Compressed Sensing and Tensors for Communication and Radar Systems

Lead Guest Editor: Fangqing Wen
Guest Editors: Wei Liu, Lei Zhang, Liangtian Wan, and Xianpeng Wang

Compressed Sensing and Tensors for Communication and Radar Systems

Compressed Sensing and Tensors for Communication and Radar Systems

Lead Guest Editor: Fangqing Wen
Guest Editors: Wei Liu, Lei Zhang, Liangtian Wan, and Xianpeng Wang

Jose M. Lanza-Gutierrez, Spain
Pavlos I. Lazaridis, United Kingdom
Kim-Hung Le, Vietnam
Tuan Anh Le, United Kingdom
Xianfu Lei, China
Jianfeng Li, China
Xiangxue Li, China
Yaguang Lin, China
Zhi Lin, China
Liu Liu, China
Mingqian Liu, China
Zhi Liu, Japan
Miguel López-Benítez, United Kingdom
Chuanwen Luo, China
Lu Lv, China
Basem M. ElHalawany, Egypt
Imadeldin Mahgoub, USA
Rajesh Manoharan, India
Davide Mattera, Italy
Michael McGuire, Canada
Weizhi Meng, Denmark
Klaus Moessner, United Kingdom
Simone Morosi, Italy
Amrit Mukherjee, Czech Republic
Shahid Mumtaz, Portugal
Giovanni Nardini, Italy
Tuan M. Nguyen, Vietnam
Petros Nicopolitidis, Greece
Rajendran Parthiban, Malaysia
Giovanni Pau, Italy
Matteo Petracca, Italy
Marco Picone, Italy
Daniele Pinchera, Italy
Giuseppe Piro, Italy
Javier Prieto, Spain
Umair Rafique, Finland
Maheswar Rajagopal, India
Sujan Rajbhandari, United Kingdom
Rajib Rana, Australia
Luca Reggiani, Italy
Daniel G. Reina, Spain
Bo Rong, Canada
Mangal Sain, Republic of Korea
Praneet Saurabh, India

Hans Schotten, Germany
Patrick Seeling, USA
Muhammad Shafiq, China
Zaffar Ahmed Shaikh, Pakistan
Vishal Sharma, United Kingdom
Kaize Shi, Australia
Chakchai So-In, Thailand
Enrique Stevens-Navarro, Mexico
Sangeetha Subbaraj, India
Tien-Wen Sung, Taiwan
Suhua Tang, Japan
Pan Tang, China
Pierre-Martin Tardif, Canada
Sreenath Reddy Thummaluru, India
Tran Trung Duy, Vietnam
Fan-Hsun Tseng, Taiwan
S Velliangiri, India
Quoc-Tuan Vien, United Kingdom
Enrico M. Vitucci, Italy
Shaohua Wan, China
Dawei Wang, China
Huaqun Wang, China
Pengfei Wang, China
Dapeng Wu, China
Huaming Wu, China
Ding Xu, China
YAN YAO, China
Jie Yang, USA
Long Yang, China
Qiang Ye, Canada
Changyan Yi, China
Ya-Ju Yu, Taiwan
Marat V. Yuldashev, Finland
Sherali Zeadally, USA
Hong-Hai Zhang, USA
Jiliang Zhang, China
Lei Zhang, Spain
Wence Zhang, China
Yushu Zhang, China
Kechen Zheng, China
Fuhui Zhou, USA
Meiling Zhu, United Kingdom
Zhengyu Zhu, China

Contents

WILEY | Hindawi

Research Article
Block Sparse Bayesian Recovery with Correlated LSM Prior

Juan Zhao [ID], Xia Bai, Tao Shan, and Ran Tao

School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Juan Zhao; juanzhao@bit.edu.cn

Compressed sensing can recover sparse signals using a much smaller number of samples than the traditional Nyquist sampling theorem. Block sparse signals (BSS) with nonzero coefficients occurring in clusters arise naturally in many practical scenarios. Utilizing the sparse structure can improve the recovery performance. In this paper, we consider recovering arbitrary BSS with a sparse Bayesian learning framework by inducing correlated Laplacian scale mixture (LSM) prior, which can model the dependence of adjacent elements of the block sparse signal, and then a block sparse Bayesian learning algorithm is proposed via variational Bayesian inference. Moreover, we present a fast version of the proposed recovery algorithm, which does not involve the computation of matrix inversion and has robust recovery performance in the low SNR case. The experimental results with simulated data and ISAR imaging show that the proposed algorithms can efficiently reconstruct BSS and have good antinoise ability in noisy environments.

1. Introduction

Compressed sensing (CS) [1] provides a new sampling and reconstruction paradigm, which can recover sparse signals from linear measurements:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{\Phi} \in R^{M \times N} (M < N)$ is the measurement matrix, $\mathbf{y} \in R^M$ is the measurement vector, $\mathbf{x} \in R^N$ is the sparse signal, and $\mathbf{n} \in R^M$ is the additive noise. Many recovery algorithms have been presented to reconstruct sparse signals, including orthogonal matching pursuit (OMP) [2] and sparse Bayesian learning (SBL) [3].

In some signal processing applications such as ISAR imaging [4] and gene expression levels [5], there are many sparse signals with block structural features; i.e., nonzero elements are often clustered. Inducing structural a priori can largely improve the reconstruction performance. Therefore, to improve the reconstruction effect of block sparse signals (BSS), many algorithms are proposed. For instance, Block-OMP [6] and Block-StOMP [7] are OMP-based approaches. Meanwhile, some block recovery algorithms

based on the Bayesian compressed sensing framework are presented, including block sparse Bayesian learning (BSBL) [8], Cluss-MCMC [9], model-based Bayesian CS via local beta process (MBCS-LBP) [10], and pattern-coupled sparse Bayesian learning (PC-SBL) [11]. Among these recovery algorithms, Bayesian algorithms have parameter learning ability and can be applied to recovery arbitrary signals with unknown sparse structures by flexibly imposing different sparse prior models.

In [12], Zhang et al. have proposed an expectation-maximization-based variational Bayesian (EM-VB) inference method, which utilizes the Laplacian scale mixture (LSM) model as a sparse prior; i.e., it is assumed that the sparse signal obeys the Laplacian prior because the Laplacian distribution can represent sparseness well. Based on this model, for the BSS with unknown block information, this paper proposes a block Bayesian recovery algorithm by inducing a correlated LSM prior model, which uses the dependence between neighboring elements of the BSS. Furthermore, to improve the computational efficiency of the proposed recovery algorithm, a fast version without matrix inversion is presented, which is suitable for noisy environments, especially in the low SNR case. The experimental

results on simulated data and ISAR imaging show that the proposed algorithms have a good reconstruction effect on BSS and can resist noise in noisy environments.

The remainder of this paper is organized as follows. In Section 2, a correlated LSM prior model for BSS is given. Then, the proposed block Bayesian recovery algorithm and the fast version are derived in Section 3. Simulation experiments are presented in Section 4. Finally, we conclude this paper in Section 5.

2. Signal Model

In the framework of sparse Bayesian learning, for the measurement model shown in (1), the noise $\mathbf{n}$ is generally assumed to obey a Gaussian prior distribution $\mathcal{N}(\mathbf{n} \mid 0, \gamma^{-1}\mathbf{I}_M)$, and a Gamma distribution for the hyperparameter γ is

$$p(\gamma ; a, b) = G(\gamma ; a, b) = \frac{b^a}{\Gamma(a)} \gamma^{a-1} \exp(-b\gamma), \quad (2)$$

where $\Gamma(a) = \int_0^{+\infty} x^{a-1} \exp(-x) dx$. The sparse signal $\mathbf{x}$ is usually assumed to obey a sparse prior distribution. In the LSM layered prior model [12], the sparse signal $\mathbf{x}$ is supposed to follow a Laplacian prior distribution:

$$p(\mathbf{x} \mid \boldsymbol{\lambda}) = \prod_{n=1}^N \mathrm{La}(x_n \mid 0, \lambda_n) = \prod_{n=1}^N \frac{1}{2\lambda_n} \exp\left(-\frac{|x_n|}{\lambda_n}\right), \quad (3)$$

where $\lambda_n (n = 1, \cdots, N)$ is the scale parameter of the Laplacian distribution for each element in the signal. Since the Inverse-Gamma (IG) distribution is conjugated to Laplacian distribution, the LSM model assumes that the scale parameter λ_n obeys the IG distribution as follows:

$$p(\boldsymbol{\lambda} ; c, d) = \prod_{n=1}^N \mathrm{IG}(\lambda_n ; c, d) = \prod_{n=1}^N \frac{d^c}{\Gamma(c)} \lambda_n^{-c-1} \exp\left(-\frac{d}{\lambda_n}\right). \quad (4)$$

In the above model, each hyperparameter λ_n controls the corresponding signal element x_n individually and each signal element is considered to be independent. Considering that nonzero elements of BSS appear in clusters, this requires a more appropriate model for BSS. For the BSS whose structure prior information is unknown, the PC-SBL algorithm [11] assumes that the hyperparameters of adjacent elements have a certain influence on its sparsity. Inspired by the PC-SBL, we assume that the block sparse signal $\mathbf{x}$ obeys the following correlated Laplacian prior distribution, i.e.,

$$\begin{aligned} p(\mathbf{x} \mid \boldsymbol{\lambda}) &= \prod_{n=1}^N \mathrm{La}(x_n \mid 0, \lambda_{n-1}, \lambda_n, \lambda_{n+1}) \\ &= \prod_{n=1}^N \frac{(1/\lambda_n) + \beta(1/\lambda_{n-1}) + \beta(1/\lambda_{n+1})}{2} \exp \\ &\quad \cdot \left(-\left(\frac{1}{\lambda_n} + \beta\frac{1}{\lambda_{n-1}} + \beta\frac{1}{\lambda_{n+1}}\right)|x_n|\right), \end{aligned} \quad (5)$$

where the parameter $\beta \in [0, 1]$ indicates the degree of correlation between adjacent elements in the signal. It can be seen from (5) that the element x_n is affected by its own hyperparameter λ_n and the neighboring ones λ_{n-1} and λ_{n+1}. For the elements at both ends x_1 and x_N, let $\lambda_0 = 0$ and $\lambda_{N+1} = 0$. The model (5) makes use of the feature of the block sparse signal, in which the scale parameters $\boldsymbol{\lambda}$ still obey the IG distribution shown in (4).

