




Low-Complexity Subspace-Aided Compressive Spectrum Sensing Over Wideband Whitespace

Haoran Qi , *Student Member, IEEE*, Xingjian Zhang , *Student Member, IEEE*,
and Yue Gao , *Senior Member, IEEE*

Abstract—Compressive sensing (CS) techniques have been proposed for wideband spectrum sensing applications to achieve sub-Nyquist-rate sampling. The complexity of CS recovery algorithm and the detection performance against noise are two of the main challenges of the implementation of compressive spectrum sensing (CSS). Greedy algorithms have been of particular interest in CSS due to low complexity. We firstly propose a novel spectrum sparsity estimation scheme directly from sub-Nyquist measurements, with which the computational effort of greedy pursuit algorithms can be saved and recovery performance improved. Besides, the spectrum sparsity estimates also enable hard detection of channel occupancy where threshold adaption for energy detection is avoided. Moreover, with the detected dimension of signal subspace, we propose to implement joint-block-sparse multiple-measurement-vector (MMV) model of CSS whose dimension can be reduced to minimum and meanwhile a large portion of noise is removed. The proposed MMV model with noise and dimension reduction further improves the detection performance and also keeps the complexity low. Finally, we generalize the hard thresholding pursuit (HTP) algorithm to recover joint-block-sparse signals. In simulations, the detection performance and complexity of the proposed CSS scheme show striking superiority against multiple benchmarking schemes.

Index Terms—Compressive spectrum sensing, hard thresholding pursuit, spectrum sparsity, subspace decomposition, hard detection.

I. INTRODUCTION

THE rapid evolution of wireless communications has demanded increasingly large data rate and service coverage, and the spectrum scarcity appears to be a major challenge of wireless communication applications. Spectrum sensing is the essential and proactive way to acquire the ambient spectrum availability information, which enables dynamic spectrum access (DSA) demanded by unlicensed secondary users (SUs) to the spectrum, while guaranteeing the quality of service of licensed primary users (PUs) [1]. While Nyquist-rate sampling is required in the conventional spectrum sensing, compressive sensing (CS) techniques have been proposed in wideband

spectrum sensing to achieve sub-Nyquist-rate sampling, which alleviates the hardware and power constraints of the conventional Nyquist-rate processing [2]–[6]. Cooperative spectrum sensing, which requires data fusion and centralized decision-making, has been proposed to mitigate the deteriorated detection performance in deep fading channel environments utilizing the SU's diversity of power spectrum [6]–[9]. A real-world practice of DSA in UK is TV white-space (TVWS) where the allocated spectrum of 470 – 790MHz has been largely vacated by the massive introduction of digital terrestrial TV (DTTV) [10]–[12]. TVWS featuring wide bandwidth and sparse spectrum usage by PUs is particularly suitable for the implementation of CSS techniques [13]. With the evolution of communication system towards the fifth-generation, recent proposals are present that cellular services at millimeter-wave (mmWave) bands, with multiple operators potentially servicing overlapping small areas and coexisting with other crucial wireless services, could employ a spectrum sharing and DSA model [14]. Real-time spectrum sensing for such mmWave DSA applications, where the bandwidth can span multiple gigahertz, poses larger challenges on the sampling rate, and sub-Nyquist techniques could be promising candidates.

Considering the demand of DSA applications, rapid spectrum reconstruction and robust spectrum availability detection are desired in compressive spectrum sensing (CSS). Optimization methods in CSS have been discussed in literature, including convex l_1 optimization [4], [5], non-convex l_v ($0 < v < 1$) optimization [15], and sparse Bayesian learning [16], [17]. However, the complexity of CS recovery algorithms, especially these optimization methods, has been one of the major bottlenecks of the implementation of CSS applications. Among these algorithms, greedy algorithms, especially the Hard Thresholding Pursuit (HTP) [18] and its derivations, have drawn great attentions due to their lower computational cost compared to these optimization-based sparse recovery schemes [19]. Most greedy algorithms for sparse recovery require the sparsity of the signal to be recovered as a priori input information. Inputting an excessively large sparsity tends to lead to unnecessary free dimensions of the signal space, which does not help detect the support of the desired signal and makes the recovery fidelity worse. However, in practice, valid estimates of sparsity in a dynamic scenario are difficult to obtain, hence the *a priori* known upper bound of sparsity based on long-time observations or the maximum allowed value of sparsity is often input to these algorithms to ensure reliable spectrum reconstruction.

Manuscript received October 19, 2018; revised April 17, 2019 and June 28, 2019; accepted July 28, 2019. Date of publication October 7, 2019; date of current version December 17, 2019. This work was presented at IEEE Global Conference on Communications, December 9–13, 2018, Abu Dhabi, U.A.E. The review of this article was coordinated by Dr. X. Huang. (*Corresponding author: Haoran Qi.*)

The authors are with the Department of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: h.qi@qmul.ac.uk; xingjian.zhang@qmul.ac.uk; yue.gao@qmul.ac.uk).

Digital Object Identifier 10.1109/TVT.2019.2937649

Another simple solution is to directly use the channel sparsity information from the geo-location database. However, such information could be inaccurate as geo-location database relies on propagation model and is designed to give conservative channel availability to protect PUs [12], [20]. Some spectrum sparsity estimation solutions have also seen in literature [21], [22]. In these works, prior to the actual spectrum sensing step, the spectrum sparsity is estimated by multiple exhaustive pilots of recovery implementations, which makes it inefficient. In [23], a novel metric for sparsity is proposed for noisy signals, and estimation is performed directly from compressive measurements. However, the measurements are required to be sampled by various measurement matrices of a special formation. Moreover, the sparsity information in [21]–[23] is proposed to be applied to convex-optimization and adaptively optimize the compression ratio, which is fundamentally different from how the sparsity estimation is proposed to be used in this paper.

Robust detection of spectrum availability is also a crucial procedure in CSS. The simple yet commonly adopted method is Neyman-Pearson (NP) energy detection (ED) on each channel's spectrum energy, which determines the channel occupancy in a soft decision manner by setting a proper threshold [24]. Some work in CSS proposes cyclostationary feature extraction on the recovered spectrum and then perform ED [17], [25]. Some prior information on the noise statistics is essential to optimal threshold setting. Practical noise estimation and threshold setting schemes for conventional non-compressive spectrum sensing have been proposed in literature [26], [27]. However, it is discovered in [28] that the statistic model of the channel energy based on the recovered spectrum differs from that of the conventional model, and parameters of the proposed statistical models have to be learned from multiple recovered spectrum samples in order to set the proper threshold. The proposed parameter learning scheme also has the realistic shortcoming that the estimation requires time-consuming recovery operations.

To address the sparsity estimation issue, we propose a novel sparsity estimation directly from time-domain sub-Nyquist samples. The proposed sparsity estimation is achieved by the subspace analysis of the auto-correlation matrix of sub-Nyquist samples. In the area of array processing, a similar problem is source enumeration. The full-row-rank “measurement matrix” in the CSS context is an analogy of the full-column-rank “array steering matrix,” and the spectrum sparsity is actually equivalent to the number of sources in array processing. Source enumeration problem has been being studied for decades, with some information theoretic criteria (ITC) proposed [29], [30] to select the optimal number of principle and noise subspace dimensions in the sense of information distance through eigen-decomposition, including Akaike's information criterion [31] and minimum description length (MDL) [32] (or equivalently Bayesian Information Criterion (BIC) in Bayesian formulation [33]). BIC has drawn particular interests due to its better asymptotic performances [34]. We choose to use an enhanced BIC strategy to achieve the spectrum sparsity estimation. A pioneering subspace-aided sparsity estimation strategy has been recently outlined in [6]. However, the specific estimation methods

applied, estimation performance or the advantages of spectrum sparsity estimates have not been discussed.

Since the proposed spectrum sparsity estimation requires multiple samples of sub-Nyquist measurement vectors, we further propose to use multiple-measurement-vector (MMV) model for joint-block-sparse signals for CS directly instead of single-measurement-vector (SMV) model. Based on the subspace decomposition of the sparsity estimation scheme, it is natural to reduce the dimensionality of the MMV model to its minimum by removing the noise subspace, which effectively achieves denoising from the measurements. Finally, we extend the HTP algorithm for recovering joint-block-sparse signals, which empirically gives better detection performance than l_1 optimization and orthogonal matching pursuit (OMP) counterparts while keeping the complexity low.

The merits of the proposed spectrum sparsity estimation scheme and the implementation of dimension-reduced MMV model are summarized as follows:

- The estimation of spectrum sparsity is directly from compressed measurements, which does not require recovery operations;
- A valid estimate of spectrum sparsity aids the greedy pursuit algorithm to reduce the complexity and enhance the recovery performance;
- Dimension reduction of the MMV model further improves the recovery performance and reduces recovery complexity.
- With a valid estimate of spectrum sparsity, hard detection (HD) of channel occupancy can be achieved without the need of dictating or adapting the threshold.

The rest of this paper is structured as follows. In Section II, system and signal models of CSS are presented. In Section III, the proposed spectrum sparsity and active channel number estimation scheme are detailed. In Section IV, the dimension reduction of the MMV model and noise removal is introduced. Then, performances of various commonly-used greedy algorithms are discussed and the HTP algorithm is generalized for joint-block-sparse signals. Monte-Carlo simulation results and further comments are given in Section V.

II. SYSTEM AND SIGNAL MODELS

A. System Model

Consider a cognitive radio system with primary users (PUs) being imaginary transceiver stations that transmits orthogonal frequency division multiplexing (OFDM) signals within C frequency channels which have equal bandwidth of B . In the upper half in Fig. 1, the model diagram of an OFDM-based radio transmitter is illustrated. The baseband signal from the c th transmitter is to be upconverted by $F_0 + (c - 1)B$ to the c th channel of the interested wideband spectrum, of which the frequency range is $[F_0, F_0 + CB]$. The OFDM system of each PU transmitter is assumed to have the symbol rate of F_s , hence the minimum frequency spacing of OFDM sub-carriers is F_s . Then, the number of sub-carriers for each PU transmitter is determined by B/F_s , which is assumed to be an integer. An

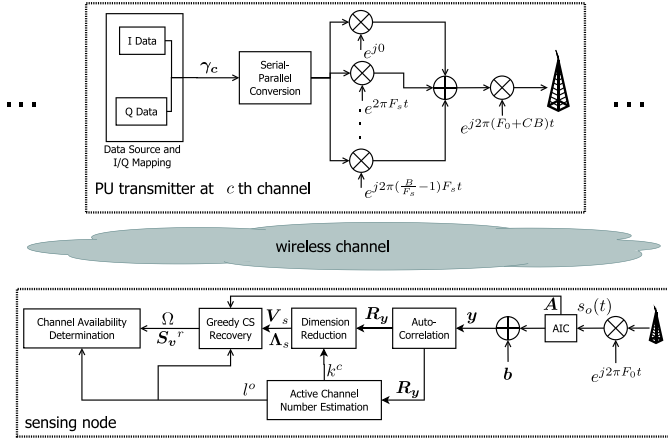


Fig. 1. The proposed system architecture: PU transmitters and a sensing node.

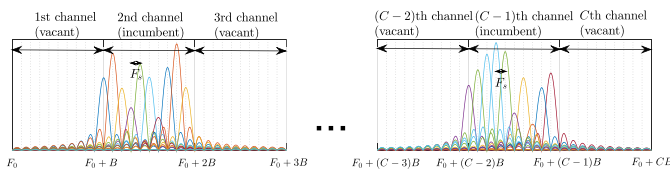


Fig. 2. A conceptual illustration of the interested TVWS wideband spectrum.

illustration of wideband OFDM spectrum with the parameters specified above is given in Fig. 2.

B. Signal Model

Here, we construct the model of the signal to be sensed. Firstly, with accordance to the PU transmitter diagram in Fig. 1, we write the continuous time-domain signal of OFDM transmitters within one symbol duration as

$$s_o(t) = \sum_{c=1}^C \sum_{d=1}^{\frac{B}{F_s}} (a_{c,d} + jb_{c,d}) e^{j2\pi[(d-1)F_s + (c-1)B]t}, \quad (1)$$

where $0 \leq t < (1/F_s)$. $a_{c,d}$ and $b_{c,d}$ represents the d th data symbol from the c th PU, of in-phase and quadrature component respectively. Practical implementations of OFDM are digital via inverse DFT (IDFT) of B/F_s points, and for a wideband system comprising C channel the Nyquist frequency is $f_{\text{NYQ}} = CB$. The discrete time-domain signal (\mathbf{s}_o) $_{\frac{CB}{F_s} \times 1}$ is equivalent to the continuous version (1) sampled at Nyquist frequency f_{NYQ} , that is

$$\mathbf{s}_o = \left[s_o(0) \ s_o\left(\frac{1}{CB}\right) \ \cdots \ s_o\left(\frac{CB-1}{CB}\right) \right]^T = \mathbf{G}\boldsymbol{\gamma}. \quad (2)$$

where $\mathbf{G}_{\frac{CB}{F_s} \times \frac{CB}{F_s}}$ is effectively a $\frac{CB}{F_s}$ -point IDFT matrix. $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1^T \ \boldsymbol{\gamma}_2^T \ \cdots \ \boldsymbol{\gamma}_C^T]^T$, where data symbol blocks ($\boldsymbol{\gamma}_c$) $_{\frac{B}{F_s} \times 1}$ ($c = 1, 2, \dots, C$) are presented by

$$\boldsymbol{\gamma}_c = \left[a_{c,1} + jb_{c,1} \ a_{c,2} + jb_{c,2} \ \cdots \ a_{c,B/F_s} + jb_{c,B/F_s} \right]^T. \quad (3)$$

In order to express the sparse channel occupation, we formally model that for some $1 \leq c \leq C$, $\boldsymbol{\gamma}_c$ are vectors with independent zero-mean complex random entries, which represents occupied channels; and for other $\boldsymbol{\gamma}_c$'s, $\boldsymbol{\gamma}_c = \mathbf{0}$ which models the vacant channels. The frequency representation (i.e. the DFT) of \mathbf{s}_o is effectively $\boldsymbol{\gamma} = \mathbf{G}^{-1}\mathbf{s}_o = \mathbf{G}^{-1}\mathbf{G}\boldsymbol{\gamma}$. On the receiver's end, the received frequency-domain signal through a general multipath channel is $\hat{\boldsymbol{\gamma}} = [\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \dots, \hat{\boldsymbol{\gamma}}_C]$, where

$$\hat{\boldsymbol{\gamma}}_c = L_c \left[\frac{1}{\sqrt{K_c + 1}} + \frac{\sqrt{K_c}}{\sqrt{K_c + 1}} \cdot \mathbf{h}_c \right] \circ \boldsymbol{\gamma}_c, \quad 1 \leq c \leq C. \quad (4)$$

\circ denotes element-wise product. L_c 's are the path loss from the transmitter occupying the c th channel and the receiver. \mathbf{h}_c 's are multipath channel gain where $\mathbf{h}_c \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. K_c is the power ratio between multipath and line-of-sight components. Denote the time-domain representation of the received signal by $\hat{\mathbf{s}}_o := \mathbf{G}^{-1}\hat{\boldsymbol{\gamma}}$.

