IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 19, 2020, accepted October 29, 2020, date of publication November 9, 2020, date of current version November 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3036656

Classification-Based One-Bit DOA Estimation for Sparse Arrays

YANPING CHEN[1], CHEN WANG[2], AND YULONG GAO[2], (Member, IEEE)
[1]School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China
[2]Department of Communication Engineering, Harbin Institute of Technology, Harbin 150001, China

Corresponding author: Yulong Gao (ylgao@hit.edu.cn)

This work was supported in part by the National Nature Science Foundation of China (NSFC) under Grant 61671176, and in part by the Civil Space Pre-research Program during the 13th Five-Year Plan under Grant B0111.

ABSTRACT Nowadays, high-resolution DOA estimation techniques have been widely used in many fields such as sonar, communication, radio astronomy and biomedicine etc. Unfortunately, conventional algorithms cannot achieve high precision due to the limitation of the array aperture, and usually suffer from high computational complexity. In this paper, we innovatively model the DOA estimation problem of incoherent signals as a binary classification problem and greatly reduce physical complexity by virtue of sparse arrays. We first propose a classification framework for DOA estimation by benefiting from one-bit quantification. In this framework, any classification algorithm can be exploited to estimate DOA, such as logistic regression used in this paper. And then, sparse array is considered to reduce the exceeding number of antennas. Moreover, an iterative grid refinement procedure is presented to achieve more accurate DOA estimation and further reduces the actual number of physical antennas. Simultaneously, we derive Cramer-Rao bound (CRB) for the proposed algorithm. Finally, simulations are conducted for correctness and validation and the results illustrate the significant performance and reduction of complexity in hardware and computation over the existing methods.

INDEX TERMS DOA estimation, classification framework, one-bit quantification, logistic regression, sparse array, grid refinement.

I. INTRODUCTION

Direction of arrival (DOA) estimation is a classic problem in array signal processing, which exploits a specific array to estimate angle of signals from the external environment. Traditional DOA estimation algorithms employ the orthogonality of signal subspace and noise subspace to estimate DOA, such as multiple signal classification (MUSIC) [1] and estimation of signal parameters via rotational invariance technique (ESPRIT) [2]. However, the subspace-based algorithms usually require a large number of sampled data points to obtain a more accurate estimate, and the performance is usually not very satisfactory for the low signal-to-noise ratio(SNR) case.

Due to spatial sparsity, compressed sensing(CS) has attracted considerable interests in DOA estimation in recent years. The inspiration of the CS-based DOA estimation algorithm comes from that considering the extended array manifold matrix as a measurement matrix and the extended signal as a sparse vector, then applying traditional CS recovery algorithms for DOA estimation, such as convex optimization algorithm [3], greedy algorithm [4], [5], and sparse Bayesian algorithm [6]. However, these methods imply the operation of grid division on the angles, which inevitably leads to picket fence effect in DOA estimation. At present, there are two kinds of ideas used to solve the picket fence effect caused by discretization in angular coordinate. One is modeling DOA in the discrete parameter space and utilizing Taylor series expansion [6] or linear interpolation model [7] to establish the relationship between angle correction and observations, and finally employing sparse Bayesian learning algorithm or other methods to estimate it. The second idea is to model DOA directly in the continuous space. For instance, the atomic norm defined in the continuous space [8], which guarantees spatial continuity and sparsity of signal parameters, and thereby avoids the picket fence effect.

In the future communication system, analog-to-digital converter(ADC) is one of the key devices for digital signal

The associate editor coordinating the review of this manuscript and approving it for publication was Hasan S. Mir.

processing. In order to reduce the computational complexity caused by the ADC with large quantization bit number, one-bit quantization has been gaining more momentum. Especially in DOA estimation, it has proved that the loss caused by one-bit quantization is relatively small [9], thus, a series of one-bit DOA estimation algorithms have emerged. For instance, Shalom and Weiss [10] derived the covariance matrix of one-bit measurements by using the arcsine theorem of statistical theory and then achieved accurate DOA estimation. Other improved algorithms, can be found in [11], [12] and [13]. Unfortunately, a massive number of physical antennas is still need to obtain high estimation accuracy. To deal with this problem, Liu and Vaidyanathan [14] proposed a one-bit sparse array DOA estimation method, which shows that sparse arrays such as nested arrays [15] and coprime arrays [16] are more robust to the deleterious effects of one-bit quantization compared to uniform linear arrays (ULAs). In [17], Chen et.al. introduced a compression matrix to extend the receive array aperture and finally achieved DOA estimation by using compressive one-bit measurements.

Actually, the result of one-bit quantization is producing two types of samples such as $+1$ and -1. For a linear expression of one-bit measurement vector model $\mathbf{q} = \text{sign}(\boldsymbol{\Phi t} + \mathbf{e})$ [18], we can obtain DOAs by the indexes of the non-zero entries of the $\mathbf{t}$, where $\mathbf{q}$, $\boldsymbol{\Phi}$, $\mathbf{t}$ and $\mathbf{e}$ represent the one-bit measurement vector, the expended array manifold matrix, the expended signal vector and the noise vector respectively. Just change our perspective, each row of the dictionary matrix $\boldsymbol{\Phi}$ can be regarded as a sample input, the sparse vector $\mathbf{t}$ can be seen as the weight vector and each quantized data $\mathbf{q}$ can be regarded as the label to be classified. Then, this sparse reconstruction problem is equivalent to a process of training parameters for a linear classification problem. Once we choose a suitable classification algorithm, the DOA estimation problem will be well solved. Whereas, the algorithm proposed in [18] suffers from massive antennas to achieve high accuracy.

In this paper, we propose a novel classification-based one-bit DOA estimation framework for sparse array, which combines the advantages of one-bit quantization and sparse array together. The algorithm takes an arbitrary sparse array as the receiver, then the degrees of freedom (DOFs) is expected to get increased by exploiting the auto-correlation information. Therefore, the number of the physical antennas can be significantly reduced. Then, by considering the one-bit observations as labels to be classified and applying classification algorithm such as logistic regression algorithm(LR) to solve the linear classifier $\mathbf{q} = \text{sign}(\boldsymbol{\Phi t} + \mathbf{e})$, in this way, we can get a rough estimation of DOAs. Due to the unavoidable picket fence effect introduced in the proposed structure, an iterative grid refinement procedure is presented to achieve more accurate DOA estimation and further reduce the number of antennas.

The main contributions of this paper are threefold as follows. First, we propose a unified framework based on classification for DOA estimation of incoherent signals of sparse array, any classification algorithm is available for the framework. Second, we have reduced physical complexity without accuracy loss. By exploiting the auto-correlation information of the observations from sparse arrays, the degrees of freedom(DOF) is highly increased compared to ULAs, and the robustness to the deleterious effects of one-bit quantization is enhanced. Moreover, the iterative grid refinement procedure proposed is under a low dimensional convex optimization process, thus, the estimation accuracy is further improved while reducing computational complexity and dependence on array size. Third, we also discuss the Cramer-Rao bound (CRB) for the algorithm, which provides some insights into the unbiasedness of DOA estimation. And a detailed analysis of complexity of the algorithm is conducted in Section IV.

The rest of this paper is organized as follows. Section II introduce classification-based DOA estimation unified framework. Section III study logistic regression algorithm in DOA estimation and then develop a classification-based one-bit off-grid DOA estimation algorithm for sparse array. Section IV presents the experimental results. Section V gives the conclusions.

The notations used in this paper are as follows. We use the lower-case letter (e.g., a), lower-case bold letter (e.g., $\mathbf{a}$), upper-case bold letter (e.g., $\mathbf{A}$) and upper-case letters in blackboard boldface (e.g., $\mathbb{A}$) to represent the scalars, vectors, matrices and sets respectively. The superscripts -1, T and H denote the inverse, the transpose and the complex conjugate transpose of a matrix respectively. In addition, we use $vec\,(\cdot)$, $sign\,(\cdot)$ and $E\,(\cdot)$ to represent the vectorization, one-bit quantization and expectation operations. We also use $\Re\,(\mathbf{A})$ and $\Im\,(\mathbf{A})$ to represent the real and imaginary parts of matrix $\mathbf{A}$. The notation $\otimes$ and $\odot$ represent the Kronecker product and Khatri-Rao product (column-wise Kronecker product). For instance, the Khatri-Rao product between two matrices $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_a]$ and $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_a]$ is given as $\mathbf{M} \odot \mathbf{N} = [\mathbf{m}_1 \otimes \mathbf{n}_1, \ldots, \mathbf{m}_a \otimes \mathbf{n}_a]$. Moreover, $tr\,(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$, $\mathbf{I}$ denotes the identity matrix. $j = \sqrt{-1}$ is the unit imaginary.