3. Block Bayesian Recovery Algorithms

In Bayesian inference, given observation $\mathbf{y}$, it needs to derive the posterior probability density for all unknown parameters $p(\mathbf{x}, \boldsymbol{\lambda}, \gamma \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x}, \gamma) p(\mathbf{x} \mid \boldsymbol{\lambda}) p(\boldsymbol{\lambda}) p(\gamma)$. Variational Bayesian inference is a widely used method to approximately solve the maximization of a posteriori, which assumes that the variables $\mathbf{x}$, $\boldsymbol{\lambda}$, and γ are independent of each other. Let $\boldsymbol{\theta} = \{\mathbf{x}, \boldsymbol{\lambda}, \gamma\}$, and then

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \approx q(\boldsymbol{\theta}) = \prod_i q(\theta_i) = q(\mathbf{x})q(\boldsymbol{\lambda})q(\gamma). \quad (6)$$

For each of these latent variables, the approximate posterior distribution may be computed in an alternating manner as follows:

$$q(\theta_i) = \frac{\exp\left(\langle \ln p(\mathbf{y}, \boldsymbol{\theta})\rangle_{q(\theta_k), k \neq i}\right)}{\int \exp\left(\langle \ln p(\mathbf{y}, \boldsymbol{\theta})\rangle_{q(\theta_k), k \neq i}\right) d\theta_i}, \quad (7)$$

where $\langle \cdot \rangle_{q(\theta_k), k \neq i}$ represents the expected operation with respect to the distributions $q(\theta_k), k \neq i$. According to (7), the proposed reconstruction algorithm is derived by alternately learning the updating rules of these latent variables.

Firstly, the approximate posterior distribution $q(\mathbf{x})$ is

$$\ln q(\mathbf{x}) \propto \langle \ln p(\mathbf{y} \mid \mathbf{x}, \gamma) + \ln p(\mathbf{x} \mid \boldsymbol{\lambda})\rangle_{q(\boldsymbol{\lambda})q(\gamma)}, \quad (8)$$

where $p(\mathbf{y} \mid \mathbf{x}, \gamma)$ is a Gaussian distribution $\mathcal{N}(\boldsymbol{\Phi}\mathbf{x}, \gamma^{-1}\mathbf{I}_M)$ and $p(\mathbf{x} \mid \boldsymbol{\lambda})$ is a Laplacian distribution shown by (5). Since these two distributions are not conjugated, a direct solution is difficult. Similar to [12], let

$$\begin{aligned} L(\mathbf{x}) &= -\langle \ln p(\mathbf{y} \mid \mathbf{x}, \gamma) + \ln p(\mathbf{x} \mid \boldsymbol{\lambda})\rangle_{q(\boldsymbol{\lambda})q(\gamma)} \propto \frac{\langle\gamma\rangle}{2} \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 \\ &\quad + \sum_{n=1}^N \left\langle \frac{1}{\lambda_n} + \beta\frac{1}{\lambda_{n-1}} + \beta\frac{1}{\lambda_{n+1}} \right\rangle |x_n|. \end{aligned} \quad (9)$$

The maximum a posteriori (MAP) estimate of the signal $\mathbf{x}$ can be obtained by $\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\min_{\mathbf{x}}\{L(\mathbf{x})\}$. The derivative of $L(\mathbf{x})$ is

$$\nabla_{\mathbf{x}} L(\mathbf{x}) = (\langle\gamma\rangle\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mathbf{D})\mathbf{x} - \langle\gamma\rangle\boldsymbol{\Phi}^T\mathbf{y}, \quad (10)$$

where

$$\mathbf{D} = \mathrm{diag} \left[\left(\left\langle \frac{1}{\lambda_n} \right\rangle + \beta \left\langle \frac{1}{\lambda_{n-1}} \right\rangle + \beta \left\langle \frac{1}{\lambda_{n+1}} \right\rangle \right) \frac{1}{|x_n|} \right], \quad (11)$$

and diag [·] denotes a diagonal matrix with the elements in the bracket. Let the derivative be equal to zero and get the approximate MAP estimate:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \langle \gamma \rangle \left(\langle \gamma \rangle \mathbf{\Phi}^T \mathbf{\Phi} + \mathbf{D} \right)^{-1} \mathbf{\Phi}^T \mathbf{y}. \quad (12)$$

So the posterior distribution can be approximated by using the second-order Taylor expansion around $\hat{\mathbf{x}}_{\mathrm{MAP}}$, i.e.,

$$\ln q(\mathbf{x}) \approx \ln q(\hat{\mathbf{x}}_{\mathrm{MAP}}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}\wedge_{\mathrm{MAP}})^T H(\hat{\mathbf{x}}_{\mathrm{MAP}})(\mathbf{x} - \hat{\mathbf{x}}_{\mathrm{MAP}}), \quad (13)$$

where $H(\hat{\mathbf{x}}_{\mathrm{MAP}}) \approx -(\langle \gamma \rangle \mathbf{\Phi}^T \mathbf{\Phi} + \mathbf{D})$. After similar simplification in [12], $q(\mathbf{x})$ can be approximated to obey the Gaussian distribution $\mathcal{N}(\mathbf{x} \mid \mathbf{\mu}, \mathbf{\Sigma}_o)$ with the mean $\mathbf{\mu} = \langle \gamma \rangle \mathbf{\Sigma}_o \mathbf{\Phi}^T \mathbf{y}$ and covariance matrix $\mathbf{\Sigma}_o = (\langle \gamma \rangle \mathbf{\Phi}^T \mathbf{\Phi} + \mathbf{D})^{-1}$. Due to that, there exists some approximation in the above derivation of the posterior distribution $q(\mathbf{x})$; the sparsity of the signal may be underestimated. So a parameter $\alpha \in [0.5, 1]$ is introduced into the computation of the covariance matrix. Thus, $q(\mathbf{x})$ is approximated to be the following Gaussian distribution:

$$q(\mathbf{x}) \approx \mathcal{N}(\mathbf{x} \mid \mathbf{\mu}, \mathbf{\Sigma}), \quad (14)$$

where the mean and covariance matrix, respectively, are

$$\begin{aligned} \mathbf{\mu} &= \langle \gamma \rangle \mathbf{\Sigma} \mathbf{\Phi}^T \mathbf{y}, \\ \mathbf{\Sigma} &= \left(\langle \gamma \rangle \mathbf{\Phi}^T \mathbf{\Phi} + \alpha \mathbf{D} \right)^{-1}. \end{aligned} \quad (15)$$

Secondly, the approximate posterior distribution $q(\mathbf{\lambda})$ is

$$\ln q(\mathbf{\lambda}) \propto \langle \ln p(\mathbf{\lambda}) + \ln p(\mathbf{x} \mid \mathbf{\lambda}) \rangle_{q(\mathbf{x})} = \ln p(\mathbf{\lambda}) + \langle \ln p(\mathbf{x} \mid \mathbf{\lambda}) \rangle_{q(\mathbf{x})}. \quad (16)$$

From (4) and (5), we have

$$\ln p(\mathbf{\lambda}) \propto \sum_{n=1}^{N} \left\{ (c+1) \ln \frac{1}{\lambda_n} - \frac{d}{\lambda_n} \right\}, \quad (17)$$

$$\begin{aligned} \langle \ln p(\mathbf{x} \mid \mathbf{\lambda}) \rangle_{q(\mathbf{x})} &\propto \sum_{n=1}^{N} \left\{ \ln \left(\frac{1}{\lambda_n} + \beta \frac{1}{\lambda_{n-1}} + \beta \frac{1}{\lambda_{n+1}} \right) \right. \\ &\quad \left. - \left(\frac{1}{\lambda_n} + \beta \frac{1}{\lambda_{n-1}} + \beta \frac{1}{\lambda_{n+1}} \right) \langle |x_n| \rangle \right\} \\ &= \sum_{n=1}^{N} \left\{ \ln \left(\frac{1}{\lambda_n} + \beta \frac{1}{\lambda_{n-1}} + \beta \frac{1}{\lambda_{n+1}} \right) \right. \\ &\quad \left. - (\langle |x_n| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle) \frac{1}{\lambda_n} \right\} \\ &\geq \sum_{n=1}^{N} \left\{ \ln \frac{1}{\lambda_n} - (\langle |x_n| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle) \frac{1}{\lambda_n} \right\}. \end{aligned}$$
$$(18)$$

So $q(\mathbf{\lambda})$ can be approximated as

$$\begin{aligned} \ln q(\mathbf{\lambda}) \propto \sum_{n=1}^{N} &\left\{ [(c+1)+1] \ln \frac{1}{\lambda_n} \right. \\ &\left. - [(\langle |x_n| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle) + d] \frac{1}{\lambda_n} \right\}. \end{aligned} \quad (19)$$

Therefore, $q(\mathbf{\lambda})$ obeys an Inverse-Gamma distribution:

$$p\left(\mathbf{\lambda}; \tilde{c}, \tilde{d} \right) = \prod_{n=1}^{N} \mathrm{IG}\left(\lambda_n; \tilde{c}, \tilde{d}_n \right), \quad (20)$$

with

$$\begin{aligned} \tilde{c} &= c+1, \\ \tilde{d}_n &= (\langle |x_n| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle) + d. \end{aligned} \quad (21)$$

We can also obtain

$$\left\langle \frac{1}{\lambda_n} \right\rangle = \frac{\tilde{c}}{\tilde{d}_n} = \frac{c+1}{(\langle |x_n| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle) + d}, \quad (22)$$

where $\langle |x_n| \rangle$ can be computed as follows according to [13]:

$$\langle |x_n| \rangle = \sqrt{\frac{2\Sigma_{n,n}}{\pi}} \exp \left(-\frac{\mu_n^2}{2\Sigma_{n,n}} \right) + |\mu_n| \, \mathrm{erf} \left(\sqrt{\frac{\mu_n^2}{2\Sigma_{n,n}}} \right), \quad (23)$$

and $\mathrm{erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$ is the error function.

