An Adaptive Compressive Wideband Spectrum Sensing Algorithm Based on Least Squares Support Vector Machine

SHIYU REN, WANTONG CHEN, DONGXIA LI, AND CHENG FANG
Department of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China
Corresponding author: Shiyu Ren (renshyshl@163.com)

ABSTRACT Most of the compressive wideband spectrum sensing algorithms need to recover the spectrum, which require high computational complexity. Recently, a novel algorithm for compressive wideband sensing without spectrum recovery (NoR) was proposed. Its computational complexity is several orders of magnitude less than that of algorithms that need spectrum recovery. However, enabling by structure-constrained assumption of sparse spectrum, NoR may fail. In order to expand its scope of application while reducing the computational complexity as much as possible, we propose an adaptive sensing (ADP) algorithm that is a powerful hybrid of the no recovery and partial recovery (PR) algorithms. The ADP algorithm adaptively chooses the no recovery or partial recovery scheme depending on the situation learned by the least squares support vector machine (LS-SVM). By simulation and analysis, compared with NoR, PR and another excellent algorithm (orthogonal matching pursuit), the ADP suits better for practical applications.

INDEX TERMS Compressive wideband spectrum sensing, adaptive sensing, folded spectrum, no spectrum recovery, partial spectrum recovery, least squares support vector machine.

I. INTRODUCTION
Compressed sensing (CS) was first used by Tian and Georgos et al. [1] to realize wideband spectrum sensing under sub-Nyquist sampling. Since then, CS-based wideband spectrum sensing method has become a research hotspot. The most classical methods fall into two broad categories: non-convex optimization-based [2] and greedy pursuit-based [3]. However, CS-based wideband spectrum sensing requires that the wideband spectrum satisfies sparsity. Then another broad category of compressive sampling sensing method without sparsity constraint appears. That is multichannel sampling-based sensing method. And multichannel sampling scheme includes multicest uniform sampling [4]–[6] and multirate sampling [7]–[9]. All in all, most of the compressive wideband spectrum sensing algorithms can be attributed to the three categories just mentioned.

Recently, intelligent algorithms are introduced into spectrum sensing by more and more experts. Here are some examples. An artificial neural network (ANN)-based joint sensing algorithm combining energy detection and cyclostationary feature detection is proposed in [10]. This joint algorithm can promote the sensing accuracy and enhance the performance stability with ANN’s advantages. In [11], the support vector machine with particle swarm optimization (PSO-SVM) is used for obtaining a nonlinear threshold to replace the linear threshold used in traditional energy detection. And the performance of the proposed algorithm is much better than that of traditional energy detection. Thilina et al. adopt SVM, K-nearest-neighbor (KNN), Gaussian mixture model (GMM) and K-means clustering learning-based classification method for cooperative spectrum sensing (CSS) [12]. And comparative simulation results clearly reveal that the proposed methods outperform the existing state-of-the-art CSS techniques. In [13], Saber et al. propose a low power consumption and low cost implementation of spectrum sensing operation by using four machine learning (ML) algorithms: the artificial
neural networks (ANN), SVM, Decision Trees (TREE), and KNN. In [14], to addresses the problem of spectrum sensing under multiple primary users condition, the authors uses multi-class SVM. Simulation results show that the proposed scheme offers significant advantage in comparison with multi-class KNN. In [15], in order to draw the spectrum map of a large-scale heterogeneous network, Huang et al. exploit least squares support vector machine (LS-SVM) to learn and analyze the historical spectrum detection results and the corresponding location information. And the proposed scheme can draw a spectrum map with the accuracy 99.3% using only about 28% users performing spectrum sensing.

From the previous introduction, we can see that the use of intelligent algorithms can improve the performance of spectrum sensing algorithm detection or expand the application scenarios. So far, however, few intelligent algorithms have been used for compressive wideband spectrum sensing.

### A. MOTIVATION

Most compressive wideband sensing approaches need spectrum recovery, either signal recovery [1]–[3] or power spectrum recovery [4], [7], [8]. However, the above approaches that use highly non-linear reconstruction methods for spectrum recovery are computationally expensive.

To reduce computational complexity, we first develop a compressive sensing algorithm with partial spectrum recovery (PR) [5]. We perform sensing on a folded spectrum model brought by multicoset sub-Nyquist sampling. It consists of two steps. In the coarse sensing step, from the folded spectrum, the aliased sub-channels that contain active subband are picked out. In the fine sensing step, only those sub-bands that fold into the selected sub-channels need recovery. The PR algorithm’s average reduced radio of computational complexity can reach 50% compared with directly performing sensing by full spectrum recovery. Not satisfied with partial spectrum recovery, we develop a novel compressive sensing algorithm with no recovery (NoR) of the spectrum [6]. The NoR performs sensing on the same folded spectrum model as the PR. It also consists of two steps, and the coarse sensing step is the same with the PR. The difference lies in the fine sensing step. Based on a particular structural assumption of the folded spectrum, we consider there is at most one active subband in each aliased sub-channel. Then the active subband identification is performed by an inner-product maximization method instead of recovery. The NoR’s computational complexity is several orders of magnitude lower than the existing spectrum recovery algorithms. However, the assumed structure will only be present under certain conditions. The NoR works well in a favorable environment, but its sensing performance significantly degrades under hostile conditions.

Not satisfied with the NoR’s assumption-restricted sensing performance, we come up with adaptively switching between the subband identification methods of PR and NoR, according to the number of active subbands in each aliased sub-channel. Then the problem points toward aliased sub-channel detection, i.e., the coarse sensing step. The detection problem becomes a multi-classification problem, other than the previous binary classification problem. Spontaneously, the classical multiple hypothesis testing based energy detection method comes up. However, in our sensing scenario, the number of signal samples is compressed, and the power of noise is enlarged in the folded power spectrum. And if the number of signal samples for sensing is small and the signal-to-noise ratio (SNR) is low, the classification problem is linearly inseparable [11]. In this situation, the performance of multiple hypothesis testing based energy detection will decrease seriously. Therefore, we think of using SVM to exploit a nonlinear decision to replace the linear threshold obtained by hypothesis testing; since among all the above works [11]–[15], SVM based algorithm always has the best performance compared with the other mentioned ML algorithms. The SVM classifier has been shown to possess excellent classification capability. Therefore, we choose the improved model LS-SVM as the classifier in our work [16], [17].