II. PROBLEM FORMULATION OF ONE-BIT DOA ESTIMATION

Consider K uncorrelated far-field narrowband signals with the directions of $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_K\}$ impinging onto an M-element sparse array with locations $\{d_1 d, d_2 d, \ldots, d_M d\}$, where d is half of the wavelength λ and $\mathbb{S} = \{d_1, d_2, \ldots, d_M\}$ is an integer set indicating the physical antenna positions. Without loss of generality, we set $d_1 = 0$. Then, the received signal vector under the l-th observation snapshot can be expressed as

$$\mathbf{x}_{\mathbb{S}} = [a_{\mathbb{S}}(\bar{\theta}_1), a_{\mathbb{S}}(\bar{\theta}_2), \ldots, a_{\mathbb{S}}(\bar{\theta}_K)]\mathbf{s}_l + \mathbf{n}_l = \mathbf{A}_{\mathbb{S}}\mathbf{s}_l + \mathbf{n}_l \quad (1)$$

where $a_{\mathbb{S}}(\bar{\theta}_k) = \left[e^{j2\pi d_1 \bar{\theta}_k}, \ldots, e^{j2\pi d_M \bar{\theta}_k}\right]^T$ is the steering vector of the $k-$th incident signal and $\bar{\theta}_k = (d/\lambda)\sin\theta_k$ represents the normalized DOA of the k-th incident signal with $k = 1, 2, \ldots, K$, $\mathbf{A}_{\mathbb{S}}$ denotes the manifold matrix in

terms of $\mathbb{S}$, $\mathbf{s}_l \in \mathbb{C}^K$ denotes the complex amplitude vector of the sources with power $\{{\sigma_1}^2, {\sigma_2}^2, \ldots, {\sigma_K}^2\}$. In addition, $\mathbf{n}_l \in \mathbb{C}^M$ denotes the uncorrelated Gaussian noise with mean 0 and variance ${\sigma_n}^2$. The covariance matrix of $\mathbf{x}_\mathbb{S}$ can be expressed as

$$\mathbf{R}_\mathbb{S} = E[\mathbf{x}_\mathbb{S}{\mathbf{x}_\mathbb{S}}^H] = \sum_{i=1}^{K} \sigma_i^2 \mathbf{a}_\mathbb{S}(\bar{\theta}_i){\mathbf{a}_\mathbb{S}}^H(\bar{\theta}_i) + \sigma_n^2\mathbf{I} \tag{2}$$

where σ_i^2 represents the power of the i-th signal. It is noted that $\mathbf{n}_l$ is assumed to be uncorrelated at different antennas, the yielding covariance matrix of $\mathbf{n}_l$ is thus diagonal and is described as $\sigma_n^2\mathbf{I} \in \mathbb{C}^{M\times M}$.

Then we can construct the correlation vector $\mathbf{x}_\mathbb{D}$ on the difference co-array by vectorizing and combining the duplicate entries in (2)

$$\begin{aligned}\mathbf{x}_\mathbb{D} &= \mathrm{vec}(\mathbf{R}_\mathbb{S}) \\ &= \sum_{i=1}^{K} \sigma_i^2 \mathbf{a}_\mathbb{D}(\bar{\theta}_i) + \sigma_n^2 \mathrm{vec}(\mathbf{I}) \\ &= \mathbf{J}\mathbf{A}_\mathbb{D}\mathbf{p} + \sigma_n^2 \mathrm{vec}(\mathbf{I}) \\ &= [\mathbf{J}\mathbf{A}_\mathbb{D}\ \mathrm{vec}(\mathbf{I})][\mathbf{p}\sigma_n^2]^T \end{aligned} \tag{3}$$

where $\mathbf{p} = [\sigma_1^2, \sigma_2^2, .., \sigma_K^2]$, $\mathbf{A}_\mathbb{D} = \left[\mathbf{a}_\mathbb{D}(\bar{\theta}_1), \ldots, \mathbf{a}_\mathbb{D}(\bar{\theta}_K)\right]^T$ denotes the corresponding manifold matrix of the difference co-array, $\mathbf{a}_\mathbb{D}(\bar{\theta}_k) = [\mathrm{e}^{\mathrm{j}2\pi d_1\bar{\theta}_k}, \ldots, \mathrm{e}^{\mathrm{j}2\pi d_{|\mathbb{D}|}\bar{\theta}_k}]$ is the steering vector of the co-array associated with the k-th source, $|\mathbb{D}|$ represents the number of different elements in the difference set $\mathbb{D}$.

Definition 1 (Difference co-array) [17]: For an array specified by an integer set $\mathbb{S}$, the difference set $\mathbb{D}$ is defined as

$$\mathbb{D} = \{n_1 - n_2 \,|\forall n_1, n_2 \in \mathbb{S}\} \tag{4}$$

It is worth noting that in the **Definition 1**, it allows repetition of its elements. Assume that the set $\mathbb{D}_u$ consists of the distinct elements of the difference set $\mathbb{D}$, then the elements in the difference co-array $\mathbb{D}_u$ directly decides the distinct values of the cross correlation terms in the covariance matrix of the observations. Specifically, the relationship between $\mathbf{A}_\mathbb{S}$ and $\mathbf{A}_\mathbb{D}$ can be expressed as

$$\mathbf{A}_\mathbb{S}^* \odot \mathbf{A}_\mathbb{S} = \mathbf{J}\mathbf{A}_\mathbb{D} \tag{5}$$

where the binary matrix $\mathbf{J}$ is defined as:

Definition 2 (The binary matrix): The binary matrix $\mathbf{J} \in \mathbb{C}^{|\mathbb{S}|^2\times|\mathbb{D}|}$, and the i-th column of $\mathbf{J}$ satisfy

$$\langle\mathbf{J}\rangle_{:,i} = vec\left(\mathbf{I}\left(i\right)\right), \quad i \in \mathbb{D} \tag{6}$$

$$\langle\mathbf{I}\left(i\right)\rangle_{n_1,n_2} = \begin{cases} 1, & \text{if } n_1 - n_2 = i \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

Then the model for $\mathbf{x}_\mathbb{D}$ form a ULA of higher DOFs, which executes like a series of the "virtual sensors" (given by $\mathbb{D}_u$). Then we can apply traditional DOA estimation algorithms, such as SS-MUSIC in [15], [16], to solve DOA estimation problem in case of more sources than the number of arrays.

In our algorithm, we only use a part of the "virtual sensors" which consists of the non-negative elements of $\mathbb{D}_u$. For instance, if a nested array has $N_1 + N_2$ sensors whose locations is $\mathbb{S}_{\mathrm{ne}} = \{1, \ldots, N_1, (N_1+1), N_2(N_1+1)\}$, the difference co-array $\mathbb{D}_u$ can be expressed as $\mathbb{D}_u = \{0, \pm1, \ldots, \pm(N_2(N_1+1)-1)\}$ [15], then the locations of virtual array in our algorithm can be described as $\mathbb{U} = \{0, 1, \ldots, (N_2(N_1+1)-1)\}$, and the observation is corresponding to $\mathbf{x}_\mathbb{U}$. More discussion on the performance of sparse arrays will be conducted in Section IV.

Usually, the observation result of the "virtual sensors" under L snapshots can be expressed as

$$\mathbf{X}_\mathbb{U} = vec[\frac{1}{L}\sum_{l=1}^{L} \mathbf{x}_\mathbb{S}{\mathbf{x}_\mathbb{S}}^H] \tag{8}$$

Note that $\mathbf{X}_\mathbb{U}$ is the high-dimensional single snapshot data and the steering matrix $\mathbf{A}_\mathbb{D}$ contains a series of single frequency complex sinusoid signal, we can construct an over-complete dictionary $\mathbf{F}$ to further express (8) as

$$\mathbf{X}_\mathbb{U} = \mathbf{F}\mathbf{P} + \mathbf{N} \tag{9}$$

where $\mathbf{F} \in \mathbb{C}^{N\times N}$ whose $(i,\ k)\ th$ element is $e^{j(i-1)(k-1)2\pi/N}$, $\mathbf{P} \in \mathbb{C}^N$ denotes the zero-padded expansion of the original signal power, $\mathbf{N} \in \mathbb{C}^N$ denotes the noisy vector, and N is the number of the virtual array. Formulation (9) implies that once we recover the sparse vector $\mathbf{P}$, we can index the spatial angle and obtain DOAs on grid.