Thirdly, the approximate posterior distribution of the noise parameter $q(\gamma)$ is

$$\begin{aligned} \ln q(\gamma) \propto \langle \ln p(\gamma) + \ln p(\mathbf{y} \mid \mathbf{x}, \gamma) \rangle_{q(\mathbf{x})} &\propto \left(a - 1 + \frac{M}{2} \right) \ln \gamma \\ &- \left(\frac{\langle \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 \rangle}{2} + b \right) \gamma. \end{aligned}$$
$$(24)$$

Input: $\mathbf{y}, \boldsymbol{\Phi}, \varepsilon$, Max_iter, β, α.
Initialize: $\boldsymbol{\mu}^0, \boldsymbol{\lambda}, \gamma, a, b, c, d$
While $\left\| \boldsymbol{\mu}^{t+1} - \boldsymbol{\mu}^t \right\|_2 > \varepsilon$ or $t \leq$ Max_iter **do**
Update:
(1) Compute the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ by (15) and (11).
(2) Compute $\boldsymbol{\lambda}^{-1}$ according to (22) and (23).
(3) Compute γ via (27).
Output: $\hat{\mathbf{x}} = \boldsymbol{\mu}$

ALGORITHM 1: Block EM-VB.

Input: $\mathbf{y}, \boldsymbol{\Phi}, \varepsilon$, Max_iter, β, α
Initialize: $\boldsymbol{\mu}^0, \boldsymbol{\lambda}, \gamma, a, b, c, d$
While $\left\| \boldsymbol{\mu}^{t+1} - \boldsymbol{\mu}^t \right\|_2 > \varepsilon$ or $t \leq$ Max_iter **do**
Update:
(1) Compute the mean μ_i and the variance σ_i^2 ($i = 1, 2, \cdots, N$), sequentially by (35).
(2) Compute $\boldsymbol{\lambda}^{-1}$ according to (22) and (23).
(3) Compute γ via (27) and (36).
Output: $\hat{\mathbf{x}} = \boldsymbol{\mu}$

ALGORITHM 2: Fast Block EM-VB.
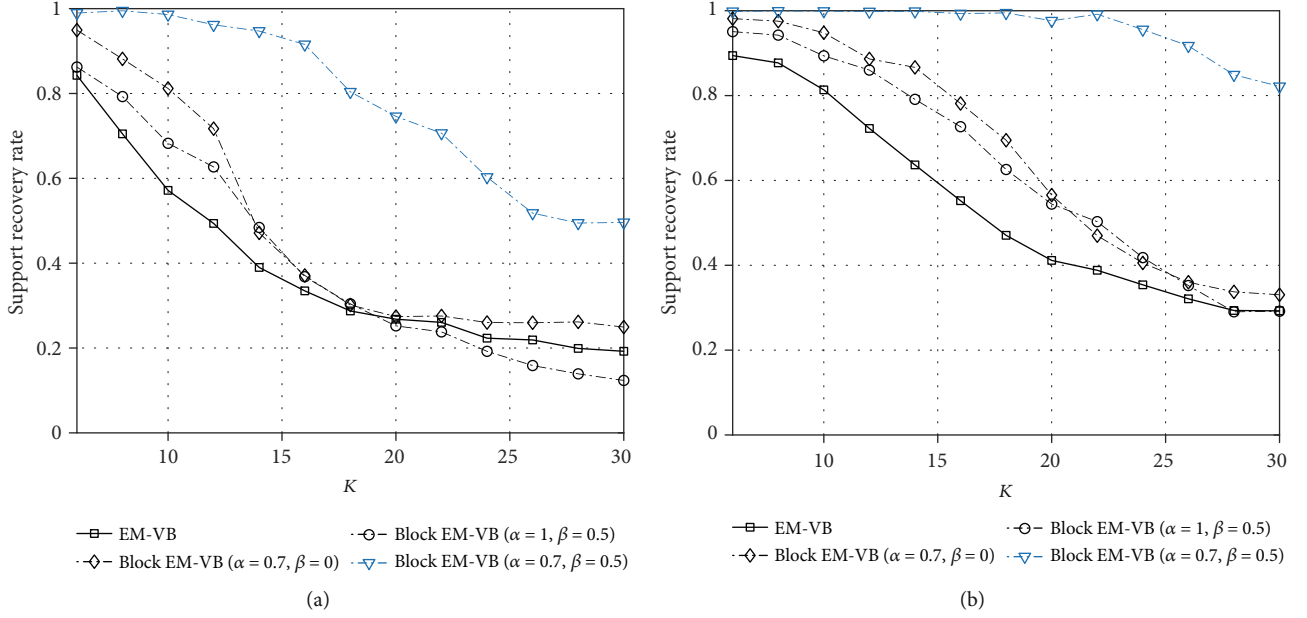
FIGURE 1: The support recovery rate versus the sparsity in the noiseless case when (a) $M = 30$ and (b) $M = 40$.

So $q(\gamma)$ obeys the Gamma distribution:

$$q(\gamma) = G\left(\gamma; \tilde{a}, \tilde{b}\right), \tag{25}$$

with

$$\tilde{a} = a + \frac{M}{2},$$
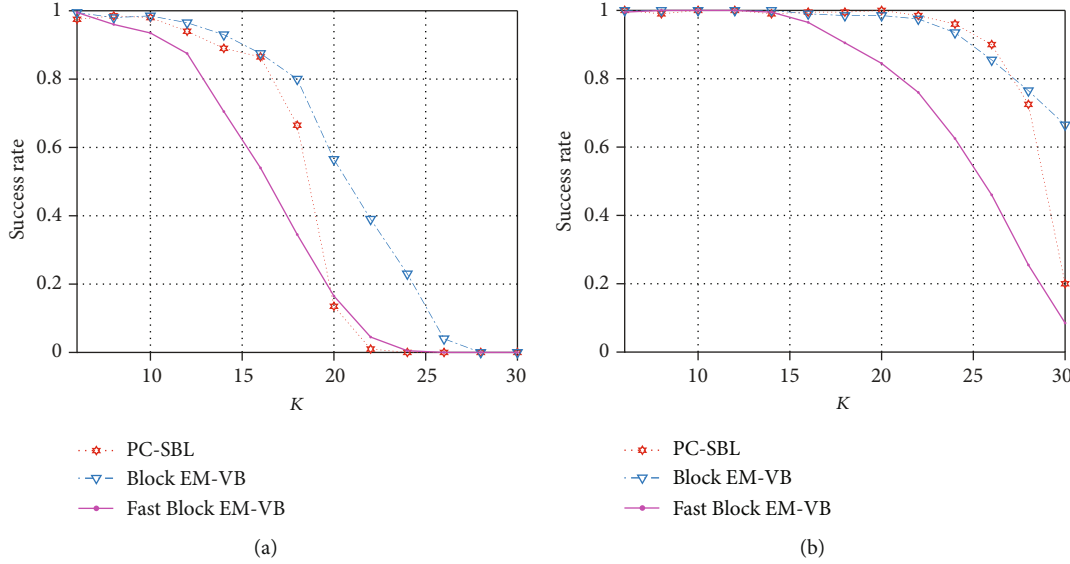
$$\tilde{b} = \frac{\left\langle \left\| \mathbf{y} - \boldsymbol{\Phi} \mathbf{x} \right\|_2^2 \right\rangle}{2} + b, \tag{26}$$

Figure 2: The success rate vs. the sparsity level in the noiseless case when (a)$M = 30$ and (b) $M = 40$.
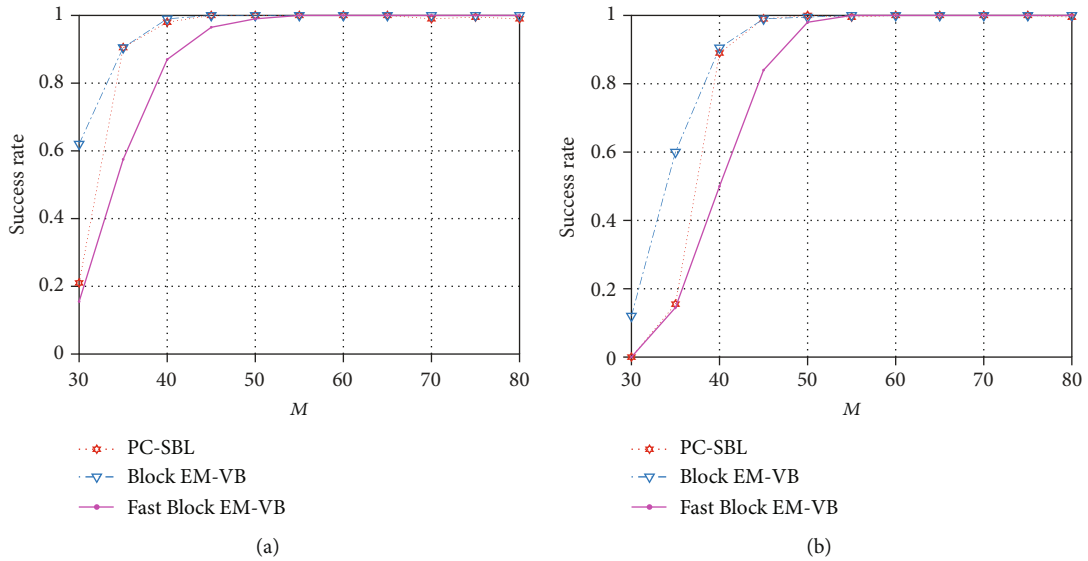
Figure 3: The success rate vs. the number of measurements in the noiseless case when (a) $K = 20$ and (b) $K = 25$.

and we can obtain

$$\langle \gamma \rangle = \frac{\tilde{a}}{\tilde{b}} = \frac{a + (M/2)}{\left(\langle \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 \rangle /2 \right) + b}, \tag{27}$$

where $\langle \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 \rangle = \|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\mu}\|_2^2 + \mathrm{Tr}[\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\boldsymbol{\Phi}]$ and $\mathrm{Tr}(\cdot)$ is the trace of a matrix.

Therefore, the whole process of the proposed algorithm is summarized in Algorithm 1, where ε is the preset error that can be tolerated. The proposed algorithm can be regarded as an extension of the EM-VB algorithm for the recovery of BSS, which is termed the Block EM-VB algorithm. It has an additional parameter α except for the block

parameter β. When $\alpha = 1$ and $\beta = 0$, the Block EM-VB algorithm reduces to the EM-VB. The parameters α in (15) and β in (22) have great influences on the recovery performance of the Block EM-VB. It is appropriate to set $\alpha < 1$ to void underestimation of the support set if the sparsity of the signal is larger, while larger α can suppress nonzero signal elements and has certain antinoise capacity in noisy environments. For the choice of β, it is similar that larger β may enhance the influence between adjacent elements and could suppress the nonzero values to make the signal more sparse in noisy environments.