Consider a simple single-node non-cooperative CSS scenario. Denote the Nyquist-rate time-domain signal as $\mathbf{s}_t_{N \times 1} = [s_t[0] \ s_t[1] \ \cdots \ s_t[N-1]]^T$ and its frequency-domain representation $\mathbf{s}_f = \mathcal{F}\mathbf{s}_t = [s_f[0] \ s_f[1] \ \cdots \ s_f[N-1]]^T$ where $\mathcal{F}_{N \times N}$ stands for the N -point discrete Fourier transform (DFT) matrix. The signal \mathbf{s}_f with sparsity, denoted by $\delta(\mathbf{s}_f) := \|\mathbf{s}_f\|_0$, can be recovered from a M -element sub-Nyquist-rate measurement vector $\mathbf{y}_{M \times 1}$ where $M < N$, with the product of \mathbf{A} and \mathcal{F}^{-1} meeting restricted isotropic property [19]. The sub-Nyquist-rate sampling can be characterized by a linear system

$$\mathbf{y} = \mathbf{A}\mathbf{s}_t + \mathbf{b} = \mathbf{A}\mathcal{F}^{-1}\mathbf{s}_f + \mathbf{b}, \quad (5)$$

where $\mathbf{A}_{M \times N}$ is the sampling matrix with structured random entries corresponding to analog-to-information converter (AIC) settings [35], and $\mathbf{b}_{M \times 1} \sim \mathcal{CN}(\mathbf{0}_{M \times 1}, \sigma^2 \mathbf{I}_M)$ is a vector of independent and identically distributed (i.i.d.) complex additive white Gaussian noise over measurements. Recovering a sparse \mathbf{s}_f from the SMV model with additive noise (5) is a typical problem in the CS area. Denote the recovered signal as \mathbf{s}_f^r . SNR is formally defined by

$$\text{SNR} := \|\mathbf{A}\mathbf{s}_t\|_2^2 / (M\sigma^2). \quad (6)$$

Consider that the practically large number of points $\frac{CB}{F_s}$ of the spectrum at receiver $\hat{\boldsymbol{\gamma}}$ makes it unnecessary and adds to the complexity for CS. Here, we consider the decimated version of the spectrum as the sparse spectrum to be recovered \mathbf{s}_f in (5), writing

$$s_f[n] := \hat{\boldsymbol{\gamma}} \left[n \cdot \frac{CB}{F_s N} \right], \quad n = 0, 1, \dots, N-1, \quad (7)$$

where it is assumed N is chosen to make $\frac{B}{F_s N}$ an integer. It is straightforward from (7) that $\delta(\mathbf{s}_f)$ non-zeros entries in \mathbf{s}_f are independent zero-mean complex random variables. The Nyquist-rate time-domain samples \mathbf{s}_t in the CS model (5) is then

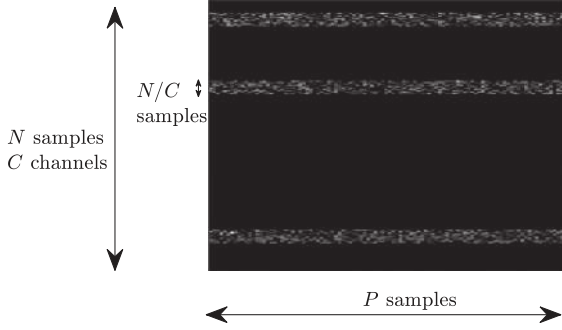


Fig. 3. A visualization of the joint-block-sparse matrix \mathbf{S}_f .

related to $\hat{\mathbf{s}}_o$ by

$$\begin{aligned}
 s_t[n] &= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} s_f[i] \cdot e^{j2\pi \frac{in}{N}} \\
 &= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \left[\sum_{m=0}^{\frac{CB}{F_s} - 1} \hat{s}_o[m] \cdot e^{-j2\pi \frac{m(i - \frac{CB}{F_s N}) F_s}{CB}} \right] e^{j2\pi \frac{in}{N}} \\
 &= \frac{1}{\sqrt{N}} \sum_{m=0}^{\frac{CB}{F_s} - 1} \hat{s}_o[m] \left[\sum_{i=0}^{N-1} e^{-j2\pi \frac{i(m-n)}{N}} \right] \\
 &= \sqrt{N} \sum_{m=0}^{\frac{CB}{NF_s} - 1} \hat{s}_o[mN + n], \quad n = 0, 1, \dots, N-1.
 \end{aligned} \tag{8}$$

III. SPECTRUM SPARSITY ESTIMATION BASED ON SUBSPACE DECOMPOSITION

Suppose P samples of measurement vectors are obtained from sub-Nyquist sampling, expressed by

$$\begin{aligned}
 \mathbf{Y} &= [\mathbf{y}^{(1)} \mathbf{y}^{(2)} \dots \mathbf{y}^{(P)}] \\
 &= \mathbf{A}\mathcal{F}^{-1} [\mathbf{s}_f^{(1)} \mathbf{s}_f^{(2)} \dots \mathbf{s}_f^{(P)}] + [\mathbf{b}^{(1)} \mathbf{b}^{(2)} \dots \mathbf{b}^{(P)}] \tag{9} \\
 &= \mathbf{A}\mathcal{F}^{-1} \mathbf{S}_f + \mathbf{B} = \mathbf{Y}_s + \mathbf{B}.
 \end{aligned}$$

The underdetermined system (9) consists a MMV model in CS literature, where the spectrum support can be determined after recovering the joint row-sparse matrix $(\mathbf{S}_f)_{N \times P}$. It is hereby assumed that the channel occupancy remains the same in the period of sample acquisition P/F_s , hence \mathbf{S}_f is joint row-sparse. Apart from the joint row-sparse property of \mathbf{S}_f , it also features a block-sparse property due to the channel structure, where every single $(\frac{N}{C} \times P)$ -element block responsible to the same channel have the same sparsity pattern - occupied or vacant. We refer the sparsity pattern of \mathbf{S}_f as “joint-block-sparse” as exemplified in Fig. 3 and the term “channel” and “block” are used interchangeably in the remaining of this paper.

Examine the auto-correlation matrix of sub-Nyquist rate measurement vectors,

$$\mathbf{R}_y = \mathbf{A}\mathcal{F}^{-1} \mathbf{R}_{s_f} (\mathcal{F}^{-1})^H \mathbf{A}^H + \sigma^2 \mathbf{I}_M, \tag{10}$$

where $\mathbf{R}_y = \mathbb{E}[\mathbf{y}\mathbf{y}^H] = \lim_{P \rightarrow \infty} \frac{1}{P} \sum_{p=1}^P \mathbf{y}^{(p)} (\mathbf{y}^{(p)})^H$ and $\mathbf{R}_{s_f} = \mathbb{E}[\mathbf{s}_f \mathbf{s}_f^H] = \lim_{P \rightarrow \infty} \frac{1}{P} \sum_{p=1}^P \mathbf{s}_f^{(p)} (\mathbf{s}_f^{(p)})^H$. With finite sample size P , one can only have the estimation of auto-correlation matrices \mathbf{R}_y and \mathbf{R}_{s_f} denoted by $\hat{\mathbf{R}}_y$ and $\hat{\mathbf{R}}_{s_f}$, where it holds

$$\begin{aligned}
 \hat{\mathbf{R}}_y &= \mathbf{A}\mathcal{F}^{-1} \hat{\mathbf{R}}_{s_f} (\mathcal{F}^{-1})^H \mathbf{A}^H + \frac{1}{P} \mathbf{B}\mathbf{B}^H \\
 &= \frac{1}{P} \mathbf{Y}_s \mathbf{Y}_s^H + \frac{1}{P} \mathbf{B}\mathbf{B}^H,
 \end{aligned} \tag{11}$$

where $\hat{\mathbf{R}}_y = \frac{1}{P} \sum_{p=1}^P \mathbf{y}^{(p)} (\mathbf{y}^{(p)})^H = \frac{1}{P} \mathbf{Y}\mathbf{Y}^H$, and $\hat{\mathbf{R}}_{s_f} = \frac{1}{P} \sum_{p=1}^P \mathbf{s}_f^{(p)} (\mathbf{s}_f^{(p)})^H = \frac{1}{P} \mathbf{S}_f \mathbf{S}_f^H$. In Proposition 1 that follows, we present the property on the rank of $\hat{\mathbf{R}}_{s_f}$ and $\hat{\mathbf{R}}_y$ under noise-free conditions, based on which the sparsity of spectrum is proposed to be estimated.

Lemma 1: Given $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(s)} \in \mathbb{C}^r$ is a series of arbitrary vectors and $s < r$, a random vector $\boldsymbol{\alpha}^{(s+1)} \in \mathbb{C}^r$ whose entries are independent and follow an absolutely continuous distribution makes $\dim\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(s)}, \boldsymbol{\alpha}^{(s+1)}\} = \dim\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(s)}\} + 1$ with the probability of 1.

Proof: Denote the subspace dimension $\dim\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(s)}\} = v$, and naturally we have $v \leq s < r$. Consider the case that $\boldsymbol{\alpha}^{(s+1)}$ falls in the subspace $\text{span}\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(s)}\}$, where the probability equals to the integral of the probability density function that is absolutely continuous in space \mathbb{C}^r on a v -dimension hyperplane, which has the measure of 0. Hence, with the probability of 1, $\boldsymbol{\alpha}^{(s+1)}$ is linear-independent with $\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(s)}\}$, which completes the proof. ■

Proposition 1: Given $\sigma^2 = 0$, and that random vector \mathbf{s}_f has the sparsity of $\delta(\mathbf{s}_f)$, if $P \geq \delta(\mathbf{s}_f)$, it holds with probability of 1 that $\text{rank}(\hat{\mathbf{R}}_{s_f}) = \delta(\mathbf{s}_f)$. Furthermore, if $\mathbf{A}\mathcal{F}^{-1}$ has full rank and $M \geq \delta(\mathbf{s}_f)$, it holds with probability of 1 that $\text{rank}(\hat{\mathbf{R}}_y) = \text{rank}(\hat{\mathbf{R}}_{s_f}) = \delta(\mathbf{s}_f)$.

Proof: Given arbitrary $\mathbf{s}_f^{(1)}$, from Lemma 1, it holds that $\dim\{\mathbf{s}_f^{(1)} \mathbf{s}_f^{(2)}\} = 2$ with probability of 1. For $2 \leq s \leq \delta(\mathbf{s}_f) - 1$, assuming $\dim\{\mathbf{s}_f^{(1)} \dots \mathbf{s}_f^{(s)}\} = s$, again from Lemma 1, it holds that $\dim\{\mathbf{s}_f^{(1)} \dots \mathbf{s}_f^{(s+1)}\} = s + 1$. Consequently, we have first $\delta(\mathbf{s}_f)$ samples satisfying $\dim\{\mathbf{s}_f^{(1)} \mathbf{s}_f^{(2)} \dots \mathbf{s}_f^{(\delta(\mathbf{s}_f))}\} = \text{rank}[\mathbf{s}_f^{(1)} \mathbf{s}_f^{(2)} \dots \mathbf{s}_f^{(\delta(\mathbf{s}_f))}] = \delta(\mathbf{s}_f)$. Then, with $P - \delta(\mathbf{s}_f)$ more samples as columns, we naturally have $\text{rank}(\mathbf{S}_f) \geq \delta(\mathbf{s}_f)$. On the other hand, consider the sparsity of $\mathbf{s}_f^{(p)}$'s, it also holds $\text{rank}(\mathbf{S}_f) \leq \delta(\mathbf{s}_f)$. From both inequalities, we have $\text{rank}(\mathbf{S}_f) = \delta(\mathbf{s}_f)$. From $\hat{\mathbf{R}}_{s_f} = \frac{1}{P} \mathbf{S}_f \mathbf{S}_f^H$, we have $\text{rank}(\hat{\mathbf{R}}_{s_f}) = \text{rank}(\mathbf{S}_f) = \delta(\mathbf{s}_f)$.