Only reserving the sign information of the measurements, the data model based on one-bit quantization under the sparse array can be expressed as

$$\widehat{\mathbf{R}}_\mathbb{S} = \mathrm{csgn}(\Re\{\mathbf{X}_\mathbb{U}\}) + j\mathrm{csgn}(\Im\{\mathbf{X}_\mathbb{U}\}) \tag{10}$$

where

$$\mathrm{csgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leqslant 0 \end{cases} \tag{11}$$

We convert the quantized data model into a real-number case by constructing several real-number matrices

$$\begin{aligned} \mathbf{q} &= \begin{bmatrix} \Re\left\{\widehat{\mathbf{R}}_\mathbb{S}\right\} \\ \Im\left\{\widehat{\mathbf{R}}_\mathbb{S}\right\} \end{bmatrix} \quad \mathbf{\Phi} = \begin{bmatrix} \Re\{\mathbf{F}\} & -\Im\{\mathbf{F}\} \\ \Im\{\mathbf{F}\} & \Re\{\mathbf{F}\} \end{bmatrix} \\ \mathbf{t} &= \begin{bmatrix} \Re\{\mathbf{P}\} \\ \Im\{\mathbf{P}\} \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} \Re\{\mathbf{N}\} \\ \Im\{\mathbf{N}\} \end{bmatrix} \end{aligned} \tag{12}$$

where $\mathbf{q} \in \{0,1\}^{2N}$, $\mathbf{\Phi} \in R^{2N\times2N}$, $\mathbf{t} \in R^{2N}$ and $\mathbf{e} \in R^{2N}$ denote the measurements, the measurement matrix, the signal power and the noise respectively. Finally, the classification-based one-bit DOA estimation data model can be expressed as

$$\mathbf{q} = \mathrm{csgn}(\mathbf{\Phi}\mathbf{t} + \mathbf{e}) \tag{13}$$

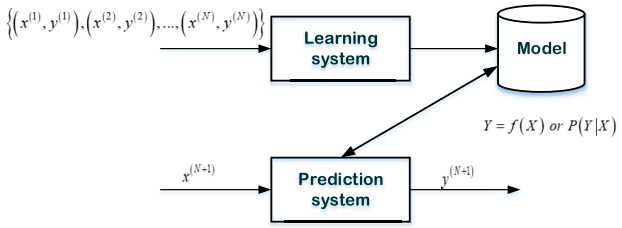


FIGURE 1. Supervised learning problem.

III. THE DESCRIPTION OF THE PROPOSED METHOD

A. DOA ESTIMATION BASED ON CLASSIFICATION

Classification is a core issue in supervised learning problem. As shown in Fig.1, classic classification problem usually involves two processes, namely, learning and classification. During the learning process, an effective learning method is used to learn a classifier ($Y = \widehat{f}(X)$ or $\widehat{P}(Y|X)$) based on the training set $\mathbb{T}_{train} = \left\{\left(x^{(1)}, y^{(1)}\right), \left(x^{(2)}, y^{(2)}\right), \ldots, \left(x^{(N)}, y^{(N)}\right)\right\}$, then the learned classifier is used to predict and classify the new samples($x^{(N+1)}$) during the classification process. Perceptron, as a special classifier, is a two-class linear classification model. Assume the feature space of the sample is $\mathcal{X} \in \mathbb{R}^N$, the output space is $\mathcal{Y} = \{+1, -1\}$, the mode of a perceptron can be described as

$$f(x) = sign(\omega \cdot x + b) \tag{14}$$

where $\omega \in \mathbb{R}^N$ and $b \in \mathbb{R}$ denote the weight and bias respectively, $x \in \mathcal{X}$ denotes the feature vector of the sample, and the symbolic function $sign(x)$ can be described as

$$sign(x) = \begin{cases} +1, & x \geqslant 0 \\ -1, & x < 0 \end{cases} \tag{15}$$

It can be found that the one-bit DOA estimation data model (13) and classification model (14) are mathematically quite similar. Actually, (13) is a linear expression where the measurement in each array element has been quantized as $+1$ or 0. Each quantized data can be regarded as an outcome from the linear classifier and each array element corresponds to two training samples containing the real and imaginary part information of the array measurements. In this way, by solving the optimal classification coefficient $\mathbf{t}$, that is, the sparse signal vector $\mathbf{t}$, the source signals location can be ascertained from the position of the non-zero elements in $\mathbf{t}$, which also helps keep a basis for the subspace to estimate what sparse combinations of columns of $\mathbf{F}$ form it. Just take LR algorithm in machine learning as an example to solve this binary classification problem, which is described in detail as follows.

Prior: Classic logistic regression model is usually defined as the following conditional probability distribution

$$P(Y = 0|x) = \frac{1}{1 + e^{\boldsymbol{\omega} \cdot x + b}} \tag{16}$$

$$P(Y = 1|x) = \frac{e^{\boldsymbol{\omega} \cdot x + b}}{1 + e^{\boldsymbol{\omega} \cdot x + b}} \tag{17}$$

where x and b denote the parameter of input sample and bias, $\boldsymbol{\omega}$ denotes weighting factor. For a given training set $\mathbb{T}_{train} = \left\{\left(x^{(1)}, y^{(1)}\right), \left(x^{(2)}, y^{(2)}\right), \ldots, \left(x^{(N)}, y^{(N)}\right)\right\}$, LR is actually a parameter estimation algorithm based on the maximum likelihood criterion. Assume $P(Y = 1|x) = \pi(x)$, $P(Y = 0|x) = 1 - \pi(x)$, the likelihood function of the input samples can be expressed as

$$\prod_{i=1}^{N} \left[\pi\left(x^{(i)}\right)\right]^{y^{(i)}} \left[1 - \pi\left(x^{(i)}\right)\right]^{1-y^{(i)}} \tag{18}$$

By taking the logarithm of (18), we can get the corresponding log-likelihood function

$$\begin{aligned} L(\boldsymbol{\omega}) &= \sum_{i=1}^{N} \left[y^{(i)} \log \pi\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - \pi\left(x^{(i)}\right)\right)\right] \\ &= \sum_{i=1}^{N} \left[y^{(i)} \log \frac{\pi\left(x^{(i)}\right)}{1 - \pi\left(x^{(i)}\right)} + \log\left(1 - \pi\left(x^{(i)}\right)\right)\right] \\ &= \sum_{i=1}^{N} \left[y^{(i)}\left(\boldsymbol{\omega} \cdot x^{(i)} + b\right) - \log\left(1 + \exp\left(\boldsymbol{\omega} \cdot x^{(i)} + b\right)\right)\right] \end{aligned} \tag{19}$$

We also introduce the concept of loss function in machine learning, just take a negative sign for (19), that is, $L_{loss}(\boldsymbol{\omega}) = \sum_{i=1}^{N} \left[\log\left(1 + \exp\left(\boldsymbol{\omega} \cdot x^{(i)} + b\right)\right) - y^{(i)}\left(\boldsymbol{\omega} \cdot x^{(i)} + b\right)\right]$ is used to represent the loss in the training process, then maximizing $L(\boldsymbol{\omega})$ is equivalent to minimizing $L_{loss}(\boldsymbol{\omega})$, i.e.

$$\min_{\boldsymbol{\omega}, b} \sum_{i=1}^{N} \left[\log\left(1 + \exp\left(\boldsymbol{\omega} \cdot x^{(i)} + b\right)\right) - y^{(i)}\left(\boldsymbol{\omega} \cdot x^{(i)} + b\right)\right] \tag{20}$$

which is a convex optimization problem so that we can update the weight coefficient by gradient descent. The process of parameters updated can be described as

$$\boldsymbol{\omega}_i = \boldsymbol{\omega}_i - \eta \frac{\partial L_{loss}(\boldsymbol{\omega}_i)}{\partial \boldsymbol{\omega}_i} \tag{21}$$

where η represents the step size of each iteration.

Specially, $\frac{\partial L_{loss}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}_i}$ can be calculated as follows.

$$\begin{aligned} &\frac{\partial L_{loss}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}_i} \\ &= -\sum_{i=1}^{N} \left[\left(y^{(i)} \frac{1}{\pi\left(x^{(i)}\right)} - \left(1 - y^{(i)}\right) \frac{1}{1 - \pi\left(x^{(i)}\right)}\right) \times \frac{\partial \pi\left(x^{(i)}\right)}{\partial \boldsymbol{\omega}_j}\right] \\ &= -\sum_{i=1}^{N} \left[\left(y^{(i)} \frac{1}{\pi\left(x^{(i)}\right)} - \left(1 - y^{(i)}\right) \frac{1}{1 - \pi\left(x^{(i)}\right)}\right) \times \pi\left(x^{(i)}\right)\left(1 - \pi\left(x^{(i)}\right)\right) \frac{\partial\left(b + \boldsymbol{\omega} \cdot x^{(i)}\right)}{\partial \boldsymbol{\omega}_j}\right] \end{aligned}$$

$$= -\sum_{i=1}^{N}\left(y^{(i)}\left(1-\pi\left(x^{(i)}\right)\right)-\left(1-y^{(i)}\right)\pi\left(x^{(i)}\right)\right)x_j^{(i)}$$
$$= -\sum_{i=1}^{N}\left(y^{(i)}-\pi\left(x^{(i)}\right)\right)x_j^{(i)}$$

Therefore, (21) can be rewritten as

$$\boldsymbol{\omega}_i \leftarrow \boldsymbol{\omega}_i + \eta\sum_j x_j^{(i)}\left(y^{(j)} - P\left(y^{(j)}=1|x_j\right)\right) \tag{22}$$

In our algorithm, we first construct a training set $\mathbb{T}$ based on the system model in Section II, which can be described as

$$\mathbb{T} = \left\{\left(\varphi^{(1)}, q^{(1)}\right), \left(\varphi^{(2)}, q^{(2)}\right), \ldots, \left(\varphi^{(i)}, q^{(i)}\right)\right\} \tag{23}$$

where $i = 1, \ldots, 2N$, $\varphi^{(i)}$ denotes the i-th sample from the i-th row vector in $\boldsymbol{\Phi}$, $q^{(i)} \in \{1, 0\}$ denotes the classification label of the i-th sample. It is worth noting that each row vector of $\boldsymbol{\Phi}$ is regarded as a sample, and the corresponding $\mathbf{q}$ is regarded as the label. Thus, the size of matrix $\boldsymbol{\Phi}$ and $\mathbf{q}$ determines the size of the training set. Since the received complex-valued data of N virtual antennas are divided into real and imaginary parts, the size of $\boldsymbol{\Phi}$ and $\mathbf{q}$ are $2N \times 2N$ and $2N \times 1$ respectively, and the total number of training samples is $2N$.