Remark 1. From (23), it can be seen that the computation of $\langle |x_n| \rangle$ is related to the erf (x) involving integral operation. In
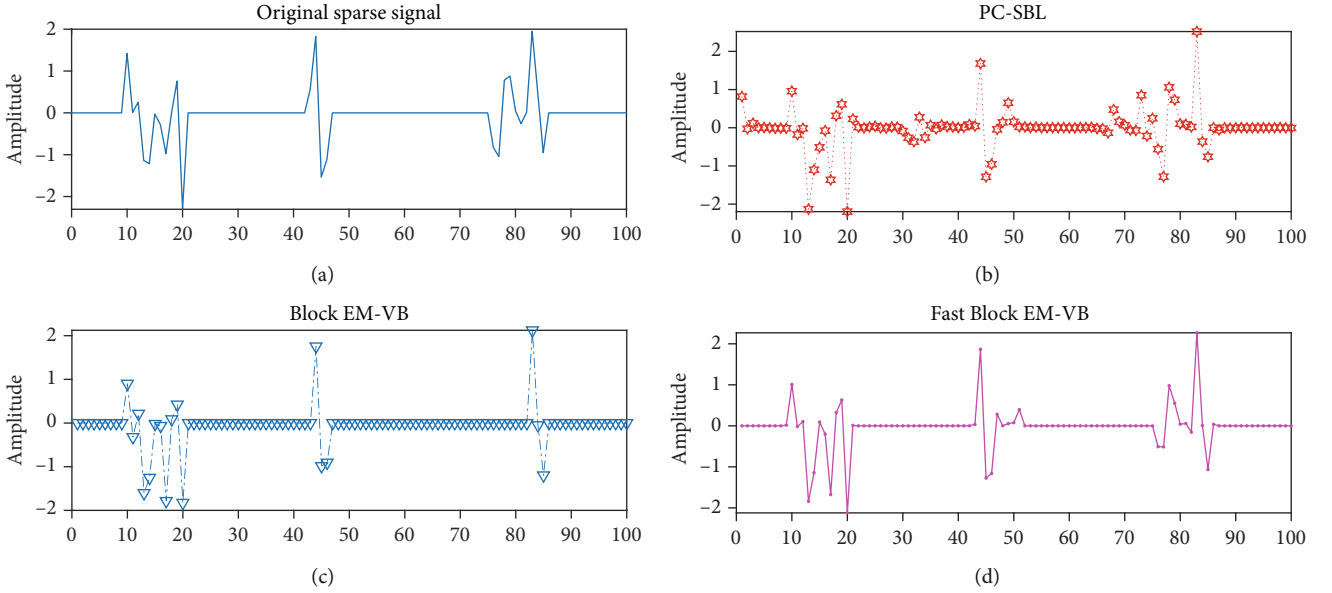
FIGURE 4: The reconstructed signal by different algorithms when $N = 100$, $M = 50$, $K = 25$, and SNR = 10 dB. (a) Original signal, (b) reconstructed signal by PC-SBL with NMSE = 0.2362, (c) reconstructed signal by Block EM-VB with NMSE = 0.2184, and (d) reconstructed signal by Fast Block EM-VB with NMSE = 0.1092.
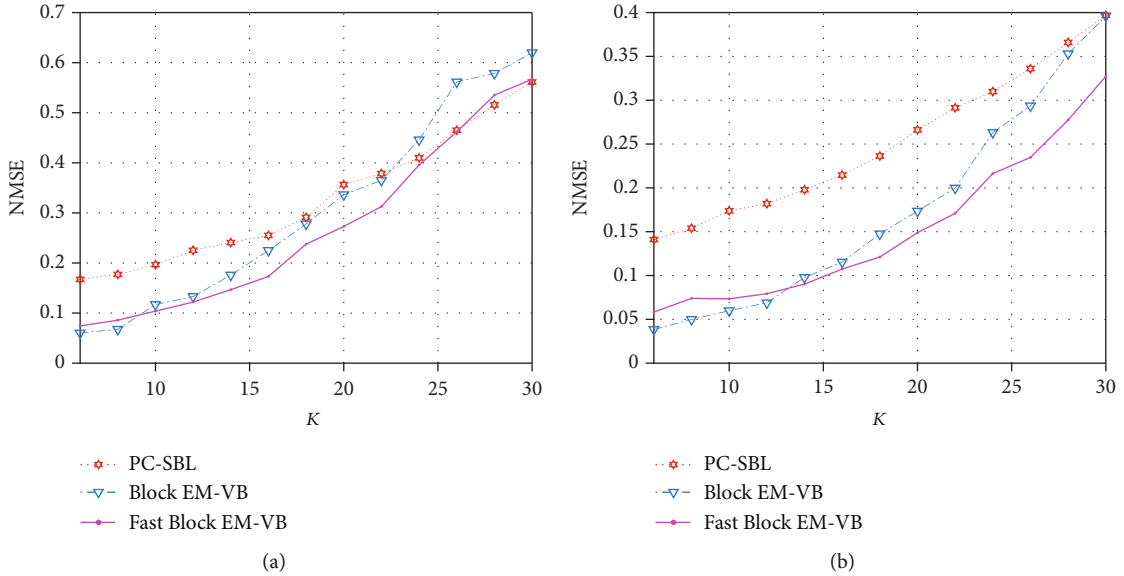
FIGURE 5: The NMSE vs. the sparsity level when SNR = 10 dB and (a) $M = 40$ and (b) $M = 50$.

practice, to reduce the complexity, erf (x) can be calculated by utilizing some approximation equations such as

$$\text{erf } (x) \approx \left\{ 1 - e^{-4x^2/\pi} \left[1 + \frac{8x^4}{\pi} \left(\frac{1}{3} - \frac{1}{\pi} \right) \right] \right\}^{1/2}. \quad (28)$$

From the process of the Block EM-VB algorithm, it can be seen that its complexity is almost the same as that of the PC-SBL.

The proposed Block EM-VB algorithm involves the matrix inversion shown in (15), which is the main computa-

tional complexity. It is better to consider the fast version of the proposed algorithm. To void computation of the inverse of the matrix, we can use the following approximate posterior distribution, which is expressed as

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \approx q(\boldsymbol{\theta}) = \prod_{i=1}^{N} q(x_i) q(\boldsymbol{\lambda}) q(\gamma). \quad (29)$$

That is to say, it assumes the independence on the posterior of each coefficient element of the signal. Similarly, by using (7), we can alternately learn the updating rules of these latent variables.

(a)

(b)
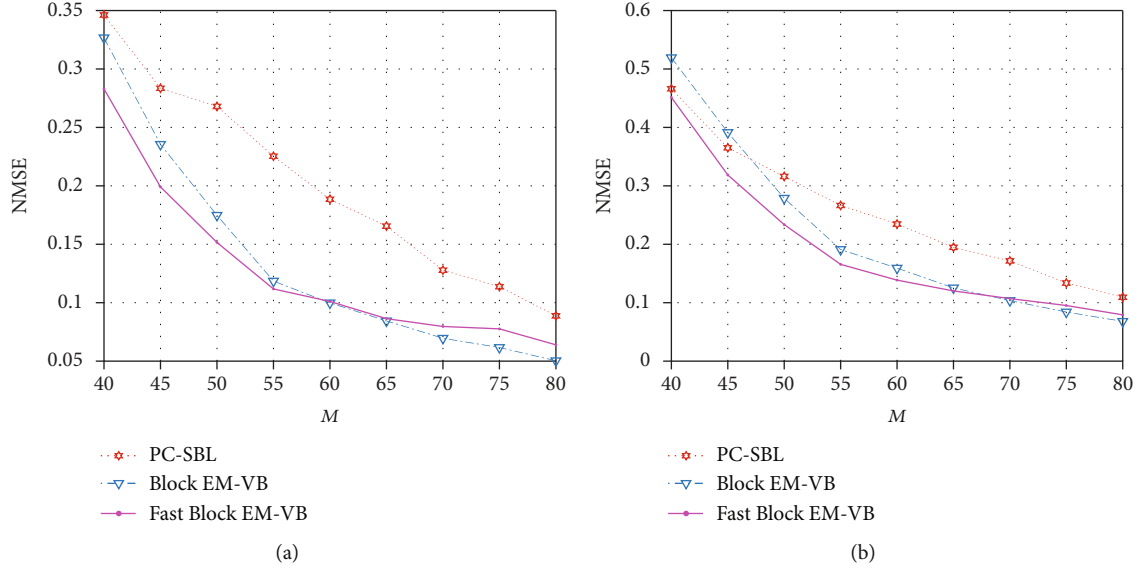
FIGURE 6: The NMSE vs. the number of measurements when SNR = 10 dB and (a) $K = 20$ and (b) $K = 25$.