The motivation of the proposed adaptive (ADP) algorithm is to address NoR’s problem of limited application scope and reduce the computational complexity as much as possible by adaptively choosing the NoR or PR scheme depending on the situation learned by the LS-SVM.

### B. CONTRIBUTIONS

The main contributions in this paper are as follows:

- Full spectrum recovery is avoided in the ADP algorithm. And this is the key for reducing the computational complexity as much as possible. As shown in FIG.1, the ADP algorithm performs sensing using the folded spectrum induced by multicoset sampling. Based on this folded spectrum structure, we split the algorithm into two steps: in the coarse sensing step, the aliased sub-channels that contain active subband are picked out; in the fine sensing step, the exact location of active subband is identified from the picked aliased sub-channels.

- Due to the excellent classification capability of LS-SVM, we use it for aliased sub-channel detection in the coarse sensing step. And by simulation, we confirm that the performance of LS-SVM based sub-channel detection scheme is much better than multiple hypothesis testing based scheme even when SNR is low and the number of samples is small. That proves the superiority of machine learning based detection scheme in the face of bad conditions.

- The ADP algorithm is the upgrade and extension of our previous NOR algorithm. The NoR’s computational complexity is several orders of magnitude lower than the existing NOR algorithm. But its performance is restricted by a particular structural assumption. To expand its scope of application, we adaptively use the subband identification methods of PR and NoR in the fine sensing step. Analysis and simulation confirm that compared with another three excellent compressive wideband spectrum sensing algorithms, the ADP is more suitable for practical applications.
The remainder of the paper is organized as follows. Section II establishes a folded spectrum model induced by multicoset sub-Nyquist sampling. In section III, the necessity for developing the ADP algorithm is explained. Then ADP’s components are introduced and analyzed. The sensing performance and computational complexity of ADP are evaluated in section IV. Conclusions are discussed in section V.

II. FOLDED SPECTRUM MODEL PROVIDED BY MULTICOSET SAMPLING

In this section, we introduce the multicoset sampling scheme. Then, the discrete-time samples obtained from the multicoset sampling scheme. The multicoset sampling employs $M$ parallel A/D cosets (or branches), and each coset samples the signal uniformly at a declined rate $\frac{1}{M}$, where $N$ is the down-sampling factor. Sub-Nyquist sampling can be obtained by $M < N$. Different cosets have different sampling offsets. We assume that $c_i T (c_i \in Z)$ is the offset of the $i$-th coset, and there is $0 \leq c_0 < c_1 < \cdots < c_{M-1} \leq N - 1$.

Practically, we perform wideband sensing in the digital domain. Then the discrete-time samples obtained from the $i$-th coset are:

$$y_i(n) = \begin{cases} x(nT), & n = kN + c_i, \\ 0, & \text{otherwise}, \end{cases}$$ (1)

where $F$ is the number of samples from each coset, and we let $U = FN$. Assume that $\bar{y}_i(k) = y_i(kN + c_i)$ is the down-sampled signal without the zero insertions. Then the $U$-point FFT of $\bar{y}_i(k)$ can be calculated by the $F$-point FFT of $y_i(k)$ [4]:

$$Y_i(k) = \left[ \frac{1}{\sqrt{U}} \sum_{l=0}^{F-1} \bar{y}_i(l) e^{-j 2\pi \frac{k l}{U}} \right] e^{-j 2\pi \frac{k c_i}{U}}, 0 \leq k \leq U - 1.$$ (2)

Note that the frequency resolution is $\Delta f = \frac{1}{FN} = \frac{1}{UT}$ and this corresponds to discretizing $[0, \frac{1}{N}]$ into $F$ grids. Then we can obtain the discrete spectrum of the wideband signal $X(k)$ as in [4]:

$$\begin{bmatrix} Y_0(k) \\ Y_1(k) \\ \vdots \\ Y_{M-1}(k) \end{bmatrix} = \sqrt{\frac{F}{N}} \begin{bmatrix} e^{j \frac{2\pi}{N} c_0 0} & e^{j \frac{2\pi}{N} c_0 1} & \cdots & e^{j \frac{2\pi}{N} c_0 (N-1)} \\ e^{j \frac{2\pi}{N} c_1 0} & e^{j \frac{2\pi}{N} c_1 1} & \cdots & e^{j \frac{2\pi}{N} c_1 (N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j \frac{2\pi}{N} c_{M-1} 0} & e^{j \frac{2\pi}{N} c_{M-1} 1} & \cdots & e^{j \frac{2\pi}{N} c_{M-1} (N-1)} \end{bmatrix}$$

$$\begin{bmatrix} X_0(k) \\ X_1(k) \\ \vdots \\ X_{N-1}(k) \end{bmatrix}, \quad k = 0, 1, \ldots, F - 1.$$ (3)

Fig.2 is the virtual model of the folded spectrum (3). In Fig.2, the discrete spectrum of the wideband $X(k) = \{X_n(k) | X_n(k) = X(k + nF), 0 \leq n \leq N - 1, 0 \leq k \leq F - 1\}$ is folded by $N$ to form the aliased spectrum $Y_i(k) = \{Y_i(k), 0 \leq k \leq F - 1\}$. So we also name $N$ as spectrum folding factor. Each cell $X_n(k)$ represents the $(nF + k)$-th subband in the whole wideband, with a bandwidth of $\frac{1}{NT}$. Each column $Y_i(k)$ represents an aliased sub-channel who is folded by $N$ subbands $\{X_0(k), X_1(k), \ldots, X_{N-1}(k)\}$, whose bandwidth is also $\frac{1}{NT}$. The colored cell means that a signal is transmitted on this subband, and we name it as “active (occupied) subband”.