Just input the training set $\mathbb{T}$ into a basic LR model, and then we can get the following optimization problem to train the parameter $\mathbf{t}$

$$\min_{\mathbf{t}, t_0} \sum_{i=1}^{2N}\left[\log\left(1+\exp\left(t_0+\mathbf{t}\cdot\varphi^{(i)}\right)\right) - q^{(i)}\left(t_0+\mathbf{t}\cdot\varphi^{(i)}\right)\right] \tag{24}$$

where $\mathbf{t}_0$ is the initial classification coefficient vector, $\varphi^{(i)}$ is the input of the i-th sample and $\mathbf{t}$ is the classification coefficient to be optimized. The process of the coefficient updated by gradient descent method can be described as

$$t_i \leftarrow t_i + \eta\sum_j \Phi_{ji}\left(q^{(j)} - P\left(q^{(j)}=1|\Phi_j\right)\right) \tag{25}$$

where t_i denotes the i-th component of the coefficient vector $\mathbf{t}$, η denotes the step size of each optimization update, Φ_{ji} denotes the element in the j-th row and i-th column of $\boldsymbol{\Phi}$, $q^{(j)}$ denotes the j-th measurement and Φ_j denotes the j-th row of $\boldsymbol{\Phi}$ (also the j-th training sample).

In logistic regression, training is completed when the loss function is less than a certain threshold. Then we can get an optimal classification coefficient vector $\mathbf{t}$ based on the training set $\mathbb{T}$ and can easily index the DOAs on grid by the position of K non-zero elements of the sparse vector $\mathbf{t}$. The power spectrum of the recovered signal can be described as

$$\|P\|_i = \sqrt{t_i^2 + t_{i+N}^2}, \quad i = 1, 2, \ldots, N \tag{26}$$

The largest K elements in $\|\mathbf{P}\|$ correspond to the K DOAs on grid, which can be described as

$$\hat{\theta}_i = arcsin(2n_i/N), \quad i = 1, \ldots, K \tag{27}$$

where n_i denotes the subscript value corresponding to the i-th largest component in $\|\mathbf{P}\|$.

It is worth pointing out that in this paper, the machine learning algorithm we use does not act like traditional supervised learning algorithms including training stage and testing stage. As we have explained in the introduction, the one-bit quantized DOA estimation problem based on CS framework can be regarded as the coefficient solving process of a binary classification problem. Once the training process is over, a set of classification coefficients can be obtained, which is a sparse vector, and we can index the coarse position of DOAs through its sparse position. On the other hand, the scenario we consider in is wireless communication whose DOA range is between 0 degrees and 90 degrees. Equation (27) implies that the grid allocation is limited by the number of virtual antenna elements essentially. For an array in which the element spacing is half wavelength and the number of elements is N, the number of meshing angle is $\frac{N}{2}$ in the range between 0 degrees and 90 degrees, thus, the number of virtual antennas determines the resolution for the dataset.

B. MULTI-RESOLUTION GRID REFINEMENT

It is worth noting that (27) implements a rough estimate of DOAs, in other words, the estimated DOAs lie on the predefined N grid. Actually, the application of LR algorithm in this paper plays a pivotal role in dimensionality reduction, which helps obtain the main component of the incident signals and determine which column vectors in $\mathbf{F}$ should be retained. Instead of having a universally sophisticated grid, we make the grid fine only around the regions where the sources exist, that is, we perform a more complete division around the column vector of $\mathbf{F}$ retained previously. Assume that we have only one source signal in the re-divided area, then the more accurate spatial angles can be obtained through the following optimization problem, which can be constructed as

$$\begin{aligned} &\min \|\mathbf{P}\|_1 \\ &\text{s.t. } \|\mathbf{X}_{\mathbb{U}} - \boldsymbol{\Psi}\mathbf{P}\|^2 \leqslant \beta^2 \end{aligned} \tag{28}$$

where $\mathbf{X}_{\mathbb{U}} \in \mathbb{C}^N$ here is unquantized measurements of the virtual array, $\boldsymbol{\Psi}$ is partial steering matrix around the coarse position. We introduce a penalty factor λ to eliminate the inequality constraint in (28), which can be described as

$$\min \|\mathbf{X}_{\mathbb{U}} - \boldsymbol{\Psi}\mathbf{P}\|^2 + \lambda\|\mathbf{P}\|_1 \tag{29}$$

where $\|\mathbf{P}\|_1 = \sum_{i=1}^{k}\sqrt{\Re(\mathbf{P})^2 + \Im(\mathbf{P})^2}$, and the optimization result has a more ideal sparsity with the increase of λ, noise effect occupies a very small percentage with low λ. The former in (29) reflects the degree of mismatch while the latter restricts the requirement for sparsity. We adopt second order core programming to convert (29) as

$$\begin{aligned} &\min\ p + \lambda q \\ &\text{s.t. } \|\mathbf{X}_{\mathbb{U}} - \boldsymbol{\Psi}\boldsymbol{P}\|^2 \leqslant p \\ &\quad \|\mathbf{P}\|_1 \leqslant q \end{aligned} \tag{30}$$

Algorithm 1 Logistic Regression Algorithm for DOA Estimation

Input: input training set $\mathbb{T}$

Output: Power spectrum $\widehat{\|\mathbf{P}\|}$ and DOAs $\widehat{\boldsymbol{\theta}}$.

1: Generate a training set $\mathbb{T}$ by (8), (10), (11) and (12);
2: Initialize the classification coefficient vector $\mathbf{t} = \mathbf{0}^{2N}$ and the optimize step size η;Set the number of iterations T_{outer};
3: **for** $i = 1, \ldots, T_{outer}$ **do**
4: Update the coefficient vector $\mathbf{t}_i$ by (25);
5: **end for**
6: Compute the power spectrum $\|\mathbf{P}\|$ by (26), the coarse DOAs $\boldsymbol{\theta}^{(0)} = [\theta_1{}^{(0)}, \theta_2{}^{(0)}, \ldots, \theta_K{}^{(0)}]$ by (27) and set $r = 0$;
7: Form the refined grids $\boldsymbol{\psi}_{\theta}^{(r)}$ around the locations of the coarse DOAs. Use the optimization (30) to obtain more accurate source locations $\boldsymbol{\theta}^{(r)}$ and set $r = r + 1$;
8: Back to Step 7. until the grid is fine enough or iteration reaches a certain number of times;
9: **return** Accurate power spectrum $\widehat{\|\mathbf{P}\|}$ and DOAs $\widehat{\boldsymbol{\theta}}$.

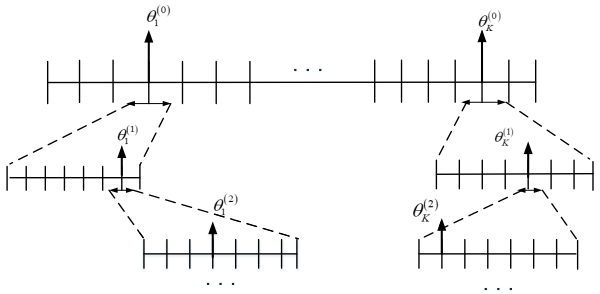


FIGURE 2. Process of grid refinement.

As depicted in Fig.2, we can get the spatial angles closer to its real location through continuous iteration and refinement. So far, the proposed algorithm is summarized as follows.