Furthermore, consider $\mathbf{Y} = \mathbf{A}\mathcal{F}^{-1} \mathbf{S}_f$, we have $\text{rank}(\mathbf{Y}) = \min\{\text{rank}(\mathbf{A}\mathcal{F}^{-1}), \delta(\mathbf{s}_f)\} = \delta(\mathbf{s}_f)$. Finally, we reach $\text{rank}(\hat{\mathbf{R}}_y) = \text{rank}(\frac{1}{P} \mathbf{Y}\mathbf{Y}^H) = \text{rank}(\mathbf{Y}) = \delta(\mathbf{s}_f)$. ■

Proposition 1 indicates that the sparsity $\delta(\mathbf{s}_f)$ can be revealed by the rank of the estimated autocorrelation of the measurements \mathbf{R}_y under the conditions that 1) $P \geq \delta(\mathbf{s}_f)$, 2) $M \geq \delta(\mathbf{s}_f)$ and that 3) $\mathbf{A}\mathcal{F}^{-1}$ has full rank. Since CS theory requires the product of sampling matrix \mathbf{A} and sparsifying matrix \mathcal{F}^{-1} to be

nearly orthogonal [19], and the number of measurements to meet $M > \mathcal{O}(\delta(\mathbf{s}_f) \log N)$ [36], condition 2) and 3) can be naturally met in most practices of CS. Therefore, one need guarantee that the number of measurement vector samples is larger than the maximum sparsity of the spectrum in order to employ the proposed sparsity estimation scheme.

Perform eigendecomposition of $\mathbf{A}\mathcal{F}^{-1}\hat{\mathbf{R}}_{\mathbf{s}_f}(\mathcal{F}^{-1})^H\mathbf{A}^H$ which has the rank of $\delta(\mathbf{s}_f)$ from Proposition 1 and yields $\delta(\mathbf{s}_f)$ eigenvalues. With the presence of noise, $\hat{\mathbf{R}}_y$ is presented by

$$\hat{\mathbf{R}}_y = \sum_{i=1}^{\delta(\mathbf{s}_f)} \lambda_i \mathbf{v}_i \mathbf{v}_i^H + \frac{1}{P} \mathbf{B}\mathbf{B}^H, \quad (12)$$

where $\lambda_1, \dots, \lambda_{\delta(\mathbf{s}_f)}$ denote the eigenvalues in the signal's subspace spanned by corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_{\delta(\mathbf{s}_f)}$. With relatively large signal-to-noise ratio (SNR), the eigenvalue profile asymptotically converges with the increasing number of samples P to such that the dominant eigenvalues correspond to signal power and that small and equal perturbing eigenvalues $\lambda_{\delta(\mathbf{s}_f)+1} = \dots = \lambda_M = \sigma^2$ represent the noise power. In practice, we use the eigendecomposition of estimated auto-correlation matrix $\hat{\mathbf{R}}_y$

$$\hat{\mathbf{R}}_y = \sum_{i=1}^M \hat{\lambda}_i \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i^H, \quad (13)$$

where $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_M$ and $\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_M$ denote the descending-ordered eigenvalues and corresponding eigenvectors of the estimated auto-correlation matrix. Then, the sparsity can be obtained by differentiating two groups of eigenvalues by their amplitudes.

A. A Bayesian-Information-Criteria-Based Estimator

In practical estimation-based eigendecomposition (13), only limited number of samples can be used for estimation and SNR can be low, hence, the noise subspace eigenvalues are not identical and can be difficult to be discriminated from signal subspace eigenvalues. The above-described sparsity estimation problem is an analogy to the source enumeration which is a fundamental problem in array processing. A recently-proposed algorithm based on information theoretic criteria (ITC) for estimating the dimensionality of signal subspace is the enhanced Bayesian Information Criterion (BICe) [37] which defines the ITC metric as

$$\begin{aligned} \text{BICe}(k) = & -2 \log f(\mathbf{Y}|\hat{\lambda}_1, \dots, \hat{\lambda}_M; k) \\ & -2 \log f(\hat{\lambda}_1, \dots, \hat{\lambda}_M; k) + C_k \log P, \end{aligned} \quad (14)$$

$k = 1, \dots, M$, where $f(\cdot)$ denotes probability density function and $C_k = k(2M - k)$ [29] is the number of free parameters related to k . Then the estimation k^o of $\delta(\mathbf{s}_f)$ is determined by

$$k^o = \arg \min_k \text{BICe}(k). \quad (15)$$

Additionally, note \mathbf{s}_f and \mathbf{b} are independent and both zero-mean random vectors, we approximately regard \mathbf{y} as a zero-mean Gaussian random vector as a result of central limit theorem. Hence, the posterior probability of P independent observations

\mathbf{Y} , given the estimated eigenvalues parameterized by k , is expressed by the multivariate Gaussian model [38]

$$\begin{aligned} f(\mathbf{Y}|\hat{\lambda}_1, \dots, \hat{\lambda}_M; k) &= \frac{\prod_{p=1}^P \exp \text{tr} \left((\mathbf{y}^{(p)})^H \cdot \hat{\mathbf{R}}_y \cdot \mathbf{y}^{(p)} \right)}{\pi^M \cdot \prod_{i=1}^k \hat{\lambda}_i \cdot (\hat{\sigma}^2)^{M-k}} \\ &= \frac{1}{\pi^M \cdot \prod_{i=1}^k \hat{\lambda}_i \cdot (\hat{\sigma}^2)^{M-k}}, \end{aligned} \quad (16)$$

where $\hat{\lambda}_1, \dots, \hat{\lambda}_k$ and $\hat{\sigma}^2 := \sum_{i=k+1}^M \hat{\lambda}_i / (M - k)$ are maximum-likelihood (ML) estimates for $\lambda_1, \dots, \lambda_k$ and σ^2 , respectively.

The joint probability density $f(\hat{\lambda}_1, \dots, \hat{\lambda}_M; k)$ is proposed [37] to be approximated by the product of the probability density of signal subspace eigenvalues and the probability density of noise subspace eigenvalues, using the ML estimates to approximate true values in conditions, that is

$$\begin{aligned} f(\hat{\lambda}_1, \dots, \hat{\lambda}_M; k) &= f(\hat{\lambda}_1, \dots, \hat{\lambda}_k | \lambda_1 = \hat{\lambda}_1, \dots, \lambda_k = \hat{\lambda}_k) \\ &\quad \cdot f(\hat{\lambda}_{k+1}, \dots, \hat{\lambda}_M | \sigma^2 = \hat{\sigma}^2). \end{aligned} \quad (17)$$

The probability densities $f(\hat{\lambda}_1, \dots, \hat{\lambda}_k | \lambda_1, \dots, \lambda_k)$ and $f(\hat{\lambda}_{k+1}, \dots, \hat{\lambda}_M | \sigma^2)$ are derived in [38] and [34] respectively. By relating (16) and (17) into (14), after terms irrelevant to k being removed, the optimization of BICe can be rearranged as the following

$$\begin{aligned} k^o &= \arg \min_k \text{BICe}(k) \\ &= \arg \min_k \{ 2(M - k)(P + M - k) \log \hat{\sigma}^2 \\ &\quad - 2P \sum_{i=k+1}^M \log \hat{\lambda}_i + 2 \sum_{i=1}^k \log \hat{\lambda}_i + 2 \sum_{i=1}^{M-k} \log \Gamma(i) \\ &\quad + P \sum_{i=k+1}^M \left(\frac{\hat{\lambda}_i - \hat{\sigma}^2}{\hat{\sigma}^2} \right)^2 - 4 \sum_{i=k+1}^M \sum_{j=i+1}^M \log |\hat{\lambda}_i - \hat{\lambda}_j| \\ &\quad + (4Mk - 2k^2 - k) \log P \}, \end{aligned} \quad (18)$$

where $\Gamma(\cdot)$ denotes the real Gamma function.

B. Initial Performance Analysis

As a generalization of the original BIC detector, the BICe-based detector is reported to inherit *asymptotic consistency property* [37] that both the overestimation and underestimation probability converge to zero with P increasing to infinity, regardless the SNR. Furthermore, overestimation probability converges much more rapidly, and is reportedly small enough empirically and analytically [39] to be neglected.

The eigenvalue profiles of varying SNR and P are firstly shown in Fig. 4(a) and (b) respectively. With decreasing SNR, it is a natural result that the increasing eigenvalues from noise subspace lead to more difficulty in discriminating the signal subspace eigenvalues from the eigenvalue profile. With decreasing P , the number of noise subspace eigenvalues are also harder to be determined as the noise subspace eigenvalues tend to be more inconsistent in the eigenvalue profile. Moreover, we illustrate

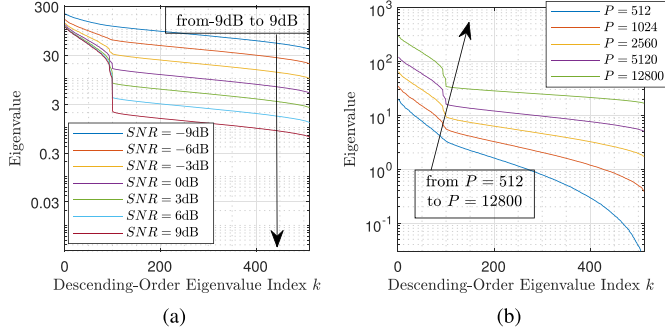


Fig. 4. Descending-ordered eigenvalue profiles: (a) with varying SNR from -9 dB to 9 dB and $P = 5120$; (b) with varying P and $SNR = 0$ dB.

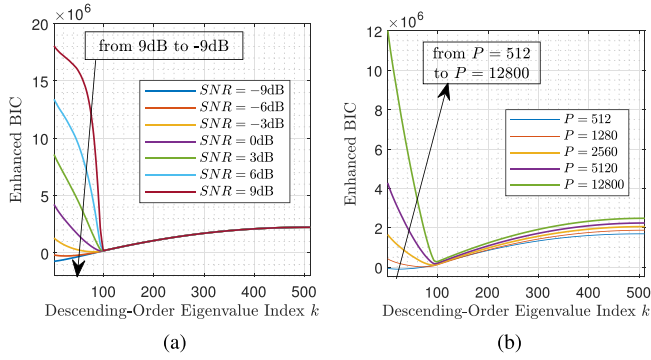


Fig. 5. Enhanced BIC metric against sparsity values k : (a) with varying SNR from -9 dB to 9 dB and $P = 5120$; (b) with varying P and $SNR = 0$ dB.

the BICe metric against $1 \leq k \leq M$ with changing SNR and P in Fig. 5(a) and (b). With SNR and P above certain level, the proposed BICe-based detection method can accurately detect the sparsity of spectrum $k^o = \text{rank}(\mathbf{R}_{s_f}) = \delta(\mathbf{s}_f)$. However, with either further decreasing SNR or P , the k values corresponding to minimum BICe metrics tend to be underestimated, that is $k^o < \delta(\mathbf{s}_f)$. Additionally, in Fig. 5, overestimation is never spotted with finite times of simulations and the smallest eligible $P = M = 512$. In this sense, we can reasonably match $P = 512$ to the qualitative description “large sample size” in terms of negligible overestimation probability.

C. Estimation of Active Channel Number

As the sparse presentation \mathbf{S}_f is a joint-block-sparse matrix, having estimated the sparsity of the spectrum, one finally needs the estimated active channel number as the number of blocks to be detected in the recovery algorithm. We hereby propose the estimation of active channel number based on the constraint that the dimension of signal subspace only can be multiples of the number of spectrum bins of each channel, which is $\delta(\mathbf{s}_f) = l \cdot N/C$ with the actual number of occupied channels $0 \leq l \leq C$ being some integer. Hence, with negligible overestimation probability, we estimate the active channel number l^o from the detected sparsity of spectrum by

$$l^o := \left\lceil \frac{k^o \cdot C}{N} \right\rceil. \quad (19)$$

This ceiling operation in active channel number detection effectively enhances the its robustness against low SNR and P in which cases the spectrum sparsity are possibly underestimated, by amending the underestimated spectrum sparsity $\delta(\mathbf{s}_f) - N/C < k^o < \delta(\mathbf{s}_f)$ as the case where the number of active channel is correctly detected. With the estimated active channel number l^o , one can get the compensated version of spectrum sparsity as

$$k^c := \frac{N \cdot l^o}{C} = \frac{N \cdot \left\lceil \frac{k^o \cdot C}{N} \right\rceil}{C}. \quad (20)$$

IV. LOW-COMPLEXITY CSS SCHEME WITH ESTIMATED ACTIVE CHANNEL NUMBER

A. Noise and Dimension Reduction Based on Subspace Decomposition

The introduction of MMV model (9) can potentially improve the detection performance as the model includes multiple samples of random measurement vectors compared to SMV case (5). From the evaluations in Section III-C, the proposed channel sparsity estimation may require a large number of measurement vectors P to achieve satisfactory performance. To enable quick spectrum sensing updates and rapid spectrum access, one could use a smaller number of measurement vectors, i.e. $P' < P$ to form the following MMV model

$$\mathbf{Y}' = \mathbf{A}\mathbf{F}^{-1}\mathbf{S}'_f + \mathbf{B}' = \mathbf{Y}'_s + \mathbf{B}', \quad (21)$$

where $\mathbf{Y}' = [\mathbf{y}'^{(1)} \mathbf{y}'^{(2)} \dots \mathbf{y}'^{(P')}]$, $\mathbf{S}'_f = [\mathbf{s}_f^{(1)} \mathbf{s}_f^{(2)} \dots \mathbf{s}_f^{(P')}]$, and $\mathbf{B}' = [\mathbf{b}'^{(1)} \mathbf{b}'^{(2)} \dots \mathbf{b}'^{(P')}]$. The sample acquisition period for (21) is then P'/Fs .