**FIGURE 1.** The diagram of the ADP algorithm.

**FIGURE 2.** The virtual display of folded spectrum $Y_i(k)$. 

---

S. Ren et al.: Adaptive Compressive Wideband Spectrum Sensing Algorithm Based on LS-SVM
And we name the aliased sub-channel as “active (occupied) sub-channel”, if it contains any active subband.

III. ADAPTIVE SENSING ALGORITHM

To realize adaptive sensing, we first detect the situation of the aliased sub-channels, and then adjust the following subband identification scheme accordingly. Before elaborating on the adaptive sensing procedure, we first explain the reason for developing it.

A. NECESSITY FOR ADAPTATION

Most CS-based sensing algorithms need spectrum recovery with high computation. So we propose a novel algorithm NoR [6] with no spectrum recovery, whose computational complexity is several orders of magnitude lower. However, the realization of NoR is enabled by a structural assumption on the sparse spectrum. We introduce this particular structure named “ADFS” in the following.

**Definition 1:** Approximate Disjoint Folded Subband (ADFS). We define that the folded spectrum $X(k)$ owns an ADFS structure if there exists only a single occupied subband $X_n(k)$ in each active aliased sub-channel $Y_i(k)$. That is:

$$Y_i(k) \approx \sqrt{\frac{F}{N}} e^{\frac{2\pi i}{N} \gamma n} X_n(k). \quad (4)$$

The reflection of this structure in Fig.2 is that no more than two colored cells appear in a same column.

Assume that the sparsity order, i.e., the number of occupied subbands is $D$. Then according to [6], if we have $U \gg D$ and $F \sim O(\sqrt{U})$, the probability of occupied subband overlapping approaches to 0. Here, $F \sim O(\sqrt{U})$ means that $F$ is larger than or at least around $\sqrt{U}$. When $U \gg D$, $F \sim O(\sqrt{U})$, the ADFS can be satisfied.

When ADFS cannot be satisfied, the performance of NoR degrades severely. However, we do not want the computation-efficient sensing algorithm to be handicapped by the specific condition. Therefore, we further the low-computation sensing algorithm by the idea of an adaptive sensing procedure.

B. ALIASED SUB-CHANNEL DETECTION

This subsection introduces how to detect whether there is none, or one, or even more active subband in each sub-channel. Multiple hypothesis testing based energy detection is a typical and excellent detection method of solving the above problem. Its nature is a linear classifier. But its performance will decrease seriously, if the SNR of the received signal is low and the samples for sensing is small. Because in this situation the classification is linearly inseparable [11]. Looking forward to better performance, we study whether LS-SVM can classify this linearly inseparable problem efficiently. We propose both the Multiple hypothesis testing and the LS-SVM based aliased sub-channel detection method.

Then their sensing performance are compared in the following sections.

1) MULTIPLE HYPOTHESIS TESTING BASED SUB-CHANNEL DETECTION

As shown in Fig.2, in the folded spectrum obtained from each coset, there are $F$ aliased sub-channels $Y_i(k) = \{Y_i(0), Y_i(1), \ldots, Y_i(F-1)\}$. Each aliased sub-channel $Y_i(k)$ is folded by $N$ subbands $\{X_0(k), X_1(k), \ldots, X_{N-1}(k)\}$. The occupancy of each aliased sub-channel can be identified by solving the following multiple hypothesis testing problem:

- $H_0$: No primary signal on any of the $N - 1$ subbands
- $H_1$: Primary signal present on only one subband
- $H_2$: Primary signal present on at least two subbands

(5)

Assume that $X_i(k) = S_i(k) + N_i(k)$, and $S_i(k), N_i(k)$ are the discrete spectrum of the signal and noise, respectively. Suppose that the signals are zero-mean and independent, and the noise is stationary and white with power spectrum density $\sigma^2$. Define the set $\Omega = \{a \mid X_a(k) \text{ denotes an active subband} \}$. Furthermore, using (3), we can model the aliased sub-channel distribution by [14]:

$$\sqrt{\frac{N}{F}} Y_i(k) \sim \begin{cases} H_0: N(0, N\sigma^2) \\ H_1: N(X_a(k), N\sigma^2) \\ H_2: N\left(\sum_{a \in \Omega} X_a(k), N\sigma^2\right) \end{cases} \quad (6)$$

The energy of the aliased sub-channel at each sampling coset is:

$$E_i(k) = |Y_i(k)|^2, \quad k = 0, 1, \ldots, F - 1. \quad (7)$$

We set the test statistic as:

$$E[Y(k)] = \sum_{i=1}^{M} E_i(k), \quad k = 0, 1, \ldots, F - 1. \quad (8)$$

Using (6)(7)(8), we can model $E[Y(k)]$ as [18]:

$$\frac{1}{\sigma^2 F}E[Y(k)] \sim \begin{cases} H_0: \chi^2_{2M} \\
H_1: \chi^2_{2M}\left(\frac{2}{N} \gamma_a(k)\right) \\H_2: \chi^2_{2M}\left(\frac{2}{N} \sum_{a \in \Omega} \gamma_a(k)\right) \end{cases} \quad (9)$$

where $\gamma_a(k)$ denotes the signal-to-noise ratio (SNR) on the subband $X_a(k), \chi^2_{2M}$ denotes the central chi-square distribution, $\chi^2_{2M}\left(\frac{2}{N} \gamma_a(k)\right)$ and $\chi^2_{2M}\left(\frac{2}{N} \sum_{a \in \Omega} \gamma_a(k)\right)$ denote the non-central chi-square distribution. And these distributions have $2M$ degrees of freedom, $\frac{2}{N} \gamma_a(k)$ and $\frac{2}{N} \sum_{a \in \Omega} \gamma_a(k)$ are the noncentrality parameters. To test whether there is none, or one, or at least two active subbands in each sub-channel, we adopt the following decision rule:

$$E[Y(k)] \in \begin{cases} [0, \lambda_1), H_0 \\ [\lambda_1, \lambda_2), H_1 \\ [\lambda_2, +\infty), H_2 \end{cases} \quad (10)$$
According to [18], we have:

\[
P_{fa} = \sigma^2 F \left( \frac{M, \frac{1}{2}}{\Gamma(M)} \right),
\]

\[
P_{fa} = \sigma^2 F Q_M \left( \sqrt{\frac{2}{N}}, \sqrt{\lambda_2} \right),
\]

where \(P_{fa}\) is the probability of false alarm. And \(\Gamma(a, x)\) is the incomplete gamma function \(\Gamma(a, x) = \int_x^\infty t^{a-1}e^{-t}dt\), and \(\Gamma(n) = (n-1)!\). Besides, \(Q_u(a, x)\) is the generalized Marcum Q-function described by \(Q_u(a, x) = (1/a^u) \int_x^\infty t^{u-2}e^{-t}I_{u-1}(at)dt\), where \(I_v(a)\) is the \(v\)-th order modified Bessel function of the first kind. Once \(P_{fa}\) is set, we can obtain the values of both thresholds \(\lambda_1, \lambda_2\), it equals a linear classifier in multiple classification problem [11].

2) LS-SVM BASED SUB-CHANNEL DETECTION

In this subsection, the SVM model and its improved model LS-SVM are introduced successively. Then, the aliased sub-channel detection scheme using multiclass LS-SVM model is presented.

\(\alpha\): SVM

Assume \((x_i, y_i), i = 1, 2, \ldots, K\) is the training data set. A data \(x_i \in \mathbb{R}^d\) represents a vector in \(d\)-dimensional space. The aim of SVM is to find a hyper-plane which divides the data into two subspaces according to their corresponding labels \(y_i \in \{-1, 1\}\). For linear separable problems, the hyper-plane can be written as: \(w^T x_i + b = 0\), where \(w, b\) are the normal vector and displacement parameter of the hyper-plane respectively. The optimal hyper-plane can be obtained by maximizing the margin distance \(\gamma = \|w\|^{-1}\) between the two classes. Then quadratic programming is used to search the maximal \(\gamma\), leading to:

\[
\min_{\omega, b} \frac{1}{2}\|\omega\|^2
\]

\[
s.t. \quad y_i \left(\omega^T x_i + b\right) \geq 1, \quad i = 1, 2, \ldots, K.
\]

The above equation can be effectively solved by integrating positive Lagrange multipliers \(\alpha_i \geq 0 (i = 1, 2, \ldots, K)\), leading to:

\[
L(\omega, b, \alpha) = \frac{1}{2}\|\omega\|^2 + \sum_{i=1}^{K} \alpha_i \left(1 - y_i \left(\omega^T x_i + b\right)\right),
\]

where \(\omega\) can be solved by setting its partial derivative as zero, i.e., \(\omega = \sum_{i=1}^{K} \alpha_i y_i x_i\). So Eqn.(13) can be rewritten as:

\[
\max_{\alpha} \left\{ \sum_{i=1}^{K} \alpha_i - \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_i \alpha_j y_i y_j x_i^T x_j \right\}
\]

\[
s.t. \quad \sum_{i=1}^{K} \alpha_i y_i = 0,
\]

\[
\alpha_i \geq 0, \quad i = 1, 2, \ldots, K.
\]

Then \(\alpha\) can be derived from Eqn.(14). After that \(w, b\) can be obtained.

To deal with linear inseparable problems, a kernel function is adopted to map the low dimensional data into a feature space with high dimension. Then Eqn.(14) can be converted to:

\[
\max_{\alpha} \left\{ \sum_{i=1}^{K} \alpha_i - \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_i \alpha_j y_i y_j \kappa \left(x_i, x_j\right) \right\}
\]

\[
s.t. \quad \sum_{i=1}^{K} \alpha_i y_i = 0,
\]

\[
\alpha_i \geq 0, \quad i = 1, 2, \ldots, K.
\]

where \(\kappa (\cdot, \cdot)\) is the kernel function. As one of the most popular kernel functions, the radial basis function (RBF) kernel is considered in this paper:

\[
\kappa \left(x_i, x_j\right) = \exp \left(-\|x_j - x_i\|^2\right).
\]

where \(g\) is kernel parameter, and presenting the width of kernel function.

\(b\): LS-SVM

In SVM, the inequality constraints in Eqn.(12) require that all samples must be divided correctly. However, in real tasks, there may not be a hyperplane that can correctly divide the two types of samples in the original space. Therefore, the kernel function is used to map the samples from the original space to a higher-dimensional feature space, so that the samples are linearly separable in this feature space. However, this may produce overfitting problem.

In LS-SVM model, to avoid overfitting problem, a few classification errors are also tolerable. After replacing with flexible constraints, the classification problem can be rewritten as [19]:

\[
\min_{\omega, b} \left\{ \frac{1}{2}\|\omega\|^2 + \gamma \sum_{i=1}^{K} \left(\alpha_i y_i \left(\omega^T x_i + b\right) - 1\right) \right\}
\]

(17)

where \(\gamma\) denotes the tolerance degree of unmatched samples. \(\text{loss}_{0/1}\) represents the loss function. LS-SVM replaces the \(\text{loss}_{0/1}\) with square error \(e_i^2\), so that Eqn.(17) is modified as [19]:

\[
\min_{\omega, b} \left\{ \frac{1}{2}\|\omega\|^2 + \gamma \sum_{i=1}^{K} e_i^2 \right\}
\]

\[
s.t. \quad y_i = \omega^T x_i + b + e_i, \quad i = 1, 2, \ldots, K.
\]

The error variable \(e_i\) represents the tolerable classification error of the \(i\)-th training data.