C. THE CRB EXPRESSION OF THE PROPOSED METHOD

In this subsection, we mainly discuss the Cramer-Rao bound (CRB) for the proposed method based on sparse arrays, which offers a lower bound on the variances of unbiased estimates of parameters. For a random vector $\mathbf{x} \in \mathbb{R}^n$ with a complex normal distribution with mean zero and covariance matrix $\mathbf{C} \in \mathbb{R}^{n\times n}$, the probability density function(pdf) can be described as

$$p(\mathbf{x};\boldsymbol{\alpha}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\mathbf{C}(\boldsymbol{\alpha}))}} \exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\mathbf{x}\right) \quad (31)$$

where $\boldsymbol{\alpha}$ is a real-valued parameter vector. For dimensional random variables $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]$, the optimal asymptotic unbiased estimator $\hat{\boldsymbol{\alpha}}(\mathbf{X})$ is obtained by maximizing the likelihood [19]

$$\begin{aligned}\hat{\boldsymbol{\alpha}}(\mathbf{X}) &= \arg\max \ln p(\mathbf{X};\theta) \\ &= \arg\max \sum_{i=1}^{m} \ln p(\mathbf{x}_i;\theta)\end{aligned} \quad (32)$$

Since the maximum likelihood estimator is consistent and effective, its asymptotic performance can be analytically characterized by the CRB, which is defined as the inverse matrix of Fisher information matrix (FIM) [20]

$$\mathrm{CRB}(\boldsymbol{\alpha}) = \mathrm{FIM}^{-1}(\boldsymbol{\alpha}) \quad (33)$$

where **FIM** is guaranteed to be positive semidefinite, and the (p, l) th entry of the **FIM** can be expressed as [20]

$$\mathbf{FIM}(\boldsymbol{\alpha})_{p,\ell} = \mathrm{tr}\left(\mathbf{C}^{-1}(\boldsymbol{\alpha})\frac{\partial\mathbf{C}(\boldsymbol{\alpha})}{\partial\alpha_p}\mathbf{C}^{-1}(\boldsymbol{\alpha})\frac{\partial\mathbf{C}(\boldsymbol{\alpha})}{\partial\alpha_\ell}\right) \quad (34)$$

For the system model described in section II, we define

$$\boldsymbol{\alpha} = \left[\bar{\theta}_1, \ldots, \bar{\theta}_K, p_1, \ldots, p_K, p_n\right]^T \quad (35)$$

Under the assumptions in Section II, each snapshot of the signal vector $\mathbf{x}_{\mathbb{S}}$ follows the Gaussian distribution with mean zero and covariance matrix $\mathbf{R}_{\mathbb{S}}$. For L snapshots of the observation, we have the following complex Gaussian distribution

$$\left[\mathbf{x}_{\mathbb{S}}(1)^T, \mathbf{x}_{\mathbb{S}}(2)^T, \ldots, \mathbf{x}_{\mathbb{S}}(L)^T\right]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_L \otimes \mathbf{R}_{\mathbb{S}}) \quad (36)$$

By substituting (36) into (34), the (p, l) th entry of **FIM** under L snapshots can be further expressed as

$$[\mathrm{FIM}(\boldsymbol{\alpha})]_{p,l} = L\,\mathrm{tr}\left(\mathbf{R}_{\mathbb{S}}^{-1}\frac{\partial\mathbf{R}_{\mathbb{S}}}{\partial\alpha_p}\mathbf{R}_{\mathbb{S}}^{-1}\frac{\partial\mathbf{R}_{\mathbb{S}}}{\partial\alpha_l}\right) \quad (37)$$

Since $\mathrm{tr}(\mathbf{XY}) = \left(\mathrm{vec}(\mathbf{X})^H\right)^H \mathrm{vec}(\mathbf{Y})$ and $\mathrm{vec}(\mathbf{XYZ}) = \left(\mathbf{Z}^T \otimes \mathbf{X}\right)\mathrm{vec}(\mathbf{Y})$, (37) can be further simplified as

$$\begin{aligned}&\mathbf{FIM}(\boldsymbol{\alpha})_{p,\ell} \\ &= L\mathrm{vec}^H\left(\frac{\partial\mathbf{R}_{\mathbb{S}}}{\partial\alpha_p}\right)\left(\mathbf{R}_{\mathbb{S}}^{-T}\otimes\mathbf{R}_{\mathbb{S}}^{-1}\right)\mathrm{vec}\left(\frac{\partial\mathbf{R}_{\mathbb{S}}}{\partial\alpha_\ell}\right) \\ &= L\left[\left(\mathbf{R}_{\mathbb{S}}^T\otimes\mathbf{R}_{\mathbb{S}}\right)^{-\frac{1}{2}}\frac{\partial\mathbf{r}_{\mathbb{S}}}{\partial\alpha_p}\right]^H\left[\left(\mathbf{R}_{\mathbb{S}}^T\otimes\mathbf{R}_{\mathbb{S}}\right)^{-\frac{1}{2}}\frac{\partial\mathbf{r}_{\mathbb{S}}}{\partial\alpha_l}\right]\end{aligned} \quad (38)$$

where $\mathbf{r}_{\mathbb{S}} = \mathrm{vec}(\mathbf{R}_{\mathbb{S}})$. We are interested in the information of DOAs in the parameter vector, then $\boldsymbol{\alpha}$ can be divided into $\boldsymbol{\alpha} = \left[\bar{\theta}_1, \ldots, \bar{\theta}_K\,|p_1, \ldots, p_K, p_n\right]^T$. Hence, (38) is finally expressed as

$$\mathbf{FIM}(\boldsymbol{\alpha}) = L\left[\mathbf{G}\ \boldsymbol{\Delta}\right]^H\left[\mathbf{G}\ \boldsymbol{\Delta}\right] \quad (39)$$

where $\mathbf{G}$ and $\boldsymbol{\Delta}$ are defined as

$$\begin{aligned}\mathbf{G} &= \left(\mathbf{R}_{\mathbb{S}}^T\otimes\mathbf{R}_{\mathbb{S}}\right)^{-\frac{1}{2}}\left[\frac{\partial\mathbf{r}_{\mathbb{S}}}{\partial\bar{\theta}_1}, \ldots, \frac{\partial\mathbf{r}_{\mathbb{S}}}{\partial\bar{\theta}_K}\right] \\ \boldsymbol{\Delta} &= \left(\mathbf{R}_{\mathbb{S}}^T\otimes\mathbf{R}_{\mathbb{S}}\right)^{-\frac{1}{2}}\left[\frac{\partial\mathbf{r}_{\mathbb{S}}}{\partial p_1}, \ldots, \frac{\partial\mathbf{r}_{\mathbb{S}}}{\partial p_K}, \frac{\partial\mathbf{r}_{\mathbb{S}}}{\partial p_n}\right]\end{aligned} \quad (40)$$

If $\mathbf{FIM}(\boldsymbol{\alpha})$ is invertible, the CRB for the normalized DOAs $\bar{\theta} = [\bar{\theta}_1, \ldots, \bar{\theta}_K]^T$ can be given as the inverse of the Schur complement of the block $\boldsymbol{\Delta}^H\boldsymbol{\Delta}$ of $\mathbf{FIM}(\boldsymbol{\alpha})$ [21]

$$\mathrm{CRB}(\bar{\theta}) = \frac{1}{L}\left(\mathbf{G}^H\boldsymbol{\Pi}_{\boldsymbol{\Delta}}^{\perp}\mathbf{G}\right)^{-1} \quad (41)$$

where $\boldsymbol{\Pi}_{\boldsymbol{\Delta}}^{\perp} = \mathbf{I} - \boldsymbol{\Delta}\left(\boldsymbol{\Delta}^H\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}^H$ denotes the orthogonal projection onto the null space of $\boldsymbol{\Delta}^H$.

Specifically, the proposed classification-based DOA estimation algorithm performs one-bit quantization on some items in $\mathbf{R}_{\mathbb{S}}$, i.e., the elements at the positive position in the co-array is quantized as $+1$ or 0. The effect of this operation is that the terms in the lower triangle of $\mathbf{R}_{\mathbb{S}}$ are quantized as $+1$ or 0 by (11), accordingly, $\mathbf{r}_{\mathbb{S}}$ also contains quantized elements. Since $\mathbf{R}_{\mathbb{S}}$ with no element quantized is an Hermite matrix with conjugate symmetry, the quantization of the lower triangle of $\mathbf{R}_{\mathbb{S}}$ will not affect its rank. From the previous discussion, the existence of the CRB is equivalent to the nonsigularity of $\mathbf{FIM}(\boldsymbol{\alpha})$. To further discuss the existence of CRB, we define a augmented coarray manifold(ACM) matrix $\mathbf{V}_{\mathbb{D}}$

Definition 3 (ACM matrix) [17]: The augmented co-array manifold (ACM) matrix is defined as

$$\mathbf{V}_{\mathbb{D}} = [\mathbf{J}diag\,(\mathbb{D})\,\mathbf{A}_{\mathbb{D}}\;\mathbf{W}_{\mathbb{D}}] \tag{42}$$

where $\mathbf{W}_{\mathbb{D}} = [\mathbf{J}\mathbf{A}_{\mathbb{D}}\;\mathrm{vec}(\mathbf{I})]$, $\mathbf{A}_{\mathbb{D}} = \left[\mathbf{a}_{\mathbb{D}}(\bar{\theta}_1), \ldots, \mathbf{a}_{\mathbb{D}}(\bar{\theta}_K)\right]^T$, $\mathbb{D}$ and $\mathbf{J}$ are defined in **Definition 1** and **Definition 2** respectively.