Firstly, the approximate posterior distribution $q(x_i)$ $(i = 1, 2, \cdots, N)$ is

$$\ln q(x_i) \propto \langle \ln p(\mathbf{y} \mid \mathbf{x}, \gamma) + \ln p(x_i \mid \lambda_{i-1}, \lambda_i, \lambda_{i+1}) \rangle_{q(\gamma)q(\boldsymbol{\lambda})q(x_j)_{j\neq i}}, \tag{30}$$

where $p(\mathbf{y} \mid \mathbf{x}, \gamma) = \mathcal{N}(\boldsymbol{\Phi}\mathbf{x}, \gamma^{-1}\mathbf{I}_M)$ and $p(x_i \mid \lambda_{i-1}, \lambda_i, \lambda_{i+1})$ $= (((1/\lambda_i) + \beta(1/\lambda_{i-1}) + \beta(1/\lambda_{i+1}))/2) \exp(-((1/\lambda_i) + \beta(1/\lambda_{i-1}) + \beta(1/\lambda_{i+1}))|x_i|)$. Let

$$\begin{aligned} l(x_i) &= -\langle \ln p(\mathbf{y} \mid \mathbf{x}, \gamma) + \ln p(x_i \mid \lambda_{i-1}, \lambda_i, \lambda_{i+1}) \rangle_{q(\gamma)q(\boldsymbol{\lambda})q(x_j)_{j\neq i}} \propto \frac{\langle \gamma \rangle}{2} \langle \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 \rangle_{q(x_j)_{j\neq i}} \\ &+ \left\langle \frac{1}{\lambda_i} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle |x_i| \propto \frac{\langle \gamma \rangle}{2} \langle -2\mathbf{y}^T \boldsymbol{\Phi}\mathbf{x} + \mathbf{x}^T \boldsymbol{\Phi}^T \boldsymbol{\Phi}\mathbf{x} \rangle_{q(x_j)_{j\neq i}} \\ &+ \left\langle \frac{1}{\lambda_i} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle |x_i| \propto \frac{\langle \gamma \rangle}{2} \left(-2\mathbf{y}^T \varphi_i x_i + \varphi_i^T \varphi_i x_i^2 + 2x_i \varphi_i^T \sum_{j\neq i} \varphi_j \langle x_j \rangle \right) \\ &+ \left\langle \frac{1}{\lambda_i} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle |x_i|, \end{aligned} \tag{31}$$

where φ_i is the i-th column of $\boldsymbol{\Phi}$. The derivative of $l(x_i)$ is

$$\begin{aligned} \nabla_{x_i} l(x_i) &= \frac{\langle \gamma \rangle}{2} \left(-2\mathbf{y}^T \varphi_i + 2\varphi_i^T \varphi_i x_i + 2\varphi_i^T \sum_{j\neq i} \varphi_j \langle x_j \rangle \right) \\ &+ \left\langle \frac{1}{\lambda_i} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle \frac{x_i}{|x_i|}. \end{aligned} \tag{32}$$

Let $\nabla_{x_i} l(x_i) = 0$ and obtain the following approximate MAP estimate:

$$\tilde{x}_i = \langle \gamma \rangle \left(\langle \gamma \rangle \varphi_i^T \varphi_i + \left\langle \frac{1}{\lambda_i} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle \frac{1}{|x_i|} \right)^{-1} \left(\mathbf{y} - \sum_{j\neq i} \varphi_j \langle x_j \rangle \right)^T \varphi_i. \tag{33}$$
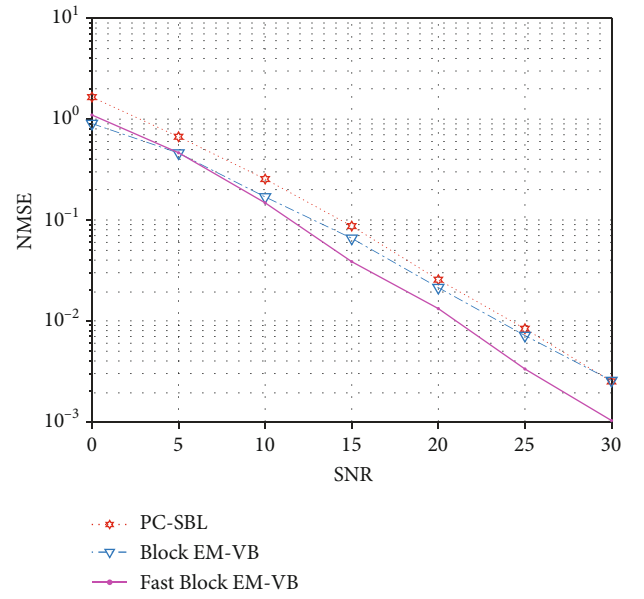
FIGURE 7: The comparison of different algorithms under different SNR when $N = 100$, $K = 20$, and $M = 50$.

Similarly, $q(x_i)$ is approximated to obey the following Gaussian distribution:

$$q(x_i) \approx \mathcal{N}(x_i \mid \mu_i, \sigma_i^2), \tag{34}$$

where

$$\mu_i = \langle \gamma \rangle \sigma_i^2 \left(\mathbf{y} - \sum_{j\neq i} \varphi_j \langle x_j \rangle \right)^T \varphi_i,$$

$$\sigma_i^2 = \left(\langle \gamma \rangle \varphi_i^T \varphi_i + \left\langle \frac{1}{\lambda_i} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle \frac{\alpha}{|x_i|} \right)^{-1}, \tag{35}$$

in which a parameter α is also introduced to avoid underestimation of the sparsity of the signal. Thus, $\langle x_i \rangle = \mu_i$ $(i = 1, 2, \cdots, N)$ can be computed in a sequential manner.

Secondly, the approximate posterior distribution $q(\boldsymbol{\lambda})$ still obeys an Inverse-Gamma distribution shown in (20) and the computation of $\langle \boldsymbol{\lambda}^{-1} \rangle$ is the same as (22).

Thirdly, the approximate posterior distribution $q(\gamma)$ obeys the Gamma distribution $G(\gamma; \tilde{a}, \tilde{b})$ and we can obtain $\langle \gamma \rangle = \tilde{a}/\tilde{b} = (a + M/2)/(\langle \| \mathbf{y} - \boldsymbol{\Phi}\mathbf{x} \|_2^2 \rangle/2 + b)$, where

$$
\begin{aligned}
\langle \| \mathbf{y} - \boldsymbol{\Phi}\mathbf{x} \|_2^2 \rangle &= \| \mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\mu} \|_2^2 + \mathrm{Tr}\left[\boldsymbol{\Sigma}\boldsymbol{\Phi}^T \boldsymbol{\Phi} \right] \\
&= \| \mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\mu} \|_2^2 + \sum_{i=1}^{N} \sigma_i^2 \varphi_i^T \varphi_i,
\end{aligned}
\tag{36}
$$

and $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_N)^T$ and $\boldsymbol{\Sigma} \triangleq \mathrm{diag}\left[\sigma_i^2\right]$.

Compared with the Block EM-VB algorithm, the above algorithm has low computational complexity due to without matrix inversion in each iteration. It can be called Fast Block EM-VB, and its process is summarized in Algorithm 2. In noisy environments, the correlations between adjacent signal elements may be weakened, especially in the low SNR case, so it is appropriate to assume independence on the posterior of each coefficient element, which implies that the Fast Block EM-VB is suitable to recover the BSS under the low SNR case.

4. Simulation Experiments

In this section, some simulation experiments are carried out to demonstrate the performances of the proposed Block EM-VB algorithm and its fast version. A comparison with other algorithms such as EM-VB [12] and PC-SBL [11] is also given.

4.1. Performance Analysis via Simulated Data. In the following simulation, let the length of the sparse signal be $N = 100$, and an arbitrary block Gaussian sparse signal is randomly generated with its nonzero entries randomly distributed in $B = 3$ blocks, and the measurement matrix is a random Gaussian matrix. The number of Monte Carlo simulations is 200. The parameters in EM-VB and in the Block EM-VB and its fast version are set as $a = b = c = d = 10^{-6}$, Max_iter = 200, and $\varepsilon = 10^{-5}$. The parameters α and β in the Block EM-VB and its fast version will be set adaptively according to the noiseless and noisy cases because their performances are sensitive to the choice of α and β. The parameters of PC-SBL are set the same as those in [11].

First, we discuss the influences of two parameters α and β on the performance of the proposed Block EM-VB and make a comparison with the EM-VB, which is a special case of the Block EM-VB when $\alpha = 1$ and $\beta = 0$. To demonstrate the effect of these two parameters, the performances of the Block EM-VB with $\alpha = 0.7$ and $\beta = 0$ and $\alpha = 1$ and $\beta = 0.5$ are also given. The support recovery rate is used to evaluate the performance of the Block EM-VB algorithm with different parameters. The recovered support of the sparse signal is defined as supp $(\widehat{\mathbf{x}}) = \{i, |\widehat{x}_i| > 0.001\}$, and then the support
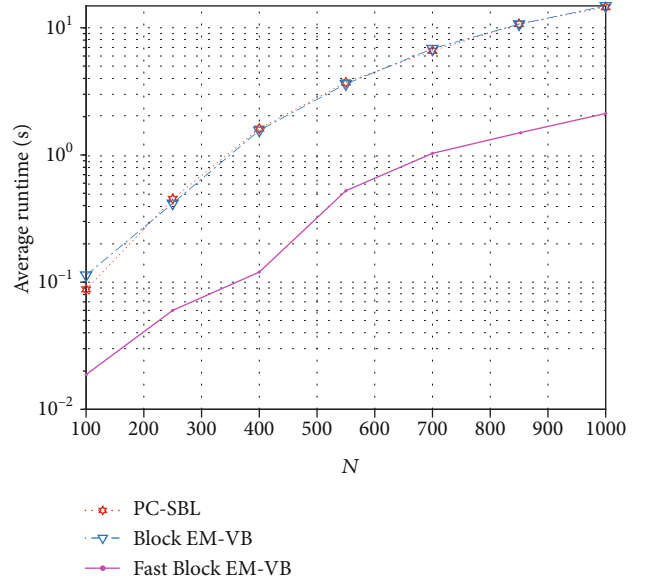
FIGURE 8: The average runtimes vs. the length of signal N.

recovery rate is defined by $|\mathrm{supp}\,(\widehat{\mathbf{x}}) \cap \mathrm{supp}\,(\mathbf{x})|/|\mathrm{supp}\,(\widehat{\mathbf{x}}) \cup \mathrm{supp}\,(\mathbf{x})|$, where $|\cdot|$ denotes the number of elements in a set. If the overlap between the estimated support and the true support is more, the recovery rate is closer to 1. Figure 1 plots the support recovery rates of different algorithms versus the sparsity level K when the number of measurements is $M = 30$ and $M = 40$, respectively. It can be seen that the parameters of the Block EM-VB algorithm have an important influence on the recovery performance. The appropriate parameter $\alpha < 1$ can avoid the underestimation of the support of the sparse signals, and the block parameter $\beta > 0$ is helpful to recover the block sparse signals. Thus, the proposed Block EM-VB algorithm with appropriate parameters has better performance than the EM-VB.

Then, we make a comparison between the Block EM-VB and its fast version with the PC-SBL. In the noiseless case, the success rate is used to evaluate the performances of these different algorithms. When a trial satisfies $\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2/\|\mathbf{x}\|_2^2 < 10^{-6}$, it is regarded as a successful trial. The success rate is defined as the percentage of successful trials in the total of independent trials. In the noisy case, the reconstruction performance of each algorithm is evaluated by the normalized mean square error (NMSE), where NMSE = $\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2/\|\mathbf{x}\|_2^2$.

