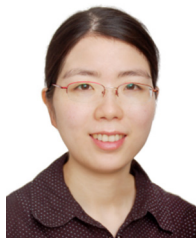
REFERENCES

- [1] O. H. Toma, M. López-Benítez, D. K. Patel, and K. Umebayashi, "Estimation of primary channel activity statistics in cognitive radio based on imperfect spectrum sensing," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2016–2031, Apr. 2020.
- [2] R. Senanayake, P. J. Smith, P. A. Dmochowski, A. Giorgetti, and J. S. Evans, "Mixture detectors for improved spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4335–4348, Jun. 2020.
- [3] Q. Yang *et al.*, "Location based joint spectrum sensing and radio resource allocation in cognitive radio enabled LTE-U systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2967–2979, Mar. 2020.
- [4] L. Lu, W. Xu, Y. Cui, M. Dai, and J. Long, "Block spectrum sensing based on prior information in cognitive radio networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–5.
- [5] B. Hamdaoui, B. Khalfi, and M. Guizani, "Compressed wideband spectrum sensing: Concept, challenges, and enablers," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 136–141, Apr. 2018.
- [6] Y. Luo, J. Dang, and Z. Song, "Optimal compressive spectrum sensing based on sparsity order estimation in wideband cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12094–12106, Dec. 2019.
- [7] Y. Wang and Z. Tian, "Big data in 5G," in *Encyclopedia of Wireless Networks*. New York, NY, USA: Springer, 2018.
- [8] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, pp. 1357–1360.
- [9] Y. Wang, Z. Tian, and C. Feng, "Sparsity order estimation and its application in compressive spectrum sensing for cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2116–2125, Jun. 2012.
- [10] Y. Vasavada and C. Prakash, "Sub-Nyquist spectrum sensing of sparse wideband signals using low-density measurement matrices," *IEEE Trans. Signal Process.*, vol. 68, pp. 3723–3737, Jun. 2020.
- [11] Y. Wang, Z. Tian, and C. Feng, "A two-step compressed spectrum sensing scheme for wideband cognitive radios," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Miami, FL, USA, Dec. 2010, pp. 1–5.
- [12] Y. Wang, Z. Tian, and C. Feng, "Cooperative spectrum sensing based on matrix rank minimization," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3000–3003.
- [13] Y. Wang, Z. Tian, and C. Feng, "Collecting detection diversity and complexity gains in cooperative spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2876–2883, Aug. 2012.
- [14] F. Li and X. Zhao, "Block-structured compressed spectrum sensing with Gaussian mixture noise distribution," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1183–1186, Aug. 2019.
- [15] M. Denil and N. de Freitas, "Recklessly approximate sparse coding," 2012, *arXiv:1208.0959*.
- [16] H. Xu, Z. Wang, H. Yang, D. Liu, and J. Liu, "Learning simple thresholded features with sparse support recovery," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 970–982, Apr. 2020.
- [17] A. Kusupati *et al.*, "Soft threshold weight reparameterization for learnable sparsity," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Jul. 2020, pp. 5544–5555.
- [18] K. Wu, Y. Guo, Z. Li, and C. Zhang, "Sparse coding with gated learned ISTA," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Addis Ababa, Ethiopia, Apr. 2020, pp. 1–27.
- [19] K. Liu, J. Yang, C. Sun, and H. Chi, "Supervised adaptive threshold network for instance segmentation," 2021, *arXiv:2106.03450*.
- [20] R. T. Dabou, I. Kamwa, J. Tagoudjeu, and F. C. Mugombozi, "Sparse signal reconstruction on fixed and adaptive supervised dictionary learning for transient stability assessment," *Energies*, vol. 14, no. 23, p. 7995, Nov. 2021.
- [21] S. Shao, L. Xing, W. Yu, R. Xu, Y.-J. Wang, and B.-D. Liu, "SSDL: Self-supervised dictionary learning," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shenzhen, China, Jul. 2021, pp. 1–6.
- [22] X. Song, Y. Chen, Z.-H. Feng, G. Hu, T. Zhang, and X.-J. Wu, "Collaborative representation based face classification exploiting block weighted LBP and analysis dictionary learning," *Pattern Recognit.*, vol. 88, pp. 127–138, Apr. 2019.
- [23] A.-K. Seghouane, A. Iqbal, and K. Abed-Meraim, "A sequential block-structured dictionary learning algorithm for block sparse representations," *IEEE Trans. Comput. Imag.*, vol. 5, no. 2, pp. 228–239, Jun. 2019.
- [24] N. Kumar and R. Sinha, "Improved structured dictionary learning via correlation and class based block formation," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5082–5095, Oct. 2018.
- [25] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.
- [26] Y. Huang, Y. Quan, T. Liu, and Y. Xu, "Exploiting label consistency in structured sparse representation for classification," *Neural Comput. Appl.*, vol. 31, no. 11, pp. 6509–6520, Apr. 2018.
- [27] M. Liao and X. Gu, "Face recognition approach by subspace extended sparse representation and discriminative feature learning," *Neurocomputing*, vol. 373, pp. 35–49, Jan. 2020.
- [28] S.-J. Kim and G. B. Giannakis, "Cognitive radio spectrum prediction using dictionary learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 3206–3211.
- [29] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 4, pp. 648–664, Dec. 2018.
- [30] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.
- [31] Q. Bai, J. Wang, Y. Zhang, and J. Song, "Deep learning-based channel estimation algorithm over time selective fading channels," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 1, pp. 125–134, Mar. 2020.
- [32] Y. Cui, W. Xu, Y. Tian, and J. Lin, "Perturbed block orthogonal matching pursuit," *Electron. Lett.*, vol. 54, no. 22, pp. 1300–1302, Nov. 2018.
- [33] X. Zhang, W. Xu, J. Lin, and Y. Dang, "Block normalized iterative hard thresholding algorithm for compressed sensing," *Electron. Lett.*, vol. 55, no. 17, pp. 957–959, Aug. 2019.
- [34] X. Zhang, W. Xu, Y. Cui, L. Lu, and J. Lin, "On recovery of block sparse signals via block compressive sampling matching pursuit," *IEEE Access*, vol. 7, pp. 175554–175563, 2019.
- [35] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [36] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
- [37] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [38] L. Lu, W. Xu, Y. Wang, and Z. Tian, "Recovery conditions of sparse signals using orthogonal least squares-type algorithms," *IEEE Trans. Signal Process.*, unpublished.
- [39] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [40] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.