If and only if $\mathbf{V}_{\mathbb{D}}$ has full column rank, $\mathbf{FIM}(\boldsymbol{\alpha})$ is nonsingular [21], then the CRB for the normalized DOAs $\bar{\theta} = \left[\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_K\right]$ can be expressed as

$$\mathrm{CRB}(\bar{\theta}) = \frac{1}{4\pi^2 L}\left(\mathbf{G}_0^H \boldsymbol{\Pi}^{\perp}_{\mathbf{M}_0\mathbf{W}_{\mathrm{D}}}\mathbf{G}_0\right)^{-1} \tag{43}$$

where

$$\mathbf{G}_0 = \mathbf{M}_0\mathbf{J}\,\mathrm{diag}(\mathbb{D})\mathbf{A}_{\mathbf{D}}\,\mathrm{diag}\,([p_1, \ldots, p_K]) \tag{44}$$

$$\mathbf{M}_0 = \left(\tilde{\boldsymbol{R}}_{\mathbb{S}}^T \otimes \tilde{\boldsymbol{R}}_{\mathbb{S}}\right)^{-\frac{1}{2}} \tag{45}$$

$$\mathbf{W}_{\mathbb{D}} = [\mathbf{J}\mathbf{A}_{\mathbb{D}}\;\mathrm{vec}(\mathbf{I})] \tag{46}$$

$$\mathbf{R}_{\mathbb{S}} = \mathbf{A}_{\mathbb{S}}\mathbf{R}_{\mathbf{ss}}\mathbf{A}_{\mathbb{S}}^H + p_n\mathbf{I} \tag{47}$$

$$\mathbf{A}_{\mathbb{D}} = \left[\mathbf{a}_{\mathbb{D}}(\bar{\theta}_1), \mathbf{a}_{\mathbb{D}}(\bar{\theta}_2), \ldots, \mathbf{a}_{\mathbb{D}}(\bar{\theta}_K)\right]^T \tag{48}$$

Here, p_k is the power of the k-th source, $k = 1, 2, \ldots, K$, $p_n = {\sigma_n}^2$ is the power of noise, $\tilde{\boldsymbol{R}}_{\mathbb{S}}$ is the covariance matrix whose elements in the lower triangle are quantized as $+1$ or 0 by(11).

IV. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed algorithm from various perspectives including the mean absolute error (MAE) and measurement success rate (MSR). SNR is defined in terms of the signal power p_s and noise power p_n as

$$\mathrm{SNR} = 10\log\left(\frac{p_s}{p_n}\right) \tag{49}$$

All sources are assumed to have the same power. In addition, the MAE and MSR of the estimated DOAs are defined as

$$\mathrm{MAE}_{\theta} = \frac{1}{RK}\sum_{r=1}^{R}\sum_{k=1}^{K}\left|\theta_k - \tilde{\theta}_k^{(r)}\right| \tag{50}$$

$$\mathrm{MSR}_{\theta_i} = \frac{m_k}{R} \tag{51}$$

respectively, where R is the number of Monte Carlo trials and $\tilde{\theta}_k^{(i)}$ is the estimated angle of the k-th source in the r-th trial. m_k denotes the number of successful trails of the k-th source in R Monte Carlo trials, where m_k is effective if and only if $\left|\theta_k - \tilde{\theta}_k\right| \leqslant 0.2°$.

It is worth noting that the proposed algorithm based on CS adopts classification algorithm in machine learning to solve the sparse recovery problem of DOA. Hence, when comparing the performance of the proposed algorithm and other algorithms, we choose the classic CS-based DOA estimation algorithms such as OMP [5], SBL [6]. On the other hand, an over-complete dictionary is constructed in our algorithm, which implies the angle division. Thus, L1-SVD [3] is also taken into account due to the similar operation.

A. NUMBER OF DOFs

In this subsection, we mainly discuss two types of sparse arrays, coprime array [16] and nested array [15]. A coprime array is comprised of two ULAs with M and N antenna elements, while M and N ($N > M$ without loss of generality) are coprime, as shown in Fig. 3(b). The total number of a coprime array is $(N + 2M - 1)$ whose locations is given by $\mathbb{S}_{\mathrm{coprime}} = \{0, M, \ldots., (N-1)M, N, \ldots, (2M-1)N\}$. Similarly, a two-level nested array is also a series of two ULAs with locations $\mathbb{S}_{\mathrm{nested}} = \{1, \ldots, N_1, (N_1+1), \ldots, N_2(N_1+1)\}$, as depicted in Fig. 3(c), the total number of a nested array is $(N_1 + N_2)$. Accordingly, the locations of the virtual array of the coprime array and nested array can be excessed as $\mathbb{D}_{\mathrm{coprime}} = \{0, \pm1, \ldots., \pm(MN + M - 1)\}$ and $\mathbb{D}_{\mathrm{nested}} = \{0, \pm d, \ldots, \pm(N_2(N_1+1)-1)\}$ respectively, where $\mathbb{D}_{\mathrm{coprime}}$ and $\mathbb{D}_{\mathrm{nested}}$ are given by **Definition 1** respectively. $\mathbb{U}_{\mathrm{coprime}} = \{0, 1, \ldots., (MN + M - 1)\}$ and $\mathbb{U}_{\mathrm{nested}} = \{0, 1, \ldots, (N_2(N_1+1)-1)\}$ are the non-negative part of $\mathbb{D}_{\mathrm{coprime}}$ and $\mathbb{D}_{\mathrm{nested}}$ respectively. We select a certain number of antennas to observe the expansion of sparse arrays which is actually caused by the second-order statistics of the measurements, as shown in Fig.4. It can be seen that a nested array will expand more DOFs under the same number of physical antennas, also, the co-array is a continuous ULA. However, there are some holes of the co-array of a coprime array. Just take nested arrays as an example in subsequent simulations.

Next, we consider the effect of the number of antennas on the MSR. We select θ from the range $10° \sim 60°$ randomly with 500 independent Monte Carlo trails under each number of antennas. As shown in Fig.5, the process of grid refinement can highly increase the MSR of the estimated DOAs and further reduce the dependence on the array size. Especially, the MSR with grid refinement is up to 100% when $N \geqslant 16$, while the rough estimation only comes to 50%.

B. ESTIMATION ACCURACY

In this subsection, we first examine the DOA estimation accuracy between the proposed algorithm and other methods,

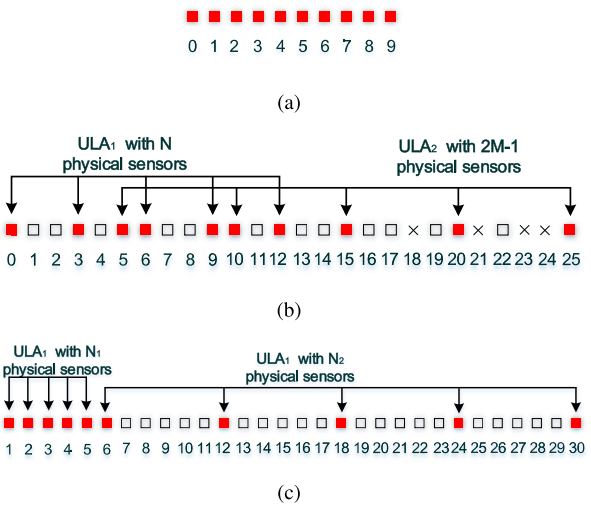


FIGURE 3. The array configurations for (a) ULA with 10 sensors, (b) a coprime array with $M = 3$ and $N = 5$, and (c) a nested array with $N1 = N2 = 5$. Here red squares denote physical antennas, white squares represent extended virtual antennas, × denotes hole of the co-array with no physical or virtual antennas.

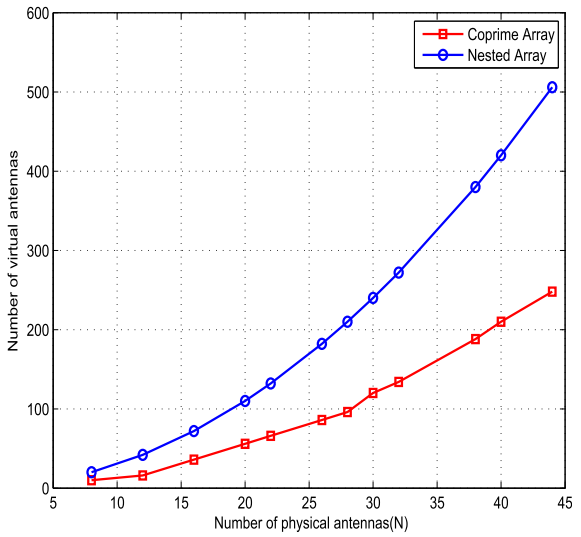


FIGURE 4. The number of virtual antennas versus the number of physical antennas.The array configurations for a coprime array with $M = [3\ 5\ 5\ 5\ 5\ 5\ 7\ 7\ 11\ 11\ 13]$ and $N = [3\ 3\ 7\ 11\ 13\ 17\ 19\ 17\ 19\ 17\ 19\ 19]$, and a nested array with $N1 = N2 = [4\ 6\ 8\ 10\ 11\ 13\ 14\ 15\ 16\ 19\ 20\ 22]$.