In the noiseless case, let $\alpha = 0.7$ and $\beta = 0.5$ in the Block EM-VB and $\alpha = 0.5$ and $\beta = 0.5$ in the Fast Block EM-VB. Figure 2 plots the success rate of individual recovery algorithm versus the sparsity level K when the number of measurement is $M = 30$ and $M = 40$, respectively. Then, let the sparsity $K = 20$ and $K = 25$, and the success rate of each algorithm versus the number of measurements M is shown in Figure 3. From these results, it is observed that the Block EM-VB is superior to the PC-SBL when the number of measurements is less or the sparsity is smaller. The Fast Block EM-VB is inferior to these two algorithms due to the independent assumption.

Then, we consider the noisy case where the Gaussian noise is added to the measurements with the signal-to-
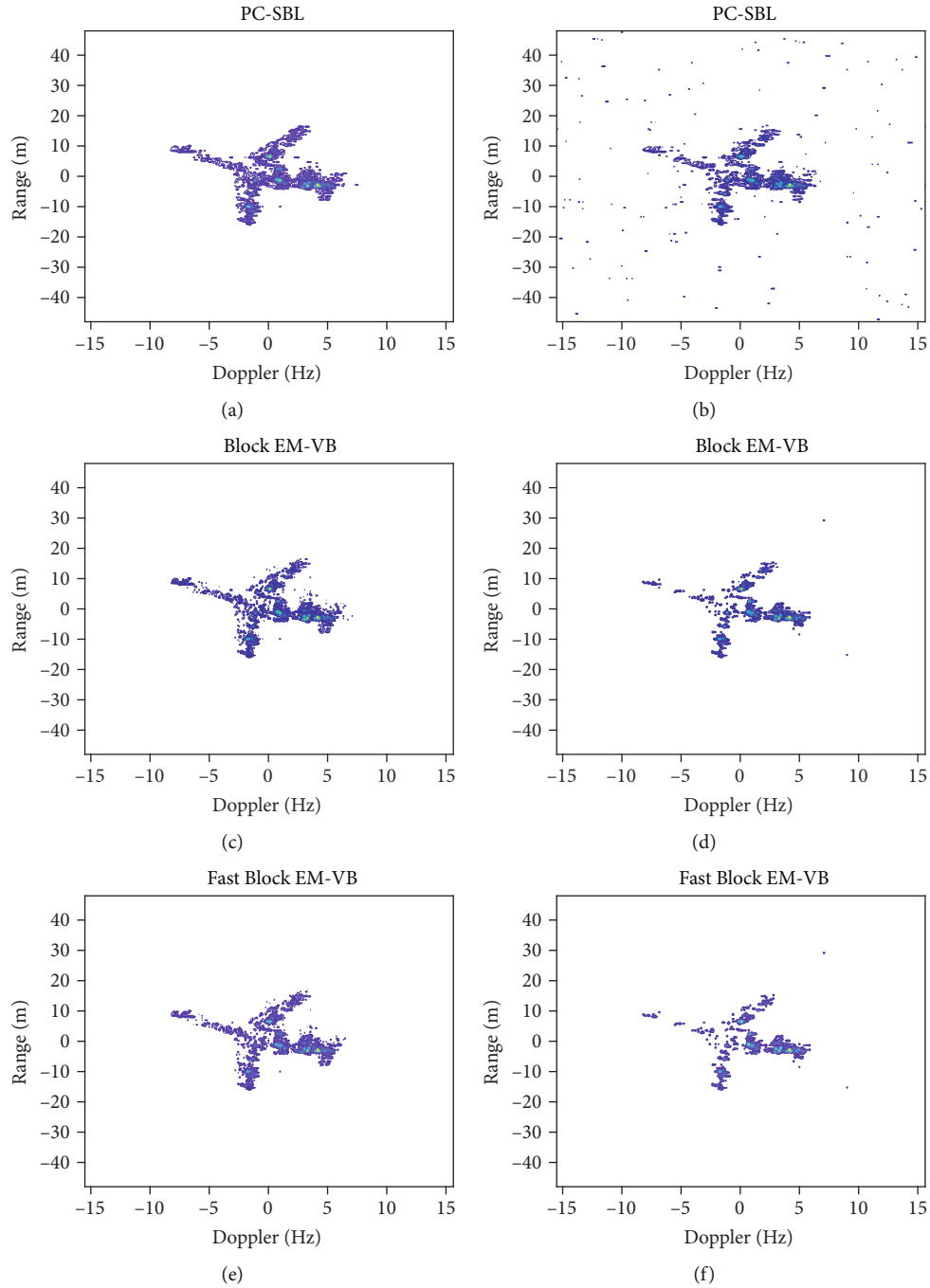
FIGURE 9: Reconstructed images with different algorithms (the left column is the noiseless case, and the right column is the noisy case).

noise ratio (SNR) defined as $20 \log_{10}(\|\mathbf{\Phi x}\|_2/\|\mathbf{n}\|_2)$. It should be noted that the setting of parameter α in the proposed algorithms needs to consider the trade-off between removing the noise and maintaining the nonzero elements of the signal. A larger value of α tends to suppress noise while losing the nonzero values of the signal. Compared with the Block EM-VB, the fast version needs to select smaller α to ensure the recovery of nonzero elements of the sparse signal because of its inherent denoising ability. Let SNR = 10 dB and $\alpha = 0.85$ and $\beta = 0.5$ in the Block

EM-VB and $\alpha = 0.5$ and $\beta = 0.5$ in the Fast Block EM-VB. The reconstruction results of different algorithms when $K = 25$ and $M = 50$ are given in Figure 4. It can be seen that the Block EM-VB algorithm and its fast version have better reconstruction performance than the PC-SBL in the low SNR environment.

The NMSE of each algorithm versus the sparsity K in the case of SNR = 10 dB is shown in Figure 5, where the numbers of measurements are given as $M = 40$ and $M = 50$, respectively. Figure 6 plots the NMSE of individual recovery

algorithm versus the number of measurements M in the case of SNR = 10 dB, where $K = 20$ and $K = 25$, respectively. It can be found that the Block EM-VB algorithm is superior to the PC-SBL when the number of the measurements is larger or the sparsity is less, and the latter is better than the former when the number of the measurements is less and the sparsity is larger. It is also observed that the Fast Block EM-VB is better than the Block EM-VB when the number of measurements is less or the sparsity is larger. In addition, it is shown that the Fast Block EM-VB outperforms the PC-SBL when the sparsity is less or the number of the measurements is larger.

The performances of each algorithm in different SNR cases are shown in Figure 7, where $K = 20$ and $M = 50$. For the Block EM-VB, the parameters $\beta = 0.5$ and α is set to vary with the SNR, i.e., $\alpha = 0.85$ when the SNR varies from 0 dB to 15 dB, $\alpha = 0.75$ when the SNR changes from 20 dB to 25 dB, and $\alpha = 0.73$ for the case of 30 dB. The parameters in the Fast Block EM-VB algorithm are still set as $\alpha = 0.5$ and $\beta = 0.5$. From Figure 7, it can be seen that the Block EM-VB algorithm and the fast version have good reconstruction performance in the case of low SNR when compared with the PC-SBL. The Fast Block EM-VB especially has robust recovery performance in respect of noise immunity.

Finally, the average runtimes of these algorithms versus the length of signals N by 5 independent trials are given in Figure 8, where $M = N/2$, $K = N/10$, SNR = 10 dB, and the number of nonzero blocks $B = K/5$. It validates that the Block EM-VB almost has considerable computational complexity as the PC-SBL and the Fast Block EM-VB has the highest computational efficiency compared with other recovery algorithms, which makes it have a potential advantage in practical application.

4.2. Application in ISAR Imaging. The inverse synthetic aperture radar (ISAR) imaging is appropriately implemented under the framework of sparse signal recovery due to the sparse characteristic of the target [14]. In this experiment, the "Yak-42" dataset is used, in which the number of range cells is 256 and the number of pulses is 256. 128 pulses are randomly sampled to simulate the sparse aperture data. Here, we use the MATLAB code provided in [14], where the PC-SBL adopts a pruning operation. For a fair comparison, the proposed Block EM-VB and the fast version also use a similar pruning operation and the parameters in these two algorithms are set as Max_iter = 200 and $\varepsilon = 10^{-6}$. Image entropy is usually used to measure image quality in ISAR imaging. The smaller image entropy means better reconstruction performance. The image entropy is defined as

$$\text{Entropy} = -\sum_i \sum_j \left(\frac{|\mathbf{x}(i,j)|^2}{E} \right) \log \left(\frac{|\mathbf{x}(i,j)|^2}{E} \right), \qquad (37)$$

where $E = \sum_i \sum_j |\mathbf{x}(i,j)|^2$ is the energy of the radar image $\mathbf{x}$.

Here, the image is reconstructed by each range cell. Figures 9(a), 9(c), and 9(e) give the reconstruction results of these algorithms in the noiseless case, where we set

TABLE 1: The PSNR results of different algorithms.

	PC-SBL	Block EM-VB	Fast Block EM-VB
Noiseless	5.6384	5.3275	5.4201
Noisy	5.5457	5.0689	5.0059

TABLE 2: Comparison of the runtimes (s) of different algorithms.

	PC-SBL	Block EM-VB	Fast Block EM-VB
Noiseless	27.7130	7.8391	3.3582
Noisy	30.3940	31.6483	6.8828

$\alpha = 1$ and $\beta = 1$ for the Block EM-VB and $\alpha = 0.8$ and $\beta = 1$ for the Fast Block EM-VB. In the noisy case, the data are corrupted by additive Gaussian noise and let $\alpha = 0.85$ and $\beta = 1$ for the Block EM-VB and $\alpha = 0.6$ and $\beta = 1$ for the Fast Block EM-VB. Figures 9(b), 9(d), and 9(f) demonstrate the results in the case of SNR ≈ 3 dB (the noise variance $\gamma^{-1} = 0.0025$). The entropy values of these algorithms are shown in Table 1. From these reconstruction results, it can be seen that the image obtained by the Block EM-VB and its fast version has better quality in the noisy case, which implies that the Block EM-VB and its fast version have strong noise immunity ability. Table 2 gives the corresponding runtimes of these algorithms, which demonstrates that the Fast Block EM-VB has the highest computational efficiency and can be used in real-time processing.