Using error variable \(e_i\) makes the restriction in LS-SVM more flexible than in SVM. And the flexible restriction in LS-SVM can avoid the over fitting problem brought by the improper restriction in SVM. Also, the restriction in LS-SVM is equation, so the Lagrange multiplier can be obtained by solving linear equation. Thus the calculation in LS-SVM is more convenient and much faster than in SVM. Therefore,
The sub-channel detection scheme chooses LS-SVM as the classification model in this paper. After adopting the kernel function, the classification result of test data $x$ using LS-SVM is [19]:

$$\text{sgn} \left[ \sum_{i=1}^{K} \alpha_i y_i \kappa (x, x_i) + b \right].$$

(19)

**C. MULTICLASS LS-SVM BASED SUB-CHANNEL DETECTION SCHEME**

This scheme is to detect whether there is none, or one, or at least two active subbands in each sub-channel, which is a multiclass classification problem. SVM (LS-SVM) was originally designed for binary classification. Among the methods solving multiclass classification problem based on binary classification, “one-against-one” is more suitable for practical use than other methods [20]. So multiclass LS-SVM using one-against-one method is adopted in our scheme. As shown in Fig.3, this scheme encompasses two modules, i.e., Offline and Online.

- **Offline module**
  - Firstly, collect the training data. As mentioned above, the aliased sub-channels have three classes:

  \textit{Class 1:} No primary signal on any of the $N - 2$ subbands  
  \textit{Class 2:} Primary signal present on only one subband  
  \textit{Class 3:} Primary signal present on at least two subbands

  (20)

  Here, the data collected $\Omega^{(q)} = \{(x_p^{(q)}, y_p^{(q)}), q = 1, 2, 3, p = 1, 2, \ldots, K/3\}$ is the set consisting of energy of each sub-channel and its real classification result of the historical wideband signal, where $q$ represents the index of class and $K$ is the total number of training data from three classes.

  - Secondly, construct the training data set. Find all the possible sets that combine the data from two different classes, so that there forms totally three training data sets $T_{i,j}$ ($i < j, 1 \leq i, j \leq 3$).

  - Next, three classifiers are trained by using the three training data sets respectively. According to Eqn(19), the decision function $f_{i,j}(x)$, $i < j, 1 \leq i, j \leq 3$ of each classifier can be derived by using LS-SVM.

- **Online module**
  - Assume $x^{*} = \{x^{*}(k), 0 \leq k \leq F - 1\}$ as the energy vector of the sub-channels to be detected.

  - Next substitute $x^{*}(k)$ into the three decision functions. If $f_{i,j}(x^{*}(k)) = 1$, then $x^{*}(k)$ is decided as Class $i$ and Class $i$ gets a vote; or else $x^{*}(k)$ is decided as Class $j$ and Class $j$ gets a vote.

  - Finally, count the votes of three classes respectively. The class with the most votes is the classification result of the $k$-th sub-channel.

**C. SUBBAND IDENTIFICATION**

In this step, we identify the occupancy of subbands in each aliased sub-channel. For those aliased sub-channels that none active subband exists, there is no need of subband identification. If there are at least two active subbands in the aliased sub-channel, the partial power spectrum reconstruction scheme in PR is used. And if there is only active subband, the scheme is switched to subband classification without reconstruction in NoR.
1) PARTIAL POWER SPECTRUM RECONSTRUCTION

In this subsection, the power values of those subbands who fold into the active aliased sub-channel are reconstructed by exploiting the statistical properties of the spectrum.

Assume that \( \hat{r}_k(k) = E\{X_k(k)X_k^*(k)\}, n = 0, 1, \ldots, N-1, k = 0, 1, \ldots, E - 1 \) is the power value of the \((nF + k)\)-th subband. Let \( \hat{r}_k = [\hat{r}_k(0), \hat{r}_k(1), \ldots, \hat{r}_k(N-1)]^T \).

From (3), we can obtain the autocorrelation matrix of \( Y(k) \) by \( \hat{R}_Y(k) = E\{Y(k)Y^H(k)\} \). Define \( \text{vec}(\cdot) \) is the operation that \( \text{vec}(\Lambda) = [a_1^T, a_2^T, \ldots]^T \), where \( a_i^T \) is the \( i \)-th column of \( \Lambda \). And let \( \hat{r}(k) = \text{vec}(\hat{R}_Y(k)) \). Then, according to (4), it has been deduced that:

\[
\hat{r}_Y(k) = (C^* \otimes C) \hat{B}_X(k),
\]

in which \( (B)_{ij} = \begin{cases} 1, & i = (j-1)N + j, 1 \leq j \leq N, \\ 0, & \text{otherwise} \end{cases} \) and “\( \otimes \)” is the Kronecker product. Define \( \hat{P} = (C^* \otimes C) B \).

When \( P \) has full column rank, we can estimate the power value of each subband by:

\[
\hat{r}_X(k) = P^\dagger \hat{r}_Y(k),
\]

where “\( \dagger \)” denotes the pseudo inverse operation. The existence of \( P^\dagger \) can be decided by the sampling offsets \( c_0, c_1, \ldots, c_{M-1} \). Here, we eliminate its discussion.

Holding the set \( \kappa \), who stores the indexes of the active aliased sub-channels containing two or more active subbands, we only need to estimate the power values of the \((nF + h)\)-th subband. The occupancy of each subband can be detected by the following binary hypothesis testing problem [5]:

\[
H_0: \hat{r}_X(h) \sim N\left(\sigma^2, \nu^2_i\right) \\
H_1: \hat{r}_X(h) \sim N\left(P_X(h) + \sigma^2, \nu^2_i\right)
\]

where \( P_X(h) \) is the theoretical power value of the signal transmitted on the \((nF + h)\)-th subband, and \( \nu^2_i = \frac{1}{N} \nu^2_i \).