However, noting that the sparse recovery's complexity via an MMV model goes linearly with the number of measurement vectors P' [6], sparse recovery via an MMV model (21) with a large P' becomes impractical. Besides that the subspace-aided approach in Section V provides the estimate of spectrum sparsity, we hereby note that by a similar subspace decomposition operation one can also reduce a great portion of noise from the measurements and also the dimension of CS recovery problem. Similar to (13), we have the estimated autocorrelation of \mathbf{Y}' , which writes $\hat{\mathbf{R}}'_y := \frac{1}{P'} \sum_{p=1}^{P'} \mathbf{y}'^{(p)} (\mathbf{y}'^{(p)})^H = \frac{1}{P'} \mathbf{Y}' \mathbf{Y}'^H$, and perform eigendecomposition of $\hat{\mathbf{R}}'_y$. With the compensated sparsity estimate k^c , one can separate the subspace of $\hat{\mathbf{R}}'_y$ into signal and noise subspace

$$\begin{aligned} \hat{\mathbf{R}}'_y &= \sum_{i=1}^{k^c} \lambda'_i \mathbf{v}'_i \mathbf{v}'_i{}^H + \sum_{i=k^c+1}^{\text{rank } \hat{\mathbf{R}}'_y} \lambda'_i \mathbf{v}'_i \mathbf{v}'_i{}^H \\ &= \mathbf{V}_s \mathbf{\Lambda}_s \mathbf{V}_s^H + \mathbf{V}_n \mathbf{\Lambda}_n \mathbf{V}_n^H, \end{aligned} \quad (22)$$

where λ'_i 's are descending-ordered eigenvalues and \mathbf{v}'_i are corresponding eigenvectors, for $1 \leq i \leq \text{rank } \hat{\mathbf{R}}'_y$; $\mathbf{\Lambda}_s = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_{k^c})$ and $\mathbf{\Lambda}_n = \text{diag}(\hat{\lambda}_{k^c+1}, \hat{\lambda}_{k^c+2}, \dots, \hat{\lambda}_{\text{rank } \hat{\mathbf{R}}'_y})$. The components in noise subspace can be easily identified and removed after k^c is obtained. In order to preserve

the estimated dimension of signal subspace, i.e. k^c , it is required that $\text{rank } \hat{\mathbf{R}}'_y \geq k^c$. On the other hand, from the dimension of $\hat{\mathbf{R}}'_y$ itself, its rank satisfies $\text{rank } \hat{\mathbf{R}}'_y = \min\{M, P'\}$. Hence, it is necessary that the number of measurement vectors for sparse recovery to meet $P' \geq k^c$.

Then, the support of spectrum is proposed to be directly determined by the components in signal subspace $\mathbf{V}_s \boldsymbol{\lambda}_s \mathbf{V}_s^H$, that is

$$\mathbf{V}_s(\mathbf{A}_s)^{\frac{1}{2}} = \mathbf{A}\mathcal{F}^{-1}\mathbf{S}_v + \mathbf{B}_v, \quad (23)$$

where $(\mathbf{S}_v)_{N \times k^c}$ is the signal to be recovered by CS algorithm, and \mathbf{B}_v denotes the perturbing term due to noise residue in the signal subspace. It is noted that the dimension of the joint-block-sparse matrix \mathbf{S}_v in (23) instead of \mathbf{S}'_f (21) can be greatly reduced from $N \times P'$ to $N \times k^c$. The relationship between the sparse signal of the two MMV models, \mathbf{S}_v and \mathbf{S}'_f in (23) and (21) respectively, is expressed by

$$\text{supp}(\mathbf{S}_v) \subseteq \text{supp}(\mathbf{S}'_f) = \text{supp}(\mathbf{S}_f), \quad (24)$$

where the equality holds when the spectrum sparsity is correctly estimated, i.e. $k^c = \delta(\mathbf{s}_f)$. With underestimated spectrum sparsity when SNR and P are small, the model (23) can still be used with possibly worse detecting performance as the support of \mathbf{S}_v is a subset of \mathbf{S}_f . The rationale of (24) is provided in Proposition 2 and Corollary 1 below.¹

Proposition 2: For arbitrary matrices $\mathbf{M} \in \mathbb{C}^{a \times b}$ and $\mathbf{N} \in \mathbb{C}^{a \times c}$, if $\mathbf{M}\mathbf{M}^H = \mathbf{N}\mathbf{N}^H$, then $\text{span}(\mathbf{M}) = \text{span}(\mathbf{N})$.

Proof: $\forall \mathbf{x} \in \mathbb{C}^b$ s.t. $\mathbf{M}^H \mathbf{x} = \mathbf{0}$, one has $\mathbf{M}\mathbf{M}^H \mathbf{x} = \mathbf{0}$, which means $\text{null}(\mathbf{M}^H) \subseteq \text{null}(\mathbf{M}\mathbf{M}^H)$. Similarly, $\forall \mathbf{x} \in \mathbb{C}^b$ s.t. $\mathbf{M}\mathbf{M}^H \mathbf{x} = \mathbf{0}$, the quadratic form satisfies $\mathbf{x}^H \mathbf{M}\mathbf{M}^H \mathbf{x} = \mathbf{0}$, then $\mathbf{M}^H \mathbf{x} = \mathbf{0}$, which is equivalent to $\text{null}(\mathbf{M}\mathbf{M}^H) \subseteq \text{null}(\mathbf{M}^H)$. Thus, it holds that $\text{null}(\mathbf{M}\mathbf{M}^H) = \text{null}(\mathbf{M}^H)$. For matrix \mathbf{N} , again, we have $\text{null}(\mathbf{N}\mathbf{N}^H) = \text{null}(\mathbf{N}^H)$ and it follows that $\text{null}(\mathbf{M}^H) = \text{null}(\mathbf{N}^H) = \text{null}(\mathbf{N}\mathbf{N}^H) = \text{null}(\mathbf{M}\mathbf{M}^H)$. Finally, from rank-nullity theorem, it holds that $\text{span}(\mathbf{M}) = \text{null}(\mathbf{M}^H)^\perp = \text{span}(\mathbf{N}) = \text{null}(\mathbf{N}^H)^\perp$, which completes the proof. ■

Corollary 1: If the MMV model (21) is transformed to (23) by zeroing the noise subspace in (22), and the dimensionality of signal subspace is not overestimated, i.e. $k^c \leq \delta(\mathbf{s}_f)$, then (24) is met. Equality in (24) is guaranteed if $k^c = \delta(\mathbf{s}_f)$.

Proof: Consider the case $\mathbf{B}_v = \mathbf{0}$, and it is noted that $\mathbf{V}_s \boldsymbol{\lambda}_s \mathbf{V}_s^H = \mathbf{Y}'_s \mathbf{Y}'_s{}^H$ when $k^c = \delta(\mathbf{s}_f)$. From Proposition 2 one has $\text{span}(\mathbf{V}_s(\mathbf{A}_s)^{\frac{1}{2}}) = \text{span}(\mathbf{Y}_s)$. Since columns of the two matrices span the same k^c -dimension subspace in \mathbb{C}^M , they can be linked via a unique linear transform characterized by full-(column)-rank matrix $\mathbf{L}_{P \times k^c}$, that is $\mathbf{V}_s(\mathbf{A}_s)^{\frac{1}{2}} = \mathbf{Y}_s \mathbf{L}$. There exists a matrix $\mathbf{S}_v = \mathbf{S}_f \mathbf{L}$ that satisfies $\mathbf{V}_s(\mathbf{A}_s)^{\frac{1}{2}} = \mathbf{A}\mathbf{S}_v = \mathbf{Y}_s \mathbf{L} = \mathbf{A}\mathbf{S}_f \mathbf{L}$. And such \mathbf{S}_v apparently has the same support as \mathbf{S}_f (i.e. equality in (24)) with row sparsity of $\delta(\mathbf{s}_f)$.

For underestimated dimensionality of signal subspace $k^c < \delta(\mathbf{s}_f)$, the detected signal subspace has reduced dimensionality

compared to that of \mathbf{Y}_s , that is $\text{span}(\mathbf{V}_s(\mathbf{A}_s)^{\frac{1}{2}}) \subset \text{span}(\mathbf{Y}_s)$. Then, the relationships $\mathbf{V}_s(\mathbf{A}_s)^{\frac{1}{2}} = \mathbf{Y}_s \mathbf{L}$ and $\mathbf{S}_v = \mathbf{S}_f \mathbf{L}$ hold where there exists a matrix $\mathbf{L}_{P \times \delta(\mathbf{s}_f)}$ with the rank of k^c . Similarly, this leads to $\text{supp}(\mathbf{S}_v) \subseteq \text{supp}(\mathbf{S}_f)$. ■

B. Evaluations on Greedy Algorithms

Here we consider four algorithms frequently referred to in literature, namely Orthogonal Matching Pursuit (OMP), Compressive Sampling Matching Pursuit (CoSaMP) [40], Subspace Pursuit (SP) [41], and Hard Thresholding Pursuit (HTP) [18], all of which also require estimated sparsity as input. For simplicity, we focus on the basic SMV model (5) for the purpose of comparing these algorithms and k^c is a natural choice of input sparsity estimate of \mathbf{s}_f . It is noted that the complexity of these algorithms is dominated by the correlation multiplication step with the complexity of $\mathcal{O}(NM)$, and the least-square estimation step as it refers to the computation of the pseudo-inverse of a $M \times k^c$ column-full-rank matrix - an $\mathcal{O}((k^c)^2 M)$ operation.² Without the knowledge of the signal sparsity, a safe practice is inputting an overestimated or the maximum allowed value as estimated sparsity in order to reduce the residue as much as possible. The difference between OMP and these three algorithms is that the OMP cumulatively detects the support in the desired signal, while CoSaMP, SP and HTP updates a certain number of supports batch by batch until a satisfactory support is detected upon convergence. For OMP, the maximum allowed input of sparsity estimate is M with which a unique solution is marginally guaranteed in least-square estimation step. The complexity reduction due to a valid estimated sparsity is apparent - the explicit number of iterations is reduced from M to k^c .

In other three algorithms, pruning operation is essential to update the most responsible supports in each iteration. Specifically, CoSaMP has the pruning procedure where k^c most prominent supports out of $3k^c$ remain in each iteration. This pruning procedure dictates the maximum allowed input $k^c = M/3$, which implies sparsity can not be too much overestimated. Similarly, SP proposes to prune k^c out of $2k^c$ supports and then recalculates the least-squares in each iteration. In terms of HTP, the pruning process selects the k^c most relevant supports out of M . The complexity reduction comes from two aspects - one is the reduced dimensionality of the least-square estimation problem in each iteration from $\mathcal{O}(M^3)$ to $\mathcal{O}((k^c)^2 M)$, and another is the reduced average number of iterations as a valid estimate of sparsity aids to find the correct subspace dimension of the signal to be recovered.

In addition to the benefits of reduced complexity, the estimated sparsity information also aids to reliably recover the signal and find the correct signal support. Informally, for general greedy algorithms, inputting an overestimated signal's sparsity results in excessive free dimensions in the least-squares step, hence it leads to more inclusion of falsely-detected non-zero entries caused by noise perturbations in inaccurate measurements.

¹The support of a joint-sparse matrix $\text{supp}(\cdot)$ is defined as the set of the indices of the rows whose 0-norm are non-zero.

²In this paper, the pseudo-inverse of a $a \times b$ matrix ($b < a$) is considered to be implemented by the SVD-based algorithm, which gives the complexity of $\mathcal{O}(ab^2)$.

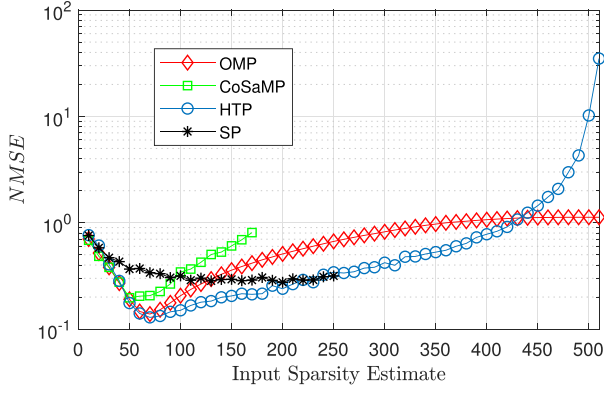


Fig. 6. NMSEs of four greedy algorithms on the SMV model (5) against different input sparsity estimates k^c . Parameters are set as $N = 1280$, $M = 512$, $SNR = 6\text{dB}$ and $\delta(\mathbf{s}_f) = 64$.

For direct empirical results, experiments are conducted for the SMV spectrum recovery problem (5) using the four interested greedy algorithms in order to compare their performance on recovery fidelity and running time. To provide fairness, we use the same stopping criteria for CoSaMP, SP and HTP - that the detected support is the same in two consequent iterations, or the algorithm reaches the maximum number of iterations of 100. These improvements on recovery fidelity illustrated in Fig. 6 is evaluated by normalized minimum-squared error (NMSE), defined by

$$NMSE = \frac{\|\mathbf{s}_f^r - \mathbf{s}_f\|_2^2}{\|\mathbf{s}_f\|_2^2}. \quad (25)$$

In Fig. 6 it is observed that all four algorithms shows the best recovery fidelity near accurate estimate of sparsity $k^c = \delta(\mathbf{s}_f) = 64$. For HTP, with the input sparsity estimate approaching M , the pruning is less effective and the recovery performance degrades with M more severely compared with OMP. It worth noting that when inputting the sparsity estimate as M , the HTP algorithm degenerates to a single correlation step (which has only 2 iterations) and the recovery performance is the worst in terms of NMSE. However, with a valid sparsity estimate, the HTP algorithm offers the lowest NMSE, which is consistent with both the theoretical superiority [18, Th. 3.8] over other three algorithms. For time complexity of these algorithms, as shown in Fig. 7(a)(b), one can see that CoSaMP is sensitive to the input sparsity. Specifically, the detected support tends not to converge with overestimated input sparsity (reaching maximum iteration number of 100). Additional stopping criteria, which terminates the iteration when the residue is less than a predefined tolerance, may be desired to stop the algorithm prior of reaching the maximum number of iteration. However, the tolerance of the residue is related to the noise level, which requires prior knowledge of the signal. On the other hand, HTP still converges with small numbers of iterations. The good convergence performance of HTP and SP leads to the considerate less recovery time of both when compared with CoSaMP. Specially, OMP does not require termination upon convergence by running an explicit number of iterations which equals to k^c , and the running time also largely exceeds that of HTP and SP.