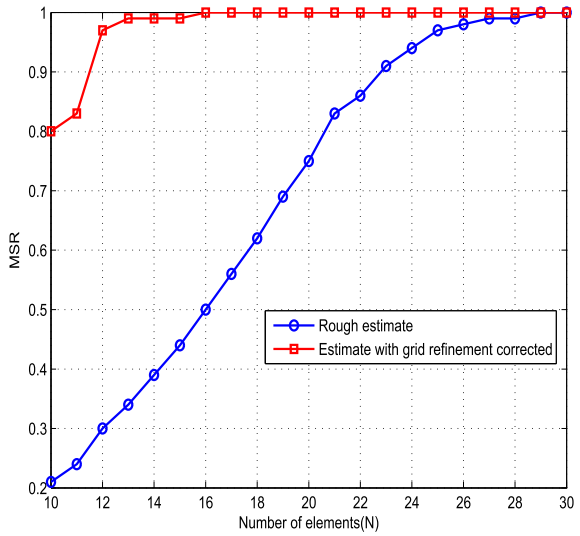


FIGURE 5. The MSR versus the number of attennas. SNR is 0dB and the number of snapshots is 128.

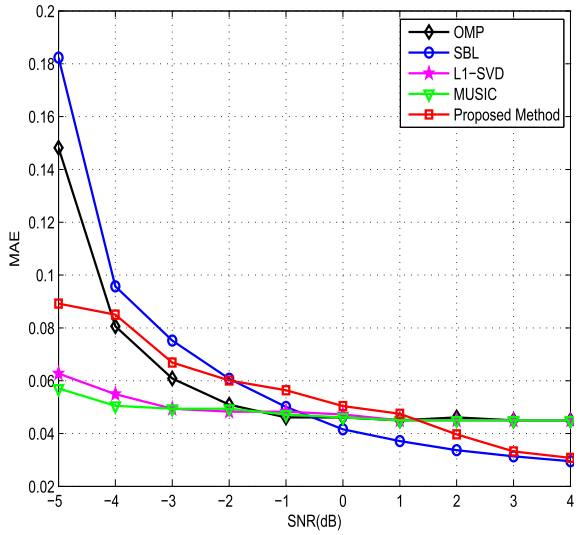


FIGURE 6. The MAE versus SNR. The proposed algorithm is equipped with 19 antennas, other methods are equipped with 100 antennas and the number of snapshots is 10.

including MUSIC [1], OMP [5], SBL [6] and L1-SVD [3]. Two configurations are taken into account, namely, the proposed algorithm with 19 antennas, other methods with 100 antennas and both the proposed algorithm and other methods with 19 antennas. For the proposed method, 19 physical antennas can expend 100 virtual antennas. Only one source is considered with the direction 47.555° and 10 snapshots, SNR is set from -5 to 4 dB and 500 independent Monte Carlo trails are conducted to calculate the MAE. Then the MAE of the two configurations versus SNR are plotted in Fig. 6 and Fig. 7. It can be observed from Fig.6 that under the first configuration, the proposed method can achieve lower MAE than SBL [6] when SNR $\leqslant -2$dB and lower than MUSIC [1], OMP [5] and L1-SVD [3] when SNR $\geqslant$ 2dB. Interestingly, the proposed method outperforms other methods mentioned in this paper on the estimate accuracy when the numbers of physical antennas are the same.

Then we consider the MSR and MAE versus different snapshots. First, three configurations are taken into account, one source with the direction of 47.5°, two sources with the direction of 47.5°, 32.4° and three sources with the direction of 47.5°, 32.4°, 10.2°. The number of snapshots is set from

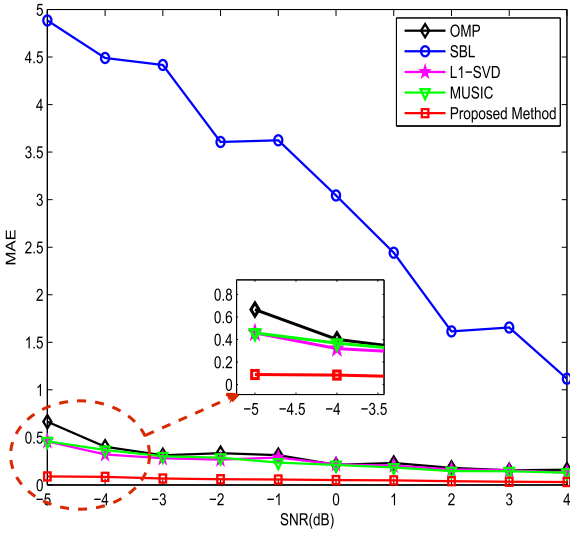


FIGURE 7. The MAE versus SNR. Both the proposed algorithm and other methods are equipped with 19 antennas, and the number of snapshots is 10.

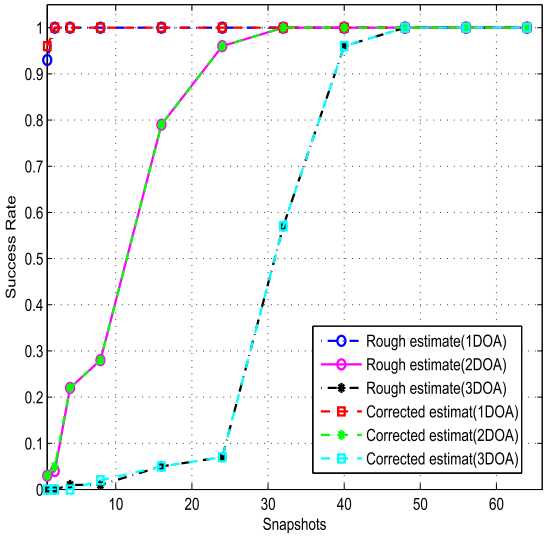


FIGURE 8. MSR versus different snapshots. SNR is set 0dB and the number of antennas is 30.

1 to 64 and 500 independent Monte Carlo trails are conducted to calculate the MSR. The result of MSR versus different snapshots is plotted in Fig.8. It is clear that the number of snapshots is needed to be guaranteed for extending effective measurements of virtual antennas. Next, one source is considered with the direction 32.485°, 19-antenna for the proposed method and 100-antenna for other methods are given to evaluate the estimation accuracy. The number of snapshots is set from 1 to 11 and 500 independent Monte Carlo trails are conducted to calculate the MAE. As depicted in Fig.9, the estimate accuracy of the proposed method is second only to L1-SVD under one snapshot, meanwhile, the proposed method performs better with the increasement of snapshots.

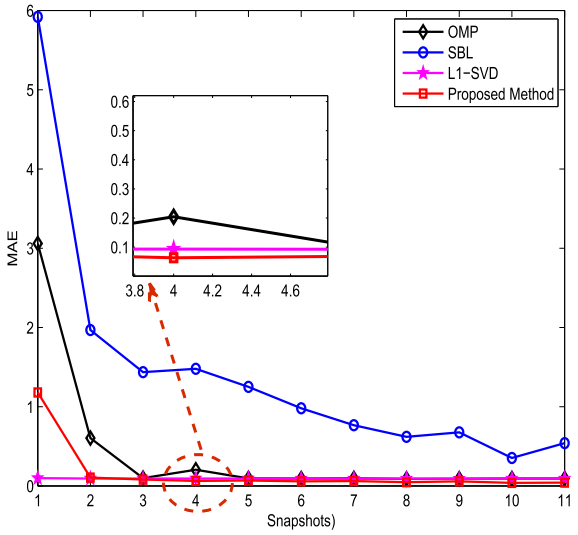


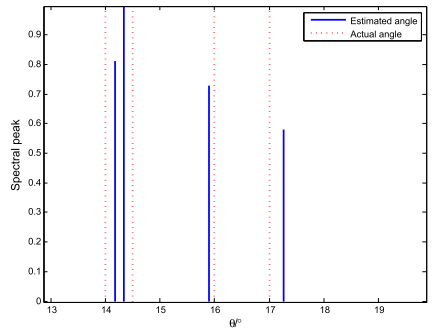
FIGURE 9. MAE versus different snapshots. The proposed algorithm is equipped with 19 antennas, other methods are equipped with 100 antennas and SNR is 0dB.

C. PREDICTION OF RESOLVABILITY

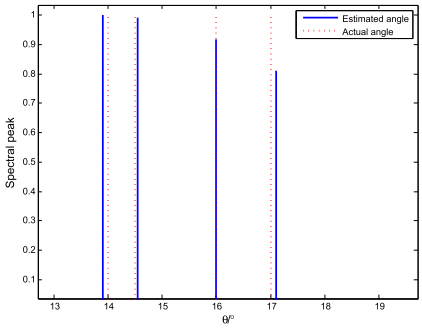
In this section, we mainly explore the estimated performance between the proposed algorithm and L1-SVD algorithm when the incident signals are quite close to each other. The simulation configurations are set as follows. Suppose that the direction of arrival of the incident signals are from 14°, 14.5°, 16° and 17°. SNR is set to 0dB and 256 snapshots are taken into account. For the proposed method, 20 antennas-nested array and 30 antennas-nested array are considered separately, whose simulation results are shown in Fig. 10(a) and Fig. 10(b) respectively. For L1-SVD algorithm, we adopt 30 antennas-ULA array and 100 antennas-ULA array separately, whose simulation results are shown in Fig. 10(c) and Fig. 10(d) respectively. As depicted in Fig.10, when the incident signals are close to each other (i.e, the angle interval is 0.5° or 1°), the proposed method can distinguish them fairly accurately with fewer physical antennas comparing with L1-SVD algorithm (L1-SVD algorithm even failed when the number of antennas is 30). It is noting that the initial angle grid allocation is essentially limited by the number of virtual array. For an extend virtual array which has M virtual antennas, the number of meshing angle is $\frac{M}{2}$, where M is increasing fast with the number of physical antennas. Thus, more physical antennas means more meshing grids, and it can be found that the proposed method has better resolution performance for extremely close two signals with the increasing number of physical antennas.