Once a constant false alarm rate \( P_{fa} \) is given, the detection threshold of the subband \( \nu_i \) can be obtained by:

\[
\nu_i = \sqrt{\nu^2_i Q^{-1}} (P_{fa} + \sigma^2).
\]

Then the occupancy of each subband can be detected: if \( \hat{r}_X(h) > \nu_i(h \in \kappa) \), the \((nF + h)\)-th subband is occupied; otherwise, all the other subbands are decided as unoccupied.

2) SUBBAND CLASSIFICATION WITHOUT RECONSTRUCTION

When there is only one active subband in the aliased sub-channel, the outputs corresponding to this sub-channel from different cosets have similar magnitude but different phases. Helping with the phase differences, the exact location of active subband in the aliased sub-channel can be identified by using the maximal inner product method.

Assume \( \Phi \) is the set, who stores the indexes of the active aliased sub-channels containing only one active subband.

And in each active aliased sub-channel, the optimal classification can be obtained by maximizing the absolute inner product between \( Y(h), h \in \Phi \) and the phase vector \( \rho(n) = [e^{2\pi i c_0 n}, e^{2\pi i c_1 n}, \ldots, e^{2\pi i c_{M-1} n}]^T \):

\[
\hat{n} = \arg \max_n \left| \sum_{i=0}^{M-1} Y_i(h) e^{-2\pi i c_i n} \right|^2.
\]

Then the \((\hat{n}F + \hat{n})\)-th subband is identified as the occupied one in the \( h \)-th sub-channel.

This classification result is unique so long as the \( N \) phase vectors \( \rho(n), n = 0, 1, \ldots, N - 1 \) are all different, and this can be fulfilled by as few as \( M = 2 \) cosets. Under ADF assumption, a better classification can be obtained by a lower coherence \( \lambda \) among different phase vectors, defined as:

\[
\lambda = \max_{n \neq n'} \frac{\|\rho(n) - \rho(n')\|^2}{\|\rho(n)\|^2 \cdot \|\rho(n')\|^2}.
\]

And the coherence \( \lambda \) can be reduced by using more sampling cosets.

IV. SIMULATION AND ANALYSIS

Wideband spectrum sensing experiments are performed in this section. These experiments fall into two categories. The ones in the first category confirm that using LS-SVM takes priority over hypothesis testing in the aliased sub-channel detection. Thus, LS-SVM based sub-channel detection scheme is finally adopted in our ADP algorithm. Then the other experiments in the second category confirm the efficiency of the ADP algorithm.

A. COMPARISON BETWEEN LS-SVM AND MULTIPLE HYPOTHESIS TESTING BASED ALIASED SUB-CHANNEL DETECTION

In this subsection, a wideband with range \([0, 960]\) MHz partitioned into \( U = 240 \) subbands is employed. And on each subband, no more than one primary user sends data. QPSK symbols are transmitted. First, to initially verify the feasibility of both the sub-channel detection schemes, we set the number of occupied subbands \( D = 6 \), the down-sampling factor \( N = 12 \), the compression ratio \( R = 2 \), the false alarm probability \( P_{fa} = 0.01 \). And the number of training data used in LS-SVM based scheme is set as \( K = 30, 150, 300 \), respectively.

The simulation results are presented in Fig.4. It is obvious that the sensing performance of LS-SVM based sub-channel detection scheme is much better than multiple hypothesis testing based scheme. Because when SNR is low and the number of samples is small, the detection problem is linearly inseparable, and the hypothesis testing based detection is not able to classify the linearly inseparable efficiently [11]. Moreover, even when the number of training data is as small as 30, the performance of LS-SVM based scheme is also good. The reason comes from an excellent character of SVM, i.e., its optimal hyper-plane is determined only by those training data.
which are close to it. In other words, most of the training data are inessential to draw the optimal hyper-plane [21].

Next, we compare the influence of compression ratio on sensing performance of the two schemes, since compressing the number of samples is the original intention of the sub-Nyquist sampling based wideband spectrum sensing. We set the compression ratio $N/M = 3, 2, 1.5$, respectively. The number of training data is $K = 150$. And the other parameters are the same with the above experiment. The results are presented in Fig.5. It is easy to find out LS-SVM based scheme also performances better than the other from the aspect of compression ratio. And the higher compression ratio is (the smaller number of samples is), the more obvious the advantage of LS-SVM based scheme is. Its reason has been explained in the analysis of above experiment.

In fact, according to IEEE802.22 standards, the spectrum detection probability of the cognitive radio users should be more than 90%. From the above simulation results, we can find the detection probability of multiple hypothesis testing based scheme is generally lower than 90% if $\text{SNR} \leq 3\text{dB}$. And this is just the probability of the aliased sub-channel detection. The occupied subbands have not been identified yet. Therefore, we adopt LS-SVM based sub-channel detection scheme in our ADP algorithm.

B. CONFIRMATION OF THE ADP ALGORITHM

In this subsection, we compare the performance of ADP, NoR, PR and orthogonal matching pursuit (OMP) algorithm.

1) SENSING PERFORMANCE

A wideband with range [0,1440] MHz partitioned into $U = 360$ subbands is employed. The locations of occupied subbands are selected randomly. We set the number of occupied subbands $D = 10$, the compression ratio $N/M = 2$, the false alarm probability $P_{fa} = 0.01$, and the down-sampling factor $N = 8, 12, 18, 24$ corresponding to $F = 45, 30, 20, 15$. And the number of training data used in ADP is set as $K = 150$. ADP, NoR and PR employ multicoset sub-Nyquist sampling. And OMP uses a compression matrix to acquire sub-Nyquist samples, whose size is decided by the number of subbands $U$ and compression ratio $N/M$ rather than $N$, $F$.