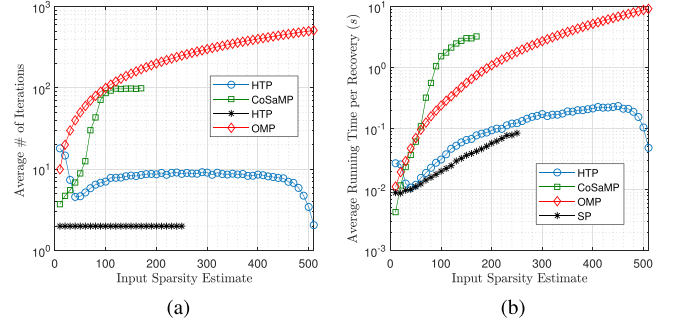


Fig. 7. (a) Average number of iterations of CoSaMP, SP and HTP on the SMV model (5) against different input sparsity estimates k^c (maximum number of iteration is 100); (b) Average running time per recovery of four greedy algorithms on the SMV model (5) against different input sparsity estimates k^c .

Algorithm 1: Joint-block-sparse Hard Thresholding Pursuit (JB-HTP).

Require: \mathbf{A} , $\mathbf{V}_s(\mathbf{\Lambda}_s)^{\frac{1}{2}}$, l^o , $maxIter$.

Ensure: \mathbf{S}_v^r , Ω .

- 1: Initialize $\mathbf{S}_v^{(i+1)} \leftarrow \mathbf{0}$, $\Omega^{(0)} \leftarrow \emptyset$, $\Omega_{old} \leftarrow \emptyset$, $i \leftarrow 0$.
 - 2: **while** ($\Omega^{(i)} \neq \Omega_{old}$ and $i < maxIter$) or $i = 0$ **do**
 - 3: $\Omega^{(i+1)} \leftarrow \{\text{indices of } l^o \text{ largest } \|\cdot\|_F \text{ block entries of } \mathbf{S}_v^{(i)} + (\mathbf{A}\mathcal{F}^{-1})^H [\mathbf{V}_s(\mathbf{\Lambda}_s)^{\frac{1}{2}} - \mathbf{A}\mathcal{F}^{-1}\mathbf{S}_v^{(i)}]\}$;
 - 4: $\Omega_{old} \leftarrow \Omega^{(i)}$;
 - 5: $i \leftarrow i + 1$;
 - 6: $(\mathbf{S}_v^{(i)})_{\text{BLK}, \Omega^{(i)}} \leftarrow \mathbf{0}$;
 - 7: $(\mathbf{S}_v^{(i)})_{\text{BLK}, \Omega^{(i)}} \leftarrow [(\mathbf{A}\mathcal{F}^{-1})_{\text{BLK}, \Omega^{(i)}}]^\dagger \mathbf{V}_s(\mathbf{\Lambda}_s)^{\frac{1}{2}}$;
 - 8: **end while**
 - 9: $\Omega \leftarrow \Omega^{(i)}$;
 - 10: $\mathbf{S}_v^r \leftarrow \mathbf{S}_v^{(i)}$;
-

C. HTP-Based Blind Block Support Detection

From both theoretical guarantees and empirical results above and in [18], we find the HTP algorithm interesting which has superiority in both recovery fidelity and convergence with inaccurate sparsity input. Consequently, we exploit the basic SMV form of HTP and revise it for solving the joint-block-sparse MMV model (23), namely JB-HTP. Having obtained the estimated active channel number l^o from subspace-decomposition-based spectrum sparsity estimation, we propose to use l^o as the input to the proposed algorithm. The routines of the proposed algorithm are illustrated in Algorithm 1. To clarify the notations in the step 4, $(\cdot)_{\text{BLK}, \Omega^{(i+1)}}$ stands for the sub-matrix whose rows/columns are indexed by the block entries in the set $\Omega^{(i+1)}$, and $(\cdot)^\dagger$ denotes pseudo-inverse. In the step 3 in the proposed algorithm, the Frobenius-norm is calculated for each block, which writes

$$\left\| \left(\mathbf{S}_v^{(i+1)} + (\mathbf{A}\mathcal{F}^{-1})^H \left[\mathbf{V}_s(\mathbf{\Lambda}_s)^{\frac{1}{2}} - \mathbf{A}\mathcal{F}^{-1}\mathbf{S}_v^{(i+1)} \right] \right)_{\text{BLK}, \{c\}} \right\|_F \quad (26)$$

for $c = 0, 1, \dots, C - 1$.

Having finished the recovery of joint-block-sparse matrix \mathbf{S}_v , one can perform ED based on the recovered matrix \mathbf{S}_v^r to

Algorithm 2: Low-complexity Blind CSS with Spectrum Sparsity Estimation.

Require: $M, N, P, \mathbf{A}, \mathbf{Y}, C, \maxIter$.

Ensure: Ω

- 1: Construct $\hat{\mathbf{R}}_y$ from P snapshots and perform EVD to obtain $\hat{\lambda}_1, \dots, \hat{\lambda}_M$ and $\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_M$ as in (13);
 - 2: Calculate BICe for $k = 1, \dots, M$ and find k^o corresponding the minimum BICe as in (18);
 - 3: Compensate for underestimation and obtain l^o and k^c as in (19) and (20) respectively;
 - 4: Construct \mathbf{V}_s and $\mathbf{\Lambda}_s$ as in (22);
 - 5: Recover the spectrum and detect incumbent blocks by $\{\mathbf{S}'_v, \Omega\} \leftarrow \text{JB-HTP}(\mathbf{A}, \mathbf{V}_s(\mathbf{\Lambda}_s)^{\frac{1}{2}}, l^o, \maxIter)$;
-

determine the channel occupancy. However, ED refers to setting an optimal threshold in the sense of detection probability, which requires extra knowledge of statistics of \mathbf{S}_v . Due to the nature of greedy pursuit, it is proposed to utilize the byproduct of the recovery algorithm - the detected block support Ω as the spectrum occupancy decision, which forms a HD scheme.

If the reconstructed spectrum $(\mathbf{S}'_f)^r_{N \times P'}$ analogous to the original joint-block-sparse spectrum \mathbf{S}'_f in the MMV model (21) is desired, it can be obtained by an additional single projection step

$$\begin{aligned} (\mathbf{S}'_f)^r_{\text{BLK}, \bar{\Omega}} &\leftarrow \mathbf{0} \\ (\mathbf{S}'_f)^r_{\text{BLK}, \Omega} &\leftarrow [(\mathbf{A}\mathbf{F}^{-1})_{\text{BLK}, \Omega}]^\dagger \mathbf{Y}'_s. \end{aligned} \quad (27)$$

To summarize this section, detailed procedures of the proposed low-complexity blind CSS with spectrum sparsity estimation are given in Algorithm 2 and a complete overview of the proposed CSS system architecture is given in Fig. 1.

D. Theoretical Guarantees and Time Complexity of JB-HTP

Here, we exhibit two conclusions of JB-HTP we are interested in - firstly the exact recovery and convergence guarantee with accurate measurements of joint-block-sparse signals, and secondly the guarantee of the order of iteration times.

Proposition 3: If the $3k^c$ -th-order restricted isotropic constant³ of the measurement matrix $\Phi := \mathbf{A}\mathbf{F}^{-1} \in \mathbb{C}^{M \times N}$ satisfies $\sigma_{\Phi, 3k^c} < 1/\sqrt{3}$, for any joint-block-sparse matrix $\mathbf{S}_v \in \mathbb{C}^{N \times k^c}$ with row sparsity of k^c , the solution $\mathbf{S}_v^{(i)}$ in the i th iteration of JB-HTP based on the MMV model (23) converges exactly to \mathbf{S}_v .

See Appendix A for proof of Proposition 3.

Proposition 4: If the $3k^c$ -th-order restricted isotropic constant of the measurement matrix $\Phi := \mathbf{A}\mathbf{F}^{-1} \in \mathbb{C}^{M \times N}$ satisfies $\sigma_{\Phi, 3k^c} < 1/\sqrt{3}$, for any joint-block-sparse matrix $\mathbf{S}_v \in \mathbb{C}^{N \times k^c}$ with row sparsity of k^c and block sparsity of l , the average number of iterations required by the JB-HTP algorithm based on the MMV model (23) is at most $\mathcal{O}(\log l) + \mathcal{O}(1)$.

³Defined in Lemma 2 in Appendix A.

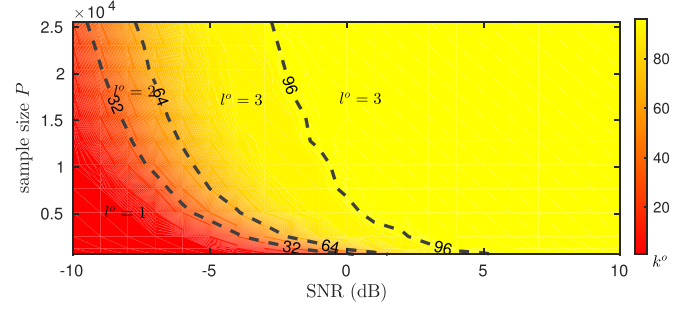


Fig. 8. Estimated sparsity k^o and estimated number of active channels l^o against SNR and the number of snapshots for estimation P .

See Appendix B for proof of Proposition 4.

Theoretically, each iteration of the proposed JB-HTP algorithm for MMV model (23) has complexity of $\mathcal{O}(k^c N M)$ from the matrix multiplication in the correlation step (line 3) and $\mathcal{O}((k^c)^2 M)$ from pseudo-inverse in the least-square step (line 4). Consequently, each iteration's complexity writes $\mathcal{O}(k^c N M) + \mathcal{O}((k^c)^2 M) = \mathcal{O}(k^c N M)$ as $k^c \ll N$ always holds. Hence, with Proposition 4, the total complexity of JB-HTP is expressed by $\mathcal{O}(k^c N M \log l)$.

V. NUMERICAL SIMULATIONS

In simulations, time-domain OFDM signals \mathbf{s}_o and \mathbf{s}_t are generated as in (2) and (8). Constant parameters are set as $B = 8$ (MHz), $C = 40$, $N = 1280$ and $CB/F_s = 7680$, and each row of \mathbf{A} is independent and generated by normalized Gaussian random vector $\mathcal{CN}(\mathbf{0}, \frac{1}{N}\mathbf{I})$. Source symbols γ are independent Gaussian. Additive white Gaussian noise channel model is adopted and received power in all occupied channels is identical. Unless specified elsewhere, the default settings of parameters in the simulation are as follows. Actual active channel number is set as $l := \text{card}(\Omega_s) = 3$, making $\delta(\mathbf{s}_f) = 96$ with central frequencies of incumbent channels set by $[f_1 \ f_2 \ f_3] = [76 \ 164 \ 244]$ (MHz), and the number of elements in each measurement vector is $M = 512$. The number of measurement vectors for sparse recovery in (21) is $P' = M = 512$.

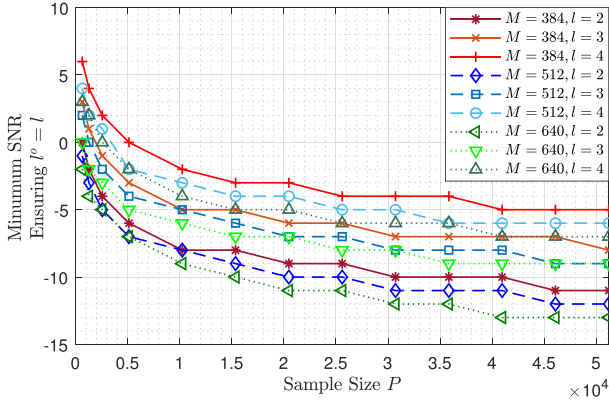
A. Performance Evaluation of the Proposed Spectrum Sparsity Estimation Scheme

Firstly, as shown in Fig. 8, it is noted that the proposed spectrum sparsity estimation scheme accurately detects the true sparsity $\delta(\mathbf{s}_f) = lN/C = 96$ with SNR and the number of measurement vectors P sufficiently large. Moreover, possible determinations of the active channel number $l^o = 1, 2, 3$ based on k^o have been shown. As a results of that the spectrum sparsity k^o is never overestimated, the proposed ceiling operation in (19) is justified and it provides more robustness against low SNR and P while still accurately detects the active channel number in high SNR and P cases.

In Fig. 9, the minimum integer SNRs with which the detection marginally gives correct active channel number with varying sample size P are shown. The results have verified

TABLE I
 SETTINGS OF THE PROPOSED CSS SOLUTION AND BENCHMARK SCHEMES

Scheme No.	CS Algorithm	CS Model	Detection Method	k^c available?	Recovery Complexity
I	BP [19]	SMV-(5)	ED	Not Applicable	$\mathcal{O}(N^3)$ [5]
II	OMP [19]	SMV-(5)	ED	No, using $k^c = \mathcal{O}(M)$	$\mathcal{O}(M^4)$
III	HTP [18]	SMV-(5)	ED	No, using $k^c = \mathcal{O}(M)$	$\mathcal{O}(M^3 \log k^c)$
IV	OMP	SMV-(5)	HD	Yes	$\mathcal{O}(k^c NM)$ [41]
V	HTP	SMV-(5)	HD	Yes	$\mathcal{O}(NM \log k^c)$ [18]
VI	M-BP [42]	MMV-(23)	ED	Yes	$\mathcal{O}(k^c N^3)$
VII	SOMP [43]	MMV-(23)	HD	Yes	$\mathcal{O}((k^c)^2 NM)$
VIII (the proposed scheme)	JB-HTP	MMV-(23)	HD	Yes	$\mathcal{O}(k^c NM \log l)$


 Fig. 9. Minimum SNRs to ensure correct detection of active channel number $l^o = l$ against numbers of snapshots P , of varying numbers of elements each measurement vector M and l .

the expectation that the decreasing M or increasing l lead to reduced dimensions in the noise subspace, hence deteriorated performance.