- [41] Z. Byambadorj, K. Asami, T. J. Yamaguchi, A. Higo, M. Fujita, and T. Iizuka, "High-precision sub-Nyquist sampling system based on modulated wideband converter for communication device testing," *IEEE Trans. Circuits Syst. I, Reg. Paper*, vol. 69, no. 1, pp. 378–388, Jan. 2022.
- [42] T. Vlačić and D. Seršić, "Sampling and reconstruction of sparse signals in shift-invariant spaces: Generalized Shannon's theorem meets compressive sensing," *IEEE Trans. Signal Process.*, vol. 70, pp. 438–451, Jan. 2022.
- [43] W. Xu, Y. Cui, Y. Wang, S. Wang, and J. Lin, "A hardware implementation of random demodulation analog-to-information converter," *IEICE Electron. Exp.*, vol. 13, no. 16, 2016, Art. no. 20160465.
- [44] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [45] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [46] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Math. Program.*, vol. 142, nos. 1–2, pp. 397–434, Dec. 2013.
- [47] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.
- [48] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [49] J. J. Meng, W. Yin, H. Li, E. Hossain, and Z. Han, "Collaborative spectrum sensing from sparse observations in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 327–337, Feb. 2011.
- [50] J. Wang and P. Li, "Recovery of sparse signals using multiple orthogonal least squares," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2049–2062, Apr. 2017.
- [51] F. Chiti, R. Fantacci, D. Marabissi, and A. Tani, "Correlation-based spectrum sensing with oversampling and optimal weights selection for OFDM-based networks coexistence in TVWS," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 511–523, Jun. 2021.
- [52] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2014.
- [53] K. Tang, Z. Su, Y. Liu, W. Jiang, J. Zhang, and X. Sun, "Subspace segmentation with a large number of subspaces using infinity norm minimization," *Pattern Recognit.*, vol. 89, pp. 45–54, May 2019.



Liyang Lu (Graduate Student Member, IEEE) received the B.S. degree in communication engineering from the Beijing University of Posts and Telecommunications, China, in 2017, where he is currently pursuing the Ph.D. degree in information and communication engineering. His area of research includes compressive sensing, cognitive radios, integrated sensing and communication, sparse representation-based classification, and signal optimization.



Wenbo Xu (Member, IEEE) received the B.S. degree from the School of Information Engineering, Beijing University of Posts and Telecommunications (BUPT), China, in 2005, and the Ph.D. degree from School of Information and Communication Engineering, BUPT, in 2010. Since 2010, she has been with BUPT, where she is currently a Professor with the School of Artificial Intelligence. Her current research interests include sparse signal processing, machine learning, and signal processing in wireless networks.



Yue Wang (Senior Member, IEEE) received the Ph.D. degree in communication and information system from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2011. He is currently a Research Assistant Professor with Electrical and Computer Engineering Department, George Mason University, Fairfax, VA, USA, where he was a Postdoctoral Researcher. Prior to that, he was a Senior Engineer with Huawei Technologies Company Ltd., China. From 2009 to 2011, he was a visiting Ph.D. student with Electrical and Computer Engineering Department, Michigan Technological University, Houghton, MI, USA. His general interests include signal processing, wireless communications, machine learning, and their applications in cyber physical systems. His specific research focuses on compressive sensing, massive MIMO, millimeter-wave communications, cognitive radios, DoA estimation, high-dimensional data analysis, and distributed optimization and learning.



Zhi Tian (Fellow, IEEE) is currently a Professor with the Electrical and Computer Engineering Department, George Mason University, Fairfax, VA, USA, since 2015. Prior to that, she was on the Faculty of Michigan Technological University, Houghton, MI, USA, from 2000 to 2014. She served as a Program Director with U.S. National Science Foundation from 2012 to 2014. Her research interest lies in the areas of statistical signal processing, wireless communications, and estimation and detection theory. Her current research focuses on compressed sensing for random processes, statistical inference of network data, distributed network optimization and learning, and millimeter-wave communications. She received the IEEE Communications Society TCCN Publication Award in 2018. She was the Chair of the IEEE Signal Processing Society Big Data Special Interest Group and the General Co-Chair of the 2016 IEEE GlobalSIP Conference, and is the Unclassified Technical Program Co-Chair of the 2022 IEEE MILCOM Conference. She was a Member-at-Large of the Board of Governors of the IEEE Signal Processing Society from 2019 to 2021. She was an IEEE Distinguished Lecturer for both the IEEE Communications Society and the IEEE Vehicular Technology Society. She served an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON SIGNAL PROCESSING.