The results are shown in Fig.6. The sensing performance of OMP is fixed as reference because $U$ and $N/M$ are fixed. In Fig.6, the sensing performance of NoR, PR and ADP are becoming worse with the increase of the down-sampling factor $N$. Because higher $N$ means more subbands can be folded into one aliased sub-channel. This will lead to a much worse influence of noise on the aliased sub-channel detection. However, it is obvious that the performance of ADP is much better than NoR and PR when confronting a higher $N$. ADP uses LS-SVM for aliased sub-channel detection, while the other two use hypothesis testing. Its reason has already been analyzed elaborately. Moreover, the detection probability of NoR decreases significantly while $N$ increases. That’s because the ADFS assumption in NoR is difficult to be satisfied when $N$ is large. This means the probability of active subband overlap increases obviously. And that will directly
lead to subband classification errors in NoR. We also find that the performance of ADP is better than OMP.

2) COMPUTATIONAL COMPLEXITY
In this paragraph, we compare the computational complexity of ADP, NoR, PR, OMP algorithms.

To measure the computational complexity of the four algorithms, we use the number of complex float point operations [22]. For a concise analysis, we eliminate the computational complexity of Fourier transform in the algorithms. To guarantee fair comparison, we set the number of subbands in the monitored wideband, the total number of samples and the compression ratio of the four algorithms are the same.

For OMP algorithm, according to [3], [7], its computational complexity is $O \left( \frac{3}{2} \left( i \times b \right) (i+1) + \frac{2}{3} i \times b \times (i+1) (i+2) \right)$, where $i$ is the iterative times, $a$ is the number of acquired sub-Nyquist samples and $b$ is the number of monitored sub-bands. Generally $i$ is set as sparsity order $D$. So the computational complexity of OMP can be given by

$$CC_{OMP} = O \left( DNMF^2 + \frac{3}{2} D(D + 1)MF \right) + \frac{1}{3} D(D + 1)(2D + 1). \quad (27)$$

For NoR algorithm, in the aliased sub-channel detection step, to obtain the decision statistic $\hat{r}_y(k)$, there are $FM$-times multiplication of the complex-valued samples. Its complexity is $O \left( FM \right)$. And in the subband classification step, there are $DN$-time multiplication of a $1 \times M$ vector and an $M \times 1$ vector, the complexity of which is $O \left( DNM \right)$. Therefore, the computational complexity of NoR algorithm can be given by

$$CC_{NoR} = O \left( FM + DNM \right). \quad (28)$$

For PR algorithm, when we estimate the power values of $N$ subbands in each aliased sub-channel, the complexity of the aliased sub-channel detection step and subband identification step are $O \left( M \right)$ and $O \left( NM \right)$, respectively. Apparently the former is much lower, so we omit the computational complexity of aliased sub-channel detection for concise analysis. Assume $\nu$ to be the number of active subbands overlapped on a same sub-channel. The probability of each subband being occupied is $\rho = \frac{\nu}{N}$. Then the probability of aliased sub-channel containing any active subband can be given by

$$\eta = \Pr \left( \nu \geq 1 \right) = 1 - \Pr \left( \nu = 0 \right) = 1 - \left( 1 - \rho \right)^N. \quad (29)$$

Therefore, the number of active aliased sub-channels is $\eta F$, and the computational complexity of PR is

$$CC_{PR} = O \left( \eta FNM^2 \right). \quad (30)$$

For ADP algorithm, in the online module of the aliased sub-channel detection step, to obtain the classification result of each sub-channel, according to (16)(19), we firstly calculate the magnitudes of $K$ numbers with complexity of $O \left( K \right)$. And secondly, there is multiplication of a $1 \times K$ vector and a $K \times 1$ vector, the complexity of which is $O \left( K \right)$. Since there are $F$ sub-channels, the complexity of aliased sub-channel detection step is $O \left( 2FK \right)$. Next, in the subband identification step, the complexity is directly related to the detection results of the above step, i.e., the number of active subbands overlapped on a same sub-channel $\nu$. In ADP, we adaptively adopt the subband identification method in NoR and PR, when $\nu = 1$, or to the corresponding method in PR when $\nu \geq 2$. Following the analysis of the above paragraph, the probability of $\nu = 1$ and $\nu \geq 2$ can be given by

$$\eta_1 = \Pr \left( \nu = 1 \right) = C_N^1 \rho \left( 1 - \rho \right)^{N-1}, \quad (31)$$
$$\eta_2 = \Pr \left( \nu \geq 2 \right) = 1 - \Pr \left( \nu = 0 \right) - \Pr \left( \nu = 1 \right) = \eta - \eta_1. \quad (32)$$

Then, according to the complexity of subband identification step in NoR and PR, we can obtain the complexity of subband identification step in ADP is $O \left( \eta_1 FNM + (\eta - \eta_1) FNM^2 \right)$. Therefore, the computational complexity of ADP can be written as

$$CC_{ADP} = O \left( 2FK + \eta_1 FNM + (\eta - \eta_1) FNM^2 \right). \quad (33)$$

According to Eqn.(27)(28)(29)(30)(31)(33) and the simulation parameters, we compare the computational complexity of these algorithms in Fig.7. Apparently, OMP has the highest complexity while NoR has the lowest. And with the increase of the down-sampling factor $N$, ADP’s complexity gradually changes from greater to less than PR’s. Moreover, we have tested that the sensing performance of ADP is much better than OMP, NoR and PR when confronting a higher $N$. And a higher $N$ means a lower sampling rate for each sampler. That will bring us a lower implementation cost apparently.

Synthetically, considering the comparison of sensing performance and computational complexity among these algorithms, our ADP algorithm suits better for practical use.