5. Conclusions

Considering the clustered structural features of nonzero elements of block sparse signals, this paper proposes the Block EM-VB algorithm for signal recovery, which is based on a correlated LSM model. Furthermore, a fast version of the Block EM-VB is presented, which can recover the block sparse signals with lower computational complexity because of no inversion in each iteration. Experimental results with simulation data and ISAR imaging demonstrate that the Block EM-VB and its fast version have good BSS reconstruction performance and noise tolerance capability, especially in the low SNR scenarios, which implies that the proposed algorithms can be potentially applied in various signal processing fields.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

There is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[3] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[4] Lu Wang, Lifan Zhao, Guoan Bi, Chunru Wan, and Lei Yang, "Enhanced ISAR imaging by exploiting the continuity of the target scene," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5736–5750, 2014.

[5] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 275–285, 2008.

[6] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[7] B. X. Huang and T. Zhou, "Recovery of block sparse signals by a block version of StOMP," *Signal Processing*, vol. 106, pp. 231–244, 2015.

[8] Z. Zhang and B. D. Rao, "Recovery of block sparse signals using the framework of block sparse Bayesian learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3345–3348, Kyoto, Japan, 2012.

[9] L. Yu, J.-P. Barbot, G. Zheng, and H. Sun, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 1, pp. 259–269, 2012.

[10] L. Yu, H. Sun, G. Zheng, and J. Pierre Barbot, "Model based Bayesian compressive sensing via local beta process," *Signal Processing*, vol. 108, pp. 259–271, 2015.

[11] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 360–372, 2015.

[12] S. Zhang, Y. Liu, X. Li, and G. Bi, "Variational Bayesian sparse signal recovery with LSM prior," *IEEE Access*, vol. 5, pp. 26690–26702, 2017.

[13] A. Winkelbauer, "Moments and absolute moments of the normal distribution," 2012, https://arxiv.org/abs/1209.4340.

[14] H. Duan, L. Zhang, J. Fang, L. Huang, and H. Li, "Pattern-coupled sparse Bayesian learning for inverse synthetic aperture radar imaging," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1995–1999, 2015.

WILEY | Hindawi

Research Article

Joint Processing of DOA Estimation and Signal Separation for Planar Array Using Fast-PARAFAC Decomposition

Zhongyuan Que [ID], Benzhou Jin [ID], and Jianfeng Li [ID]

College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Correspondence should be addressed to Zhongyuan Que; quezhongyuan@nuaa.edu.cn

A joint processing of direction of arrival (DOA) and signal separation for planar array is proposed in this paper. Through sensor array processing theory, the output data of a planar array can be reconstructed as a parallel factor (PARAFAC) model, which can be decomposed with the trilinear alternating least square (TALS) algorithm. Aiming at the problem of slow speed on convergence for the standard PARAFAC method, we introduce the propagator method (PM) to accelerate the convergence of the TALS method and propose a novel method to jointly separate signals and estimate the corresponding DOAs. Given the initial angle estimates with PM, the number of iterations of TALS can be reduced considerably. The experiments indicate that our method can carry out signal separation and DOA estimation for typical modulated signals well and remain the same performance as the standard PARAFAC method with lower computational complexity, which verifies that our algorithm is effective.

1. Introduction

Signal separation and direction of arrival (DOA) estimation are significant themes in signal processing and have been investigated in various engineering fields including wireless communication, navigation, radar, and sonar [1–5]. As fundamental issues for signal processing, they have sparked considerable attention of researchers for decades. These two problems involve multiple signals and sensors which receive a mixture of signals [6]. The goal of DOA estimation is to find the source signal location, while the signal separation is aimed at extracting desired source signals. Through the years, many classical methods have been developed to solve these problems. For DOA estimation, subspace-based methods like MUSIC and ESPRIT have been widely adopted [7, 8]. The conventional nonparametric Fourier-based methods have also been further developed [9], and the emerging sparse reconstruction-based methods like orthogonal matching pursuit (OMP) and sparse Bayesian inference (SBI) are introduced into DOA estimation [10, 11]. For signal separation, the researches focus primarily on blind source separation (BSS) methods, where independent component analysis (ICA) and

Joint Approximative Diagonalization of Eigen matrix (JADE) method are the most famous among these methods [12, 13] and have been widely applied in the separation of speech and medical signals.

Compared with conventional DOA methods, BSS methods do not require much waveform prior information and are capable of identifying the transmission parameters based on the mixture signals, which has aroused an amount of attention of researchers. Many researchers study to apply blind separation algorithms into array signal model and have made lots of works. A combined complex blind source separation DOA estimation and signal recovery method was proposed for uniform linear array (ULA) in [14], which obtains better performance by exploiting BSS to estimate the array manifold. In [15], a blind DOA estimation method based on the JADE algorithm was proposed, which introduces fourth-order cumulant and has great performance in multipath environment. In [16], the chaotic adaptive firework algorithm was applied for solving the problem of radar emitter mixed signal. In [17], a new EM-based method for broadband DOA estimation and BSS was proposed, which reduces the complexity of traditional methods. In [18], a method based

on eigenvalue decomposition for DOA estimation and blind separation of narrow-band independent signals was presented. The above studies were discussed for ULA geometry and did not involve more complex array structures like planar arrays which are more practical in actual applications.

In recent years, tensor technique has taken off in the field of data analysis and signal processing [19], in which trilinear decomposition or the parallel factor (PARAFAC) technique has been extensively investigated in radar and wireless communication fields especially [20–24]. PARAFAC is a common model for low-rank decomposition of a tensor, whose computation can be completed by alternating least squares (ALS). Many models in array signal processing can be represented as trilinear models, which enables us to utilize the trilinear alternating least square (TALS) algorithm to achieve parameter estimation and signal separation. The authors in [25, 26] studied multiparameter estimation in bistatic multiple-input multiple-output (MIMO) radar and proposed a joint direction of departure (DOD) and direction of arrival (DOA) estimation using the PARAFAC model. In [27], the authors proposed a novel 2D-DOA estimation for trilinear decomposition-based monostatic cross MIMO radar. In [28], a joint DOA and carrier frequency estimation of narrow-band sources was proposed using the unitary PARAFAC method. These methods based on TALS can separate source signals and obtain automatically paired parameters without spectral peak search, but have relatively high computational complexity. It can be seen that PARAFAC has a great potential in DOA estimation and signal separation, but its shortcoming is the standard PARAFAC method has slow speed on convergence.

Motivated by the works mentioned above, in this paper, under the basic framework of PARAFAC, we propose a method of joint two-dimensional DOA estimation and signal separation for planar arrays. We first model the output of a planar array as the PARAFAC model and then utilize the propagator method (PM) to initialize the updated matrices in the TALS method, which effectively simplifies the complexity of the algorithm. Next, perform the TALS algorithm until convergence. Finally, acquire the 2D-DOA estimates and separated signals from the direction matrices and the source matrix which is estimated by TALS. The proposed method can achieve signal separation and DOA estimation for typical modulated signals well and remains the same performance as the standard PARAFAC method but with lower complexity. The experimental results verify the effectiveness of our algorithm.

We briefly summarize our main contributions as follows:

(1) We model the output of the uniform rectangular array and reconstruct it into the PARAFAC model

(2) We propose the fast-PARAFAC decomposition method for joint 2D-DOA estimation and signal separation, which utilizes PM to initialize the updated matrices in TALS and accelerate convergence

(3) The proposed method has better performance of DOA estimation than 2D-PM and 2D-ESPRIT and can accurately separate the source signals with lower

complexity compared with the standard PARAFAC approach

(4) The proposed method can obtain separated signals and corresponding DOA estimates without an additional pairing procedure

The outline of this paper is given as follows. We discuss the data model for uniform planar array and introduce the PARAFAC model briefly in Section 2. In Section 3, the proposed algorithm is described in detail. In Section 4, the complexity analysis and advantages of the proposed method are provided. The results of numerical simulations are given in Section 5, and conclusions are drawn in Section 6.

1.1. Notation. Lower-case and upper-case boldface letters denote vectors and matrices. $\mathbb{C}$ denotes the sets of complex numbers. The superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ represent the transpose, complex conjugate, and conjugate transpose of a vector or matrix, respectively. diag $(\cdot)$ denotes a diagonal matrix that consists of the elements of the matrix. $\otimes$ denotes the Kronecker product. $D_m(\cdot)$ denotes a diagonal matrix whose diagonal elements are defined with the m-th row of the matrix. angle$(\cdot)$ denotes phase angle operator. $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the ℓ_2 and *Frobenius* norms. $(\cdot)^{-1}$ and $(\cdot)^+$ stand for the inverse and pseudo-inverse of a matrix.

2. Data Model

Consider a uniform rectangular array (URA) containing $N \times M$ sensors as depicted in Figure 1, where N and M are the numbers of elements along the x-axis and y-axis. The interelement spacings along both the x-axis and y-axis of the array are taken as half the wavelength of the waves, $d_x = d_y = \lambda/2$.