B. Performance Evaluation of the Proposed CSS Scheme

1) *Benchmark Scheme Settings:* In order to evaluate the performance of the proposed JB-HTP-based CSS scheme, we choose to compare it with a few benchmark schemes illustrated in Table I. In scheme No. I, II, and III, three CS recovery algorithms, BP, OMP and HTP are chosen to solve the sparse recovery problem of the SMV model (5) without the knowledge of the spectrum sparsity where ED scheme is used. To compare to SMV-ED schemes, using the active channel number estimation scheme proposed in Section III, SMV-HD schemes with OMP and HTP recovery algorithms are examined in scheme No. IV and V. Apart from the proposed active channel number estimation scheme, in scheme No. VI, VII, and VIII, the MMV model with dimension and noise reduction proposed in Section IV, i.e. (21), applies for sparse recovery. Particularly, in scheme No. VI, the MMV version of BP - multiple-BP (M-BP) [42] - is used to solve (9) where sparsity estimation is not applicable and hence ED has to be used. In scheme No. VI and VII, the MMV version of OMP, simultaneous-OMP (SOMP) [43], and the proposed JB-HTP are used for MMV model (21) where HD is used because a spectrum sparsity estimation is required by the formulation of the proposed model (21).

It should be noted that the recovery algorithms - OMP, HTP and SOMP in scheme No. IV, V and VII are not purposed for block-sparse signals so that the block support (i.e. active channel support) is not directly output by these recovery algorithms. In these schemes, we use an intuitive procedure to produce the active channel support as in [44]. Having recovered the spectrum $\mathbf{s}_f^r = [s_f^r[0] \ s_f^r[1] \ \dots \ s_f^r[N-1]]^T$ from (5) or $(\mathbf{S}_v^r)_{N \times k^c}$ from (23), we calculate the channel energy $\mathbf{p}_f = [p_f[0] \ p_f[1] \ \dots \ p_f[N-1]]^T$ whose entries are

$$p_f[c] = \sum_{i=cN/C}^{(c+1)N/C-1} |s_f^r[i]|^2, \quad (28)$$

for SMV modeo (5) or

$$p_f[c] = \left\| (\mathbf{S}_v^r)_{\text{BLK}, \{c\}} \right\|_F^2, \quad (29)$$

for MMV model (23), where $c = 0, 1, \dots, C-1$. We then directly determine the active channels by sorting $p_f[n]$'s and indicating the channels corresponding to the largest l^o entries, which effectively achieves hard detection, formally

$$\Omega = \{\omega_1, \dots, \omega_{l^o} | p_f[\omega_1] \geq \dots \geq p_f[\omega_{l^o}] \geq \dots \geq p_f[\omega_C]\}. \quad (30)$$

2) *Evaluation Metrics:* For the purpose of spectrum sensing in the background of DSA, the performance of detection of occupied and vacant channel is the natural choice of the metrics of the performance. In Monte-Carlo simulations, by definition, the probability of detection P_d and the probability of false alarm P_f are presented by

$$P_d = \frac{\text{total \# correctly detected channels}}{\# \text{ recovery trials} \times l},$$

$$P_f = \frac{\text{total \# falsely detected channels}}{\# \text{ recovery trials} \times (C - l)}.$$

3) *Simulation Results:* For the general effectiveness of the proposed scheme VII with JB-HTP, we exhibit in Fig. 10 the detection performance with different numbers of samples per measurement vector M and block sparsity l of the spectrum. One can see observe that the default setting ($M = 512, l = 3$) gives P_d close and approaching 1 ($P_d > 0.99$) and P_f close and approaching 0 ($P_f < 10^{-3}$) starting from the SNR as low as -20 dB. In comparison with other combination of M and l , the proposed scheme, like other CSS schemes, shows superior detection performance with larger M (i.e. larger compression

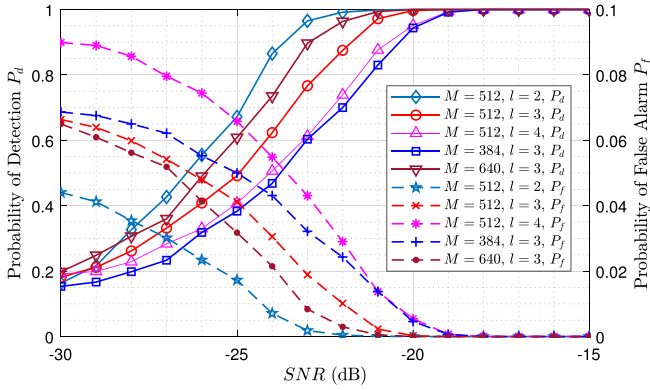


Fig. 10. Detection performance of the proposed JB-HTP scheme on the MMV model (23) against SNR . It is assumed the active channel number are correctly estimated, $l^o = l$ in all shown cases.

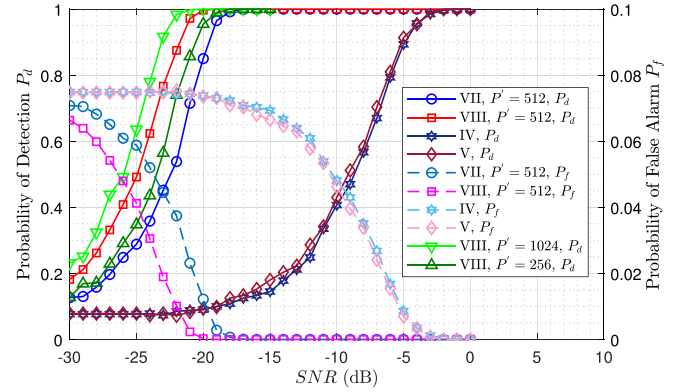


Fig. 12. Detection performance against SNR of MMV-HD schemes compared with SMV-HD counterparts. It is assumed the active channel number are correctly estimated, $l^o = l$ in all shown cases.

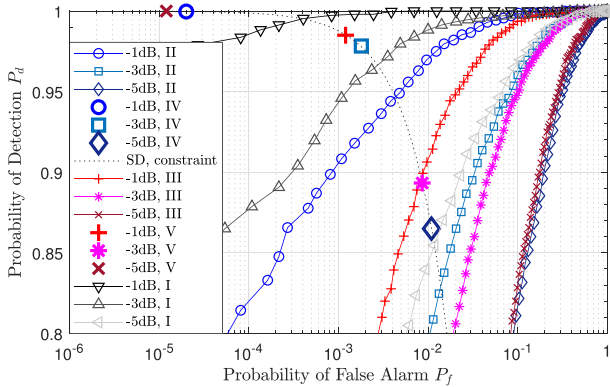


Fig. 11. ROC curves of SMV-ED schemes and scatters of SMV-HD schemes in Table I with $SNRs$ from -1 dB to -5 dB and $l^o = l = 3$.

ratio) and smaller l (i.e. smaller spectrum sparsity). It should be made clear that the estimated active channel number used here is accurate, that is $l^o = l$ being input to Algorithm 1.

Next, we examine the detection performance the SMV schemes to exhibit the benefit of using HD scheme brought by the proposed active channel number estimation. Receiver operating characteristics (ROCs) of HD scheme IV and V are given in Fig. 11 to compare with ROC curves of ED schemes - I, II, and III. Note that ROCs of HD schemes are scattered points as it is a hard decision scheme. Within the ED schemes, the BP algorithm outperforms the greedy algorithms in scheme II and III. Moreover, the advantage in detection performance of HD schemes is obvious as each ROC scatter lies in the upper-left region to its counterpart ROC curve of ED scheme, and even the ED scheme with BP, under the same SNR . Specifically, the proposed HD scheme can achieve the same P_d level while dramatically reduces P_f by orders of magnitude. This advantage may be intuitively ascribed to the aid of the extra information of the active channel number estimation to both the recovery and decisioning procedures. It is also noted that the effect of the hard decision based on l^o intrinsically constrained P_f with P_d , that is $P_f = (l^o - lP_d)/(C - l)$. This implies that

the proposed hard-decision scheme has constrained and small P_f when P_d is relatively large and l is only a fraction of C . The constraint for $l^o = l = 3$ is drawn in Fig. 11 where $P_f < 0.016$ is guaranteed when $P_d > 0.8$. This inherent constraint on P_f is desirable compared to the ED scheme, where setting a threshold with P_f constraints requires knowledge of the noise statistics.

Furthermore, we account the benefit of implementing the noise and dimension reduction (22) and using the MMV model (23) by comparing the detection performance of MMV-HD schemes - VII and VIII - with that of SMV-HD schemes - IV and V. In Fig. 12, it is clear that each MMV scheme - the proposed JB-HTP or SOMP - shows abundantly increased robustness regarding detection performance against noise. Taking $P_d > 0.99$ and $P_f < 10^{-3}$ as reference levels, quantitatively, the proposed CSS scheme (VIII) with JB-HTP shows the superior robustness against noise of as much as 20dB to achieve equivalent detection performance in comparison to both SMV-HD schemes with OMP and HTP. On the other hand, Fig. 12 indicates that the use of the proposed JB-HTP algorithm based on MMV model (23) proves to deliver better detection performance than using SOMP in scheme No. VII. Furthermore, we also experiment the proposed scheme (VIII) with different number of measurement vectors, $P' = 256, 512$, and 1024. As expected, a larger number of stacked measurement vectors in (21) leads to better separation of signal and noise components via eigendecomposition hence better detection performance after sparse recovery, but one should bear in mind that a large P' also implies longer sample acquisition time and consequently the delay of spectrum sensing results.

In Fig. 13, the detection performances of MMV schemes VI, VII and VIII are illustrated in the form of ROC scatters and curves. The results show that the proposed scheme VIII with JB-HTP outperforms the scheme VII with SOMP and the scheme VI with M-BP and ED in respect of detection performance, while the scheme VII is only slightly better than the ED scheme VI. Additionally, it is also worth considering the cases when l^o is underestimated due to low SNR or the number of

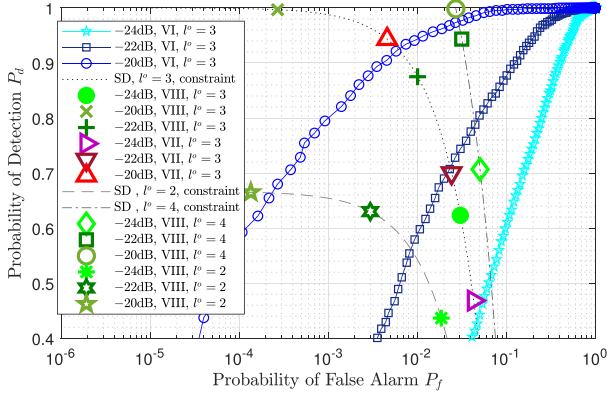


Fig. 13. ROC curves of MMV-ED schemes and scatters of MMV-HD schemes in Table I with $SNRs$ from -20dB to -24dB and $l^o = l = 3$. ROC scatters of MMV-HD schemes with underestimated active channel number $l^o = l - 1 = 2$ and overcompensated active channel number $l^o = l + 1 = 3$ are also shown for comparison.

measurement vectors P , and the possible option of purposely overestimating l^o . With underestimated l^o , any HD scheme, can only achieve relatively poor P_d upper-bounded by l^o/l , as a direct consequence of at least $l - l^o$ miss-detected active channel. However, one can deliberately compensate l^o by some integer where some levels of underestimation are prone to appear under certain $SNRs$. The price of such compensation method is that l^o is also likely to be overestimated. We experiment with the proposed scheme VII to examine the detection performance with both underestimate ($l^o = l - 1 = 2$) and overcompensate ($l^o = l + 1 = 4$) cases and results are also given in Fig. 13. In the underestimate case, P_d 's deterioration is obvious - it can only achieve no higher than 0.67 as expected. In the overcompensate case, the price is the increased P_f which is lower-bounded by $1/(C - l) = 0.027$, and P_d is somehow greater, compared to the case where $l^o = 3$ under the same $SNRs$.

Finally, we address the complexity of these benchmarked schemes I-VII and the proposed scheme VIII. The time complexity of each CSS scheme in Table I has been accounted in the order of atomic manipulations of complex addition and multiplication. In comparison within either SMV or MMV schemes, the HTP or the proposed JB-HTP scheme has the smallest order of time complexity in the recovery phase. Note there is also an additional complexity of $\mathcal{O}(M^3)$ in scheme No. IV-VIII from the active channel number estimation due to the SVD procedure. In addition, the empirical results on the average time per recovery of the 8 schemes are shown in Fig. 14 with 3 different values of M . The time required for estimating the active channel number has also been accounted for schemes where applicable. As a general trend, the average time consumed per recovery is positively related to the value of M . More importantly, the empirical time complexities, with the account of estimation of l^o , of scheme No. VIII and VI are also the smallest among MMV and SMV schemes, respectively. By comparing the proposed MMV scheme VIII with the SMV counterparts III and VI, although the analytic complexity order and recovery time of scheme VIII is found slightly greater as shown in Fig. 14, it is reminded by our earlier experiments in

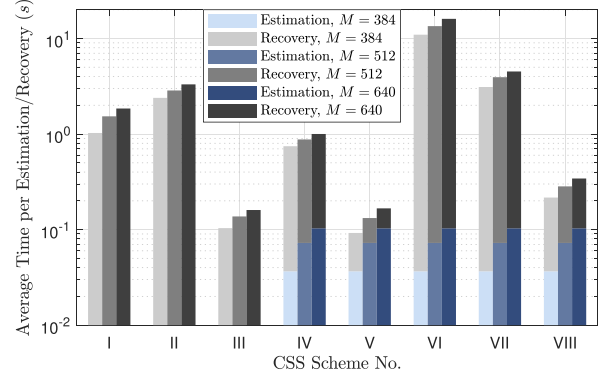


Fig. 14. Average time consumed by the CS recovery schemes in Table I together with the active channel number estimation scheme (if applicable).