V. CONCLUSION
In this paper, we have proposed a LS-SVM based adaptive compressive wideband spectrum sensing algorithm. The ADP algorithm performs sensing on the folded spectrum provided...
by multicoset sampling. It involves a coarse sensing to classify the aliased sub-channels by using LS-SVM, and then a fine sensing to identify exact locations of occupied subbands. In particular, we have provided two schemes for subband identification, which will be adaptively chosen according to the classification result of aliased sub-channels. Considering the sensing performance and computational complexity synthetically, the ADP algorithm is more suitable for practical applications compared with another three excellent compressive wideband spectrum sensing algorithms. And in our future work, we will discuss whether the ADP can work in dynamic scenarios with varying noise variance [23].

REFERENCES

[1] Z. Tian and G. B. Giannakis, “Compressed sensing for wideband cognitive radios,” in Proc. IEEE Int. Conf. Acoust., Apr. 2007, pp. 1357–1360.

[2] Y. Wang, J. Wang, and Z. Xu, “On recovery of block-sparse signals via mixed \( l_2 / l_1 \) norm minimization,” EURASIP J. Adv. Signal Process., vol. 2013, no. 1, 2013, Art. no. 76.

[3] W. Nasar, A. R. Maud, H. Wang, and H. Dia, “An efficient greedy algorithm for wide band spectrum sensing in cognitive radio networks,” in Proc. IEEE Int. Conf. Green Comput. Commun., Jul./Aug. 2018.

[4] S. Ren, Z. Zeng, C. Guo, and X. Sun, “A low computational complexity algorithm for adaptive wideband spectrum sensing,” IEICE Trans. Fundam. Electron., Commun. Comput. Sci., vol. E100.A, no. 1, pp. 294–300, 2017.

[5] S. Ren, Z. Zeng, C. Guo, and X. Sun, “A low complexity sensing algorithm for wideband sparse spectra,” IEEE Commun. Lett., vol. 21, no. 1, pp. 92–95, Jan. 2017.

[6] S. Ren, Z. Zeng, C. Guo, and X. Sun, “Wideband spectrum sensing based on corpmise sampling,” in Proc. 22nd Int. Conf. Telecommun. (ICT), Sydney, NSW, Australia, Apr. 2015, pp. 348–352.

[7] X.-L. Huang, Y. Gao, X.-W. Tang, and S.-B. Wang, “Spectrum mapping of cognitive radio networks,” IEEE Trans. Signal Process., vol. 37, no. 8, pp. 2098–2191, Nov. 2013.

[8] S. Ren, Z. Zeng, C. Guo, and X. Sun, “A low-computation compressive wideband spectrum sensing algorithm based on multirate corpmise sampling,” IEICE Trans. Fundam. Electron., Commun. Comput. Sci., vol. E100.A, no. 4, pp. 1060–1065, 2017.

[9] Y.-J. Zhang, B.-D. He, X.-H. Fu, J.-R. Xu, and K.-P. Zhou, “Raman spectra combined with PSO-LSSVM algorithm for detecting the components in ternary blended edible oil,” Spectros. Spectral Anal., vol. 37, no. 8, pp. 2440–2445, 2017.

[10] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415–425, Mar. 2002.

[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, Nov. 2016, pp. 140–142.

[12] Z. Hu, C. Feng, T. Zhang, Q. Niu, and Y. Chen, “Nonlinear joint transmit-receive processing for coordinated multi-cell systems: Centralized and decentralized,” EURASIP J. Adv. Signal Process., vol. 2015, no. 1, pp. 1–14, Dec. 2015.

[13] DOXIA LI received the M.S. and Ph.D. degrees in electrical engineering from Beijing Institute of Technology (BIT), Beijing, in 1999 and 2002, respectively. She is currently an Associate Professor with the College of Electronic Information and Automation, Civil Aviation University of China (CAUC), Tianjin, China. Her current research interests include aeronautical telecommunication networks and aeronautical mobile communication systems.

[14] CSIAO CHENG received the B.E. degree in information and communication engineering from Shandong University, China, in 2011, and the Ph.D. degree in information and communication engineering from Beijing University of Posts and Telecommunications (BUPT), China, in 2017. She is currently a Lecturer with the Department of Electronic Information and Automation, Civil Aviation University of China (CAUC), Tianjin, China. Her research interests include wireless communications and networks, with an emphasis on cognitive radios and wideband spectrum sensing.

[15] S. Ren, Z. Zeng, C. Guo, and X. Sun, “A low-computation compressive wideband spectrum sensing algorithm based on multirate corpmise sampling,” IEICE Trans. Fundam. Electron., Commun. Comput. Sci., vol. E100.A, no. 4, pp. 1060–1065, 2017.

[16] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415–425, Mar. 2002.

[17] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, Nov. 2016, pp. 140–142.

[18] Z. Hu, C. Feng, T. Zhang, Q. Niu, and Y. Chen, “Nonlinear joint transmit-receive processing for coordinated multi-cell systems: Centralized and decentralized,” EURASIP J. Adv. Signal Process., vol. 2015, no. 1, pp. 1–14, Dec. 2015.

[19] D. R. Joshi, D. C. Popescu, and O. A. Dobre, “Adaptive spectrum sensing with noise variance estimation for dynamic cognitive radio systems,” in Proc. 44th Annu. Conf. Inf. Sci. Syst., Mar. 2010, pp. 1–5.

[20] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415–425, Mar. 2002.

[21] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, Nov. 2016, pp. 140–142.

[22] Z. Hu, C. Feng, T. Zhang, Q. Niu, and Y. Chen, “Nonlinear joint transmit-receive processing for coordinated multi-cell systems: Centralized and decentralized,” EURASIP J. Adv. Signal Process., vol. 2015, no. 1, pp. 1–14, Dec. 2015.

[23] D. R. Joshi, D. C. Popescu, and O. A. Dobre, “Adaptive spectrum sensing with noise variance estimation for dynamic cognitive radio systems,” in Proc. 44th Annu. Conf. Inf. Sci. Syst., Mar. 2010, pp. 1–5.