D. ANALYSIS ON THE COMPLEXITY

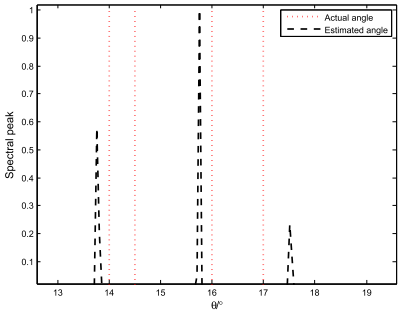
In some applications such as massive MIMO communications, a large array is usually equipped and complexity reduction is always of great significance. To deal with this challenge, we take two effective measures in the proposed algorithm, the sparse arrays and one-bit quantization. The



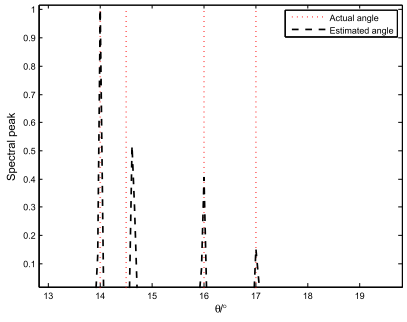
(a) The proposed method, 20 antennas.



(b) The proposed method, 30 antennas.



(c) L1-SVD algorithm, 30 antennas.



(d) L1-SVD algorithm, 100 antennas

FIGURE 10. The prediction of resolvability between the proposed method and L1-SVD algorithm.The direction of the incident signals are 14°, 14.5°, 16°, 17°.

TABLE 1. Computation time.

Different methods	1DOA	2DOAs	3DOAs
SS-MUSIC[15]	0.20021s	0.20555s	0.21178s
L1-SVD[3]	3.3606s	4.4822s	5.6529s
SBL[6]	1.3222s	1.476725s	1.3894s
OMP[5]	0.0607s	0.0836s	0.0875s
the proposed method	0.2628s	0.2657s	0.2755s

former has the advantage that it greatly reduces the physical antennas of the system and meanwhile expand the measurements with virtual array. Since the subsequent operations of one-bit quantization only retains the symbol information of the extended measurements, it will be more efficient in digital signal processing. In addition, the logistic regression algorithm applied in our algorithm actually played a role of finding the main components of the signals, so the subsequent grid refinement is working with a convex optimization problem of low-dimensional data.

To further compare the computational complexity of the proposed algorithm with other methods including SS-MUSIC [15], OMP [5], SBL [6] and L1-SVD [3], the computation time calculated by the MATLAB profiler under the environment of Intel CPU I5-6300HQ with the processor frequency 2.30 GHz and 8 GB RAM is listed in Table 1. In the simulation, SS-MUSIC and the proposed method is equipped with 19 physical antennas, i.e., 100 virtual antennas, other methods are equipped with 100 physical antennas. SNR is set to 0dB and three configurations are taken into account, one source with direction of 10°, two sources with direction of 10°, 30° and three sources with direction of 10°, 30°, 50°. It is clear that the running time of the proposed algorithm is roughly the same as that of SS-MUSIC [15] under the same conditions, whose running time are significantly less than that of SBL [6] and L1-SVD [3]. Although the running time of the proposed algorithm is longer than that of OMP, it can be seen from Fig.7 that the estimation accuracy of the proposed algorithm is higher than that of OMP, and it can be seen from Fig.8 that the performance of the proposed algorithm is also better at low snapshots. Thus, the increased running time compared with OMP is acceptable.

V. CONCLUSION

In this paper, we developed a generalized framework of sparse array for classification-based one-bit DOA estimation. DOA estimation is divided into two steps including finding the main components of the signals with classification algorithms, and then a valid grid refinement process is adopted to alleviate the grid effect. We also derive the Cramer-Rao bound for the algorithm under the proposed structure. By analysing the number of DOFs and estimation accuracy, it is revealed that given the same number of physical antennas, the proposed method outperforms the traditional CS-based DOA estimation algorithm. The superiority of the

proposed algorithm is that both the hardware complexity and computational complexity are lower with the application of sparse arrays and one-bit quantization. It is worth noting that the classification algorithm in the proposed method plays a role in dimensionality reduction and achieves rough estimation of DOAs.

ACKNOWLEDGMENT

The authors would like to thank Linxiao Su for sharing his matlab code about SBL.

REFERENCES

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
[2] C. Zhou, F. Haber, and D. L. Jaggard, "A resolution measure for the MUSIC algorithm and its application to plane wave arrivals contaminated by coherent interference," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 454–463, Feb. 1991.
[3] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
[4] G. Z. Karabulut, T. Kurt, and A. Yongacoglu, "Angle of arrival detection by matching pursuit algorithm," in *Proc. IEEE 60th Veh. Technol. Conf. VTC-Fall.*, vol. 4, Sep. 2004, pp. 154–156.
[5] S. Ganguly, I. Ghosh, R. Ranjan, J. Ghosh, P. K. Kumar, and M. Mukhopadhyay, "Compressive sensing based off-grid DOA estimation using OMP algorithm," in *Proc. 6th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Noida, India, Mar. 2019, pp. 772–775.
[6] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.
[7] X. Wu, W.-P. Zhu, and J. Yan, "Direction of arrival estimation for off-grid signals based on sparse Bayesian learning," *IEEE Sensors J.*, vol. 16, no. 7, pp. 2004–2016, Apr. 2016.
[8] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 499–512, Jan. 2015.
[9] M. Stein, K. Barbe, and J. A. Nossek, "DOA parameter estimation with 1-bit quantization bounds, methods and the exponential replacement," in *Proc. 20th Int. ITG Workshop Smart Antennas*, Mar. 2016, pp. 1–6.
[10] O. Bar-Shalom and A. J. Weiss, "DOA estimation using one-bit quantized measurements," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 868–884, Jul. 2002.
[11] X. Meng and J. Zhu, "A generalized sparse Bayesian learning algorithm for 1-bit DOA estimation," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1414–1417, Jul. 2018.
[12] K. Yu, Y. D. Zhang, M. Bao, Y.-H. Hu, and Z. Wang, "DOA estimation from one-bit compressed array data via joint sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1279–1283, Sep. 2016.
[13] C. Gianelli, L. Xu, J. Li, and P. Stoica, "One-bit compressive sampling with time-varying thresholds: Maximum likelihood and the Cramér-rao bound," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 399–403.
[14] C.-L. Liu and P. P. Vaidyanathan, "One-bit sparse array DOA estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 3126–3130.
[15] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.
[16] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the music algorithm," in *Proc. Digit. Signal Process. Signal Process. Edu. Meeting (DSP/SPE)*, Sedona, AZ, USA, Jan. 2011, pp. 289–294.
[17] T. Chen, M. Guo, and X. Huang, "Direction finding using compressive one-bit measurements," *IEEE Access*, vol. 6, pp. 41201–41211, 2018.
[18] Y. Gao, D. Hu, Y. Chen, and Y. Ma, "Gridless 1-b DOA estimation exploiting SVM approach," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2210–2213, Oct. 2017.
[19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddler River, NJ, USA: Prentice-Hall, 1993.
[20] P. Stoica, E. G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: A textbook derivation," *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 148–150, May 2001.
[21] C.-L. Liu and P. P. Vaidyanathan, "New Cramér-Rao bound expressions for coprime and other sparse arrays," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Rio de Janerio, Brazil, Jul. 2016, pp. 1–5.

YANPING CHEN received the B.S. degree from the Harbin University of Commerce, China, in 2005, the M.S. degree from the Harbin Institute of Technology, China, in 2007, and the Ph.D. degree from Harbin Engineering University, China, in 2012. She is currently a Lecturer with the School of Computer and Information Engineering, Harbin University of Commerce. Her current research interests include statistical signal processing and performance analysis of networks.

CHEN WANG received the B.S. degree in electronic information engineering from Harbin Engineering University, Harbin, China, in 2019. He is currently pursuing the M.S. degree with the Communication Engineering Center, Harbin Institute of Technology. His research interests include direction-of-arrival estimation, sparse arrays, compressive sensing, matrix completion, and Cramer-Rao bound analysis.

YULONG GAO (Member, IEEE) received the B.S. degree from the Heilongjiang Institute of Science and Technology, China, in 2001, the M.S. degree from Harbin Engineering University, China, in 2004, and the Ph.D. degree from the Harbin Institute of Technology, China, in 2007. From May 2012 to May 2013, he was a Visitor with the University of Toronto, Canada. He is currently an Associate Professor with the School of Electronics and Information Engineering, Harbin Institute of Technology. His current research interests include cognitive radio and intelligent signal processing. He is also a member of the IEEE Communications Society.