Fig. 11 and 12 that the superiority of the scheme VIII in detection performance is tremendous.

VI. CONCLUSION

This paper presents a novel greedy-pursuit-based compressive spectrum sensing (CSS) scheme with the aid of subspace-decomposition-based spectrum sparsity estimation, where the spectrum sparsity is directly estimated from the output of the sub-Nyquist measurements without recovery operations. In the proposed scheme, a multiple-measurement-vector (MMV) model with noise and dimension reduction is introduced, and a generalized version of hard thresholding pursuit (HTP) for joint-block sparse signals is proposed as the recovery algorithm. The benefits of such a novel CSS scheme are summarized in three aspects. Firstly, the estimation of spectrum sparsity effectively alleviates the computational complexity of spectrum recovery with an explicit spectrum sparsity estimate compared to the cases where such information is absent. Secondly, the use of the proposed MMV model with noise and dimension reduction further enhances the detection performance (as much as 20dB seen in simulation) while effectively keeping the complexity low. Thirdly, the use of the spectrum sparsity estimate enables hard detection when determining the occupancy of wireless channels, which improves the detection performance from that of energy detection schemes, as well as removes the need for threshold adaption. Monte-Carlo simulations have evaluated the performance of the spectrum sparsity estimation, shown the superior detection performance of the proposed scheme against multiple benchmarking schemes, and verified its low time complexity.

ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council in United Kingdom under Grant EP/R00711X/1.

APPENDIX A PROOF OF PROPOSITION 3

Proof: Before formally starting the proof, we introduce Lemma 2 directly from the definition of restricted isotropic property (RIP).

Lemma 2 ([18]): Define the RIP constant of matrix Φ with sparsity s as

$$\sigma_{\Phi,s} := \max_{\{S | \text{card}(S) \leq s\}, \forall \mathbf{x} \in \mathbb{C}^S} \frac{\langle ((\Phi)_{:,S}^H (\Phi)_{:,S} - \mathbf{I}) \mathbf{x}, \mathbf{x} \rangle}{\|\mathbf{x}\|_2^2}, \quad (31)$$

and the following relations hold

$$\| \langle (\mathbf{I} - \Phi^H \Phi) \mathbf{u}, \mathbf{v} \rangle \|_2 \leq \sigma_t \|\mathbf{u}\|_2 \|\mathbf{v}\|_2, \quad (32)$$

$$\forall \mathbf{u}, \mathbf{v} \text{ s.t. } \text{card}(\text{supp}(\mathbf{u}) \cup \text{supp}(\mathbf{v})) \leq t,$$

and

$$\| ((\mathbf{I} - \Phi^H \Phi) \mathbf{v})_U \|_2 \leq \sigma_t \|\mathbf{v}\|_2, \quad (33)$$

$$\forall U, \mathbf{v} \text{ s.t. } \text{card}(U \cup \text{supp}(\mathbf{v})) \leq t.$$

For conciseness, denote $\Phi := \mathbf{A}\mathcal{F}^{-1}$ and $\Psi := \mathbf{V}_s(\Lambda_s)^{\frac{1}{2}}$. From the least-square step (line 4) in Algorithm 1, the residue produced in the $(i+1)$ th iteration $\Phi \mathbf{S}_v^{(i+1)} - \Psi$ is orthogonal to the measurement space determined by the selected block support, $\text{span}((\Phi)_{\text{BLK}, \Omega^{(i+1)}})$. This leads to

$$(\Phi \mathbf{S}_v)^H (\Phi \mathbf{S}_v^{(i+1)} - \Psi) = \mathbf{0}, \forall \mathbf{S}_v \text{ s.t. } (\mathbf{S}_v)_{\text{BLK}, \overline{\Omega^{(i+1)}}} = \mathbf{0}, \quad (34)$$

which we rewrite as

$$(\Phi^H \Phi \mathbf{S}_v)^H (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v) = \mathbf{0}, \forall \mathbf{S}_v \text{ s.t. } (\mathbf{S}_v)_{\text{BLK}, \overline{\Omega^{(i+1)}}} = \mathbf{0}. \quad (35)$$

Then, we examine the error from the detected support

$$\begin{aligned} & \left\| (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)_{\text{BLK}, \Omega^{(i+1)}} \right\|_F^2 \\ &= \text{tr} \left((\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)^H (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)_{\text{BLK}, \Omega^{(i+1)}} \right) \\ &\stackrel{(35)}{=} \text{tr} \left((\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)^H \right. \\ &\quad \left. \cdot (\mathbf{I} - \Phi^H \Phi) (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)_{\text{BLK}, \Omega^{(i+1)}} \right) \\ &\stackrel{(32)}{\leq} \sigma_{\Phi, 2k^c} \left\| \mathbf{S}_v^{(i+1)} - \mathbf{S}_v \right\|_F \left\| (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)_{\text{BLK}, \Omega^{(i+1)}} \right\|_F, \end{aligned} \quad (36)$$

from which we arrive at

$$\left\| (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)_{\text{BLK}, \Omega^{(i+1)}} \right\|_F \leq \sigma_{\Phi, 2k^c} \left\| \mathbf{S}_v^{(i+1)} - \mathbf{S}_v \right\|_F. \quad (37)$$

Next, we account the total error at the $(i+1)$ th iteration

$$\begin{aligned} & \left\| (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v) \right\|_F^2 \\ &= \left\| (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)_{\text{BLK}, \Omega^{(i+1)}} \right\|_F^2 \\ &\quad + \left\| (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)_{\text{BLK}, \overline{\Omega^{(i+1)}}} \right\|_F^2 \\ &\leq \sigma_{\Phi, 2k^c}^2 \left\| \mathbf{S}_v^{(i+1)} - \mathbf{S}_v \right\|_F^2 + \left\| (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)_{\text{BLK}, \overline{\Omega^{(i+1)}}} \right\|_F^2, \end{aligned} \quad (38)$$

from which we obtain

$$\left\| \mathbf{S}_v^{(i+1)} - \mathbf{S}_v \right\|_F^2 \leq \frac{1}{1 - \sigma_{\Phi, 2k^c}^2} \left\| (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v)_{\text{BLK}, \overline{\Omega^{(i+1)}}} \right\|_F^2. \quad (39)$$

On the other hand, we examine the correlation step (line 3) in Algorithm 1 and instantly have the following

$$\begin{aligned} & \left\| (\mathbf{S}_v^{(i)} + \Phi^H (\Psi - \Phi \mathbf{S}_v^{(i)}))_{\text{BLK}, \Omega^{(i+1)}} \right\|_F^2 \\ &\geq \left\| (\mathbf{S}_v^{(i)} + \Phi^H (\Psi - \Phi \mathbf{S}_v^{(i)}))_{\text{BLK}, \Omega_s} \right\|_F^2, \end{aligned} \quad (40)$$

where Ω_s is the block support of \mathbf{S}_v . The above may be rewritten as

$$\begin{aligned} & \left\| (\mathbf{S}_v^{(i)} + \Phi^H (\Psi - \Phi \mathbf{S}_v^{(i)}))_{\text{BLK}, \Omega^{(i+1)} \setminus \Omega_s} \right\|_F^2 \\ &= \left\| ((\mathbf{I} - \Phi^H \Phi) (\mathbf{S}_v^{(i+1)} - \mathbf{S}_v))_{\text{BLK}, \Omega^{(i+1)} \setminus \Omega_s} \right\|_F^2 \\ &\geq \left\| (\mathbf{S}_v^{(i)} + \Phi^H (\Psi - \Phi \mathbf{S}_v^{(i)}))_{\text{BLK}, \Omega_s \setminus \Omega^{(i)}} \right\|_F^2. \end{aligned} \quad (41)$$

The right-hand side of the inequality (41) is reorganized as the following given the fact $(\mathbf{S}_v^{(i+1)})_{\text{BLK}, \overline{\Omega^{(i+1)}}} = \mathbf{0}$,

$$\begin{aligned} & \left\| (\mathbf{S}_v^{(i)} + \Phi^H (\Psi - \Phi \mathbf{S}_v^{(i)}))_{\text{BLK}, \Omega_s \setminus \Omega^{(i+1)}} \right\|_F \\ &= \left\| ((\mathbf{I} - \Phi^H \Phi) (\mathbf{S}_v^{(i)} - \mathbf{S}_v))_{\text{BLK}, \Omega_s \setminus \Omega^{(i+1)}} \right. \\ &\quad \left. + (\mathbf{S}_v - \mathbf{S}_v^{(i+1)})_{\text{BLK}, \overline{\Omega^{(i+1)}}} \right\|_F \\ &\geq \left\| ((\mathbf{I} - \Phi^H \Phi) (\mathbf{S}_v^{(i)} - \mathbf{S}_v))_{\text{BLK}, \Omega_s \setminus \Omega^{(i+1)}} \right\|_F \\ &\quad + \left\| (\mathbf{S}_v - \mathbf{S}_v^{(i+1)})_{\text{BLK}, \overline{\Omega^{(i+1)}}} \right\|_F. \end{aligned} \quad (42)$$

Focus on the last term and one can shrink it from (42) and (41)

$$\begin{aligned} & \left\| (\mathbf{S}_v - \mathbf{S}_v^{(i+1)})_{\text{BLK}, \overline{\Omega^{(i+1)}}} \right\|_F \\ &\leq \left\| ((\mathbf{I} - \Phi^H \Phi) (\mathbf{S}_v^{(i)} - \mathbf{S}_v))_{\text{BLK}, \Omega_s \setminus \Omega^{(i+1)}} \right\|_F \\ &\quad + \left\| ((\mathbf{I} - \Phi^H \Phi) (\mathbf{S}_v^{(i)} - \mathbf{S}_v))_{\text{BLK}, \Omega^{(i+1)} \setminus \Omega_s} \right\|_F \\ &\leq \sqrt{2} \left\| ((\mathbf{I} - \Phi^H \Phi) \right. \\ &\quad \left. \cdot (\mathbf{S}_v^{(i)} - \mathbf{S}_v))_{\text{BLK}, (\Omega_s \setminus \Omega^{(i+1)}) \cup (\Omega^{(i+1)} \setminus \Omega_s)} \right\|_F \\ &\stackrel{(33)}{\leq} \sqrt{2} \sigma_{\Phi, 3k^c} \left\| (\mathbf{S}_v - \mathbf{S}_v^{(i)}) \right\|_F. \end{aligned} \quad (43)$$

Congregate (39) and (43) and we finally obtain the recurrence relation of error, writing

$$\left\| \mathbf{S}_v^{(i+1)} - \mathbf{S}_v \right\|_F \leq \sqrt{\frac{2\sigma_{\Phi,3k^c}^2}{1 - \sigma_{\Phi,2k^c}^2}} \left\| \mathbf{S}_v^{(i)} - \mathbf{S}_v \right\|_F. \quad (44)$$

Convergence of the algorithm requires the constant in (44) less than one. By definition, $\sigma_{\Phi,3k^c} > \sigma_{\Phi,2k^c}$ naturally holds. This immediately leads to $\sigma_{\Phi,3k^c} < 1/\sqrt{3}$ as a guarantee of JB-HTP's convergence, which completes the proof. ■

APPENDIX B PROOF OF PROPOSITION 4

Proof: The final iteration of JB-HTP is determined at the point where $\Omega^{(i-1)} = \Omega^{(i)} = \Omega = \Omega_s$. For starter, we consider the selected block upon convergence $\forall p \in \Omega$ and others $\forall q \in \bar{\Omega}$, and naturally have the following

$$\begin{aligned} & \left\| \left(\mathbf{S}_v^{(i-1)} + \Phi^H \left(\Psi - \Phi \mathbf{S}_v^{(i-1)} \right) \right)_{\text{BLK},\{p\}} \right\|_F \\ & > \left\| \left(\mathbf{S}_v^{(i-1)} + \Phi^H \left(\Psi - \Phi \mathbf{S}_v^{(i-1)} \right) \right)_{\text{BLK},\{q\}} \right\|_F \end{aligned} \quad (45)$$

On the left-hand side, we observe

$$\begin{aligned} & \left\| \left(\mathbf{S}_v^{(i-1)} + \Phi^H \left(\Psi - \Phi \mathbf{S}_v^{(i-1)} \right) \right)_{\text{BLK},\{p\}} \right\|_F \\ & = \left\| \left(\mathbf{S}_v \right)_{\text{BLK},\{p\}} + \left(\mathbf{I} - \Phi^H \Phi \right) \left(\mathbf{S}_v^{(i-1)} - \mathbf{S}_v \right) \right\|_F \\ & \geq \epsilon - \left\| \left(\mathbf{I} - \Phi^H \Phi \right) \left(\mathbf{S}_v^{(i-1)} - \mathbf{S}_v \right) \right\|_F, \end{aligned} \quad (46)$$

where $\epsilon := \min_{p \in \Omega} \left\| \left(\mathbf{S}_v \right)_{\text{BLK},\{p\}} \right\|_F$. On the right-hand side, given the fact that $\left(\mathbf{S}_v \right)_{\text{BLK},\{q\}} = \mathbf{0}$, it holds that

$$\begin{aligned} & \left\| \left(\mathbf{S}_v^{(i-1)} + \Phi^H \left(\Psi - \Phi \mathbf{S}_v^{(i-1)} \right) \right)_{\text{BLK},\{q\}} \right\|_F \\ & = \left\| \left(\mathbf{I} - \Phi^H \Phi \right) \left(\mathbf{S}_v^{(i-1)} - \mathbf{S}_v \right) \right\|_F. \end{aligned} \quad (47)$$

Subtract (47) from (46), and one may write and further reorganize as follows,

$$\begin{aligned} & \left\| \left(\mathbf{S}_v^{(i-1)} + \Phi^H \left(\Psi - \Phi \mathbf{S}_v^{(i-1)} \right) \right)_{\text{BLK},\{p\}} \right\|_F \\ & - \left\| \left(\mathbf{S}_v^{(i-1)} + \Phi^H \left(\Psi - \Phi \mathbf{S}_v^{(i-1)} \right) \right)_{\text{BLK},\{q\}} \right\|_F \\ & \geq \epsilon - \left\| \left(\mathbf{I} - \Phi^H \Phi \right) \left(\mathbf{S}_v^{(i-1)} - \mathbf{S}_v \right) \right\|_F \\ & - \left\| \left(\mathbf{I} - \Phi^H \Phi \right) \left(\mathbf{S}_v^{(i-1)} - \mathbf{S}_v \right) \right\|_F \\ & \stackrel{(32)}{\geq} \epsilon - \sqrt{2} \left\| \left(\mathbf{I} - \Phi^H \Phi \right) \left(\mathbf{S}_v^{(i-1)} - \mathbf{S}_v \right) \right\|_F \end{aligned}$$

$$\begin{aligned} & \geq \epsilon - \sqrt{2} \sigma_{\Phi,3k^c} \left\| \mathbf{S}_v^{(i-1)} - \mathbf{S}_v \right\|_F \\ & \stackrel{(44)}{\geq} \epsilon - \sqrt{2} \sigma_{\Phi,3k^c} \left(\sqrt{\frac{2\sigma_{\Phi,3k^c}^2}{1 - \sigma_{\Phi,2k^c}^2}} \right)^{i-1} \left\| \mathbf{S}_v^{(0)} - \mathbf{S}_v \right\|_F \\ & = \epsilon - \sqrt{1 - \sigma_{\Phi,2k^c}^2} \cdot \mu^i \left\| \mathbf{S}_v^{(0)} - \mathbf{S}_v \right\|_F \\ & \stackrel{\text{Prop. 3}}{>} \epsilon - \sqrt{\frac{2}{3}} \cdot \mu^i \left\| \mathbf{S}_v^{(0)} - \mathbf{S}_v \right\|_F, \end{aligned} \quad (48)$$

where $\mu := \sigma_{\Phi,3k^c} / \sqrt{1 - \sigma_{\Phi,2k^c}^2}$. To guarantee (46) to hold, inequality (48) may be bounded as following where i can be determined as a conservative upper bound,

$$\begin{aligned} 0 < \epsilon - \sqrt{\frac{2}{3}} \cdot \mu^i \left\| \mathbf{S}_v^{(0)} - \mathbf{S}_v \right\|_F \\ i < \left\lceil \frac{\log \left(\sqrt{\frac{2}{3}} \cdot \frac{\left\| \mathbf{S}_v^{(0)} - \mathbf{S}_v \right\|_F}{\epsilon} \right)}{\log \frac{1}{\mu}} \right\rceil. \end{aligned} \quad (49)$$

The specific upper bound of iteration number relies on the original sparse signal \mathbf{S}_v which is a statistical signal, and the initial value $\mathbf{S}_v^{(0)}$. Here, we consider a common case where $\mathbf{S}_v^{(0)} = \mathbf{0}$ and use a constant ρ to denote the expected logarithm ratio between the average received energy among all active channels and the minimum received channel energy,

$$\rho := \mathbb{E} \left[\log \left(\frac{\left\| \mathbf{S}_v \right\|_F}{l\epsilon} \right) \right]. \quad (50)$$

Get expectations on both sides (49) and we obtain

$$\begin{aligned} \mathbb{E}[i] & < \mathbb{E} \left[\left\lceil \frac{\log \left(\sqrt{\frac{2}{3}} \cdot \frac{\left\| \mathbf{S}_v \right\|_F}{\epsilon} \right)}{\log \frac{1}{\mu}} \right\rceil \right] \\ & \leq \mathbb{E} \left[\frac{\log \left(\sqrt{\frac{2}{3}} \cdot \frac{\left\| \mathbf{S}_v \right\|_F}{\epsilon} \right)}{\log \frac{1}{\mu}} \right] + 1 \\ & = \frac{\log \left(\sqrt{\frac{2}{3}} \right) + \log l + \mathbb{E} \left[\log \left(\frac{\left\| \mathbf{S}_v \right\|_F}{l\epsilon} \right) \right]}{\log \frac{1}{\mu}} + 1 \\ & = \frac{\log \left(\sqrt{\frac{2}{3}} \right) + \log l + \rho}{\log \frac{1}{\mu}} + 1 = \mathcal{O}(\log l) + \mathcal{O}(1), \end{aligned} \quad (51)$$

which completes the proof. ■

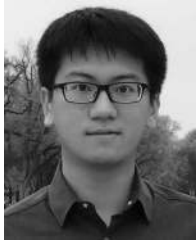
REFERENCES

- [1] A. Attar, O. Holland, and H. Aghvami, "Spectrum access and sharing," in *Cognitive Radio Communications and Networks*, A. M. Wyglinski, M. Nekovee, and Y. T. Hou, Eds. Oxford, U.K.: Academic, 2010, ch. 5, pp. 113–148.

- [2] M. Abo-Zahhad, S. Ahmed, M. Asaduzzaman Farrag, and K. Ali, "Wideband cognitive radio networks based compressed spectrum sensing: A survey," *J. Signal Inf. Process.*, vol. 09, pp. 122–151, Jan. 2018.
- [3] 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.
- [4] Z. Qin, Y. Gao, and C. G. Parini, "Data-assisted low complexity compressive spectrum sensing on real-time signals under sub-Nyquist rate," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1174–1185, Feb. 2016.
- [5] Z. Qin, Y. Gao, M. D. Plumbley, and C. G. Parini, "Wideband spectrum sensing on real-time signals at sub-Nyquist sampling rates in single and cooperative multiple nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3106–3117, Jun. 2016.
- [6] Y. Ma, Y. Gao, Y. C. Liang, and S. Cui, "Reliable and efficient sub-Nyquist wideband spectrum sensing in cooperative cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2750–2762, Oct. 2016.
- [7] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proc. IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw.*, Nov. 2005, pp. 131–136.
- [8] W. Han, J. Li, Z. Tian, and Y. Zhang, "Efficient cooperative spectrum sensing with minimum overhead in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3006–3011, Oct. 2010.
- [9] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.
- [10] O. Holland *et al.*, "To white space or not to white space: That is the trial within the Ofcom TV white spaces pilot," in *Proc. IEEE Int. Conf. Dyn. Spectr. Access Netw.*, Sep. 2015, pp. 11–22.
- [11] "Implementing TV white spaces," Office of Communications, U.K., Tech. Rep., Dec. 2015. [Online]. Available: https://www.ofcom.gov.uk/_data/assets/pdf_file/0034/68668/tvws-statement.pdf
- [12] "Statement on licence-exempting cognitive devices using interleaved spectrum," National Instruments, Austin, TX, USA, Tech. Rep., Jul. 2009. [Online]. Available: https://www.ofcom.gov.uk/_data/assets/pdf_file/0023/40838/statement.pdf
- [13] Y. Ma, Y. Gao, C. Fu, W. Rong, Z. Xiong, and S. Cui, "TV white space spectrum analysis based on machine learning," *J. Commun. Inf. Netw.*, vol. 4, no. 2, pp. 68–80, Jun. 2019.
- [14] M. Rebato and M. Zorzi, "A spectrum sharing solution for the efficient use of mmwave bands in 5G cellular scenarios," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw.*, Aug. 2018, pp. 1–5.
- [15] X. Zhang, Y. Ma, Y. Gao, and S. Cui, "Real-time adaptively regularized compressive sensing in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1146–1157, Feb. 2018.
- [16] 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.
- [17] Y. Arjouni and N. Kaabouch, "Wideband spectrum sensing: A Bayesian compressive sensing approach," *Sensors*, vol. 18, Jun. 2018, Art. no. 1839.
- [18] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM J. Numer. Anal.*, vol. 49, no. 6, pp. 2543–2563, Dec. 2011.
- [19] Z. Han, H. Li, and W. Yin, "Sparse optimization algorithms," in *Compressive Sensing for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2013, ch. 4, pp. 69–117.
- [20] Y. Ma, X. Zhang, and Y. Gao, "Joint sub-Nyquist spectrum sensing scheme with geolocation database over tv white space," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3998–4007, May 2018.
- [21] 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.
- [22] X. Zhang, Z. Qin, and Y. Gao, "Dynamic adjustment of sparsity upper bound in wideband compressive spectrum sensing," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2014, pp. 1214–1218.
- [23] M. E. Lopes, "Unknown sparsity in compressed sensing: Denoising and inference," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5145–5166, Sep. 2016.
- [24] M. Bkassiny, S. K. Jayaweera, Y. Li, and K. A. Avery, "Wideband spectrum sensing and non-parametric signal classification for autonomous self-learning cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2596–2605, Jul. 2012.
- [25] A. Toma, T. Nawaz, Y. Gao, L. Marcenaro, and C. Regazzoni, "Interference mitigation in wideband radios using spectrum correlation and neural network," *IET Commun.*, vol. 13, no. 10, pp. 1336–1347, Feb. 2019.
- [26] Z. Ye, G. Memik, and J. Grosspietsch, "Energy detection using estimated noise variance for spectrum sensing in cognitive radio networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2008, pp. 711–716.
- [27] W. Zhang, R. K. Mallik, and K. B. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5761–5766, Dec. 2009.
- [28] H. Qi, X. Zhang, and Y. Gao, "Channel energy statistics learning in compressive spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 7910–7921, Dec. 2018.
- [29] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 8, pp. 1190–1196, Aug. 1989.
- [30] Z. Lu and A. M. Zoubir, "Generalized Bayesian information criterion for source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1470–1480, Mar. 2013.
- [31] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.
- [32] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
- [33] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.
- [34] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1413–1426, Jun. 1995.
- [35] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.
- [36] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, Jun. 2007.
- [37] Z. Lu and A. M. Zoubir, "Source enumeration using the PDF of sample eigenvalues via information theoretic criteria," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 3361–3364.
- [38] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Stat.*, vol. 34, no. 1, pp. 122–148, 1963.
- [39] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1557–1567, Oct. 1989.
- [40] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.
- [41] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [42] A. Rakotomamonjy, "Algorithms for multiple basis pursuit denoising," in *Signal Processing With Adaptive Sparse Structured Representations*, R. Gribonval, Ed. Saint Malo, France: Inria Rennes-Bretagne Atlantique, Apr. 2009. [Online]. Available: <https://hal.inria.fr/inria-00369535>
- [43] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Mar. 2005, pp. v721–v724.
- [44] H. Qi, X. Zhang, and Y. Gao, "Subspace-aided low-complexity blind compressive spectrum sensing over TV whitespace," in *Proc. IEEE Global Commun. Conf.*, Dec. 2018, pp. 1–6.



Haoran Qi (S'16) received the B.Eng. degree from Beihang University, Beijing, China, and the M.Eng. degree (first class Hons.) from the University of Nottingham, Nottingham, U.K., in 2016. He is currently working toward the Ph.D. degree with the Department of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K. His research interests include spectrum sensing, cooperative cognitive radio networks, and sub-Nyquist signal processing.



Xingjian Zhang (S'16) received the B.Sc. degree (Hons.) in telecommunications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, and the Ph.D. degree in electronic engineering from the Queen Mary University of London, London, U.K., in 2014 and 2018, respectively. His research interests include cooperative wireless sensor networks, compressive sensing, real-time spectrum monitoring and analysis, and Internet of Things applications.



Yue Gao (S'03–M'07–SM'13) received the Ph.D. degree from the Queen Mary University of London (QMUL), London, U.K., in 2007. He was a Research Assistant, a Lecturer (Assistant Professor), and a Senior Lecturer (Associate Professor) with QMUL. He is currently a Reader in Antennas and Signal Processing, and the Director of the Whitespace Machine Communication Lab, School of Electronic Engineering and Computer Science, QMUL. He is leading a team developing fundamental research into practice in the interdisciplinary area on embedded artificial intelligence among smart antennas, signal processing, spectrum sharing, millimetre-wave, and Internet of Things systems. He has published more than 170 peer-reviewed journal and conference papers, two patents, one book, and five book chapters. He is an Engineering and Physical Sciences Research Council Fellow from 2018 to 2023. He was a Co-Recipient of the EU Horizon Prize Award on Collaborative Spectrum Sharing in 2016, and the Research Performance Award from the Faculty of Science and Engineering at QMUL, in 2017. He served as the Signal Processing for Communications Symposium Co-Chair for IEEE ICC 2016, the Publicity Co-Chair for the IEEE GLOBECOM 2016, the Cognitive Radio Symposium Co-Chair for the IEEE GLOBECOM 2017, and the General Chair of the IEEE WoWMoM and iWEM 2017. He is the Chair of the IEEE Technical Committee on Cognitive Networks and the IEEE Distinguished Lecturer of the Vehicular Technology Society. He is an Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE WIRELESS COMMUNICATIONS LETTERS, and *China Communications*.

tificial intelligence among smart antennas, signal processing, spectrum sharing, millimetre-wave, and Internet of Things systems. He has published more than 170 peer-reviewed journal and conference papers, two patents, one book, and five book chapters. He is an Engineering and Physical Sciences Research Council Fellow from 2018 to 2023. He was a Co-Recipient of the EU Horizon Prize Award on Collaborative Spectrum Sharing in 2016, and the Research Performance Award from the Faculty of Science and Engineering at QMUL, in 2017. He served as the Signal Processing for Communications Symposium Co-Chair for IEEE ICC 2016, the Publicity Co-Chair for the IEEE GLOBECOM 2016, the Cognitive Radio Symposium Co-Chair for the IEEE GLOBECOM 2017, and the General Chair of the IEEE WoWMoM and iWEM 2017. He is the Chair of the IEEE Technical Committee on Cognitive Networks and the IEEE Distinguished Lecturer of the Vehicular Technology Society. He is an Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE WIRELESS COMMUNICATIONS LETTERS, and *China Communications*.