Assume that K uncorrelated far-field signals individually impinge on the array from $\{(\theta_k, \phi_k)|k = 1, 2, \cdots, K\}$, where θ_k and ϕ_k are the corresponding elevation and azimuth angles of the k-th signal ($K < N \times M$, $\theta_k \in (0, 90°)$, and $\phi_k \in (0, 180°)$). The output of the rectangular array can be represented as follows [29]:

$$\tilde{\mathbf{X}} = \mathbf{AS} + \mathbf{N}, \tag{1}$$

where $\tilde{\mathbf{X}} \in \mathbb{C}^{NM \times L}$ is the output data with noise; L denotes the number of snapshots; $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K]^T \in \mathbb{C}^{K \times L}$ is the signal matrix of L snapshots; $\mathbf{N} \in \mathbb{C}^{NM \times L}$ is the additive white Gaussian noise matrix. The array manifold matrix $\mathbf{A} \in \mathbb{C}^{NM \times K}$ consists of the steering vectors and is given by [29]

$$\mathbf{A} = \left[\mathbf{a}_y(v_1) \otimes \mathbf{a}_x(u_1), \mathbf{a}_y(v_2) \otimes \mathbf{a}_x(u_2), \cdots, \mathbf{a}_y(v_K) \otimes \mathbf{a}_x(u_K)\right], \tag{2}$$

where $u_k = \sin \theta_k \cos \phi_k$ and $v_k = \sin \theta_k \sin \phi_k$; $\mathbf{a}_x(u_k)$ and $\mathbf{a}_y(v_k)$ are the steering vectors of the array, which can be represented as [30]
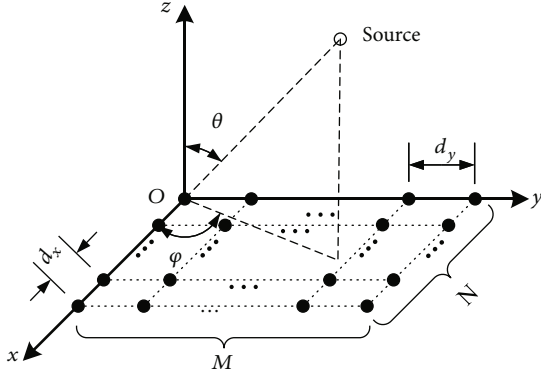
FIGURE 1: Geometry of URA with $N \times M$ sensors. The circle represents a signal source at far field, and the solid points represent the sensors.

$$\mathbf{a}_x(u_k) = \left[1, \exp\left(\frac{-j2\pi d_x u_k}{\lambda}\right), \cdots, \exp\left(\frac{-j2\pi(N-1)d_x u_k}{\lambda}\right)\right]^T,$$

$$\mathbf{a}_y(v_k) = \left[1, \exp\left(\frac{-j2\pi d_y v_k}{\lambda}\right), \cdots, \exp\left(\frac{-j2\pi(M-1)d_y v_k}{\lambda}\right)\right]^T. \tag{3}$$

More compactly, (2) can be also written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_x D_1(\mathbf{A}_y) \\ \mathbf{A}_x D_2(\mathbf{A}_y) \\ \vdots \\ \mathbf{A}_x D_M(\mathbf{A}_y) \end{bmatrix}, \tag{4}$$

where $\mathbf{A}_x = [\mathbf{a}_x(u_1), \mathbf{a}_x(u_2), \cdots, \mathbf{a}_x(u_K)]$ and $\mathbf{A}_y = [\mathbf{a}_y(v_1), \mathbf{a}_y(v_2), \cdots, \mathbf{a}_y(v_K)]$.

Before further describing the data model, we need to introduce the definition of the tensor outer product and the parallel factor (PARAFAC) model [22].

Definition 1 (Outer product [19]). The outer product of three vectors, $\mathbf{a} \in \mathbb{C}^{M \times 1}$, $\mathbf{b} \in \mathbb{C}^{N \times 1}$, and $\mathbf{c} \in \mathbb{C}^{L \times 1}$, denoted by $(\mathbf{a} \circ \mathbf{b} \circ \mathbf{c})$, is a $M \times N \times L$ tensor whose elements are defined by $(\mathbf{a} \circ \mathbf{b} \circ \mathbf{c})_{m,n,l} = a_m b_n c_l$.

Definition 2 (PARAFAC [19]). The PARAFAC model is also known as the trilinear decomposition model. A canonical PARAFAC decomposition of a three-order tensor $\mathbf{X} \in \mathbb{C}^{M \times N \times L}$ can be expressed as

$$\mathbf{X} = \sum_{f=1}^{F} \mathbf{a}_f \circ \mathbf{b}_f \circ \mathbf{c}_f, \tag{5}$$

where $\mathbf{a}_f$, $\mathbf{b}_f$, and $\mathbf{c}_f$ stand for the f-th columns of matrices $\mathbf{A} \in \mathbb{C}^{M \times F}$, $\mathbf{B} \in \mathbb{C}^{N \times F}$, and $\mathbf{C} \in \mathbb{C}^{L \times F}$. For a 3-way tensor $\mathbf{X}$, define its sliced matrices $\mathbf{X}_m \in \mathbb{C}^{N \times L}$, $\mathbf{X}_n \in \mathbb{C}^{L \times M}$, and $\mathbf{X}_l \in$

$\mathbb{C}^{M \times N}$ with the element $\mathbf{X}_{m(n,l)} = \mathbf{X}_{n(l,m)} = \mathbf{X}_{l(m,n)} = \mathbf{X}_{m,n,l}$. Then, $\mathbf{X}_m = \mathbf{B}D_m(\mathbf{A})\mathbf{C}^T$ can be viewed as "slicing" the 3-D array in a series of "slabs" (2-D arrays) and similarly for others [21].

Based on the PARAFAC model, the noiseless received data for URA can be written as [22]

$$\mathbf{X}_m = \mathbf{A}_x D_m(\mathbf{A}_y)\mathbf{S}. \tag{6}$$

Due to the symmetry of the PARAFAC model, the other two slice matrices can be obtained.

$$\mathbf{Y}_n = \mathbf{S}^T D_n(\mathbf{A}_x)\mathbf{A}_y^T, \tag{7}$$

$$\mathbf{Z}_l = \mathbf{A}_y D_l(\mathbf{S}^T)\mathbf{A}_x^T. \tag{8}$$

Define $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ as the results of the concatenation of matrices $\mathbf{X}_m$, $\mathbf{Y}_n$, and $\mathbf{Z}_l$, respectively, and then, the noise-free received signal matrices $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ can be represented as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_M \end{bmatrix} = \begin{bmatrix} \mathbf{A}_x D_1(\mathbf{A}_y) \\ \mathbf{A}_x D_2(\mathbf{A}_y) \\ \vdots \\ \mathbf{A}_x D_M(\mathbf{A}_y) \end{bmatrix} \mathbf{S},$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{S}^T D_1(\mathbf{A}_x) \\ \mathbf{S}^T D_2(\mathbf{A}_x) \\ \vdots \\ \mathbf{S}^T D_N(\mathbf{A}_x) \end{bmatrix} \mathbf{A}_y^T, \tag{9}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_L \end{bmatrix} = \begin{bmatrix} \mathbf{A}_y D_1(\mathbf{S}^T) \\ \mathbf{A}_y D_2(\mathbf{S}^T) \\ \vdots \\ \mathbf{A}_y D_L(\mathbf{S}^T) \end{bmatrix} \mathbf{A}_x^T.$$

Note that in this paper, we assume that there is no mutual coupling across the sensors. In fact, mutual coupling will degrade the performance of the algorithms. The recent researches in the presence of mutual coupling can be found in [31].

3. The Proposed Algorithm

We show how to perform signal separation and DOA estimation using our proposed algorithm in this section. The standard PARAFAC suffers from expensive computation cost due to slow convergence. To handle this problem, we introduce the propagator method (PM) to accelerate TALS by providing the initial angle estimates. Then, alternately update the LS estimates of $\mathbf{S}$, $\mathbf{A}_y$, and $\mathbf{A}_x$ until they converge.

Finally, obtain the separated signals and the corresponding DOA estimates.

Note that in practice, we need to estimate the number of sources from the received signal first. In this study, we assume that the number of sources is known in advance.

3.1. Initialization with Propagator Method.
By exploiting the property of rotational invariance, the propagator method can achieve the angle estimation with relatively low complexity [32, 33].

First, compute the data covariance matrix $\hat{\mathbf{R}}$ using the received signal data in (1) and partition it as follows [33]:

$$\hat{\mathbf{R}} = \left[\hat{\mathbf{G}}, \hat{\mathbf{H}}\right], \tag{10}$$

where $\hat{\mathbf{G}} \in \mathbb{C}^{MN \times K}$ is the first column to the K-th column of $\hat{\mathbf{R}}$ and $\hat{\mathbf{H}} \in \mathbb{C}^{MN \times (MN-K)}$ stands for the remaining columns.

Then, we can estimate the propagator $\mathbf{P}$ by

$$\hat{\mathbf{P}} = \left(\mathbf{G}\wedge^{H}\mathbf{G}\right)^{-1}\mathbf{G}\wedge^{H}\mathbf{H}, \tag{11}$$

Define [33]

$$\hat{\mathbf{P}}_c = \begin{bmatrix} \mathbf{I}_K \\ \mathbf{P}\wedge^H \end{bmatrix} = \begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_x\mathbf{\Phi}_y \\ \vdots \\ \mathbf{A}_x\mathbf{\Phi}_y^{M-1} \end{bmatrix} \mathbf{A}_x^{-1}, \tag{12}$$

where $\mathbf{\Phi}_y = \text{diag}\left\{\exp\left(-j2\pi d_y v_1/\lambda\right), \cdots, \exp\left(-j2\pi d_y v_K/\lambda\right)\right\}$ and $\mathbf{I}_K$ denotes a K-order identity matrix.

The estimates $\hat{v}_{k0}$ of v_k can be obtained by partitioning the matrix $\hat{\mathbf{P}}_c$ and eigenvalue decomposition.

After reconstructing the matrix $\hat{\mathbf{P}}_c$, another matrix $\hat{\mathbf{P}}_{cs}$ can be obtained by

$$\hat{\mathbf{P}}_{cs} = \begin{bmatrix} \mathbf{A}_y \\ \mathbf{A}_y\mathbf{\Phi}_x \\ \vdots \\ \mathbf{A}_y\mathbf{\Phi}_x^{N-1} \end{bmatrix} \mathbf{A}_x^{-1}, \tag{13}$$

where $\mathbf{\Phi}_x = \text{diag}\left\{\exp\left(-j2\pi d_x u_1/\lambda\right), \cdots, \exp\left(-j2\pi d_x u_K/\lambda\right)\right\}$. The estimates $\hat{u}_{k0}$ of u_k can also be obtained by a similar method.

3.2. Trilinear Alternating Least Square.
Trilinear alternating least square (TALS) is the most common method for trilinear model decomposition [21, 22]. The standard TALS algorithm utilizes random matrices as the initial load matrices, which usually converges slowly. In this part, the initial estimates $\hat{u}_{k0}$ and $\hat{v}_{k0}$ provided by PM are used to construct the matrices $\mathbf{A}\wedge_x^{(0)}$ and $\mathbf{A}\wedge_y^{(0)}$ as initial matrices.

Recall that we assume noise is additive Gaussian noise, and it is reasonable to employ the least square principle to estimate $\mathbf{S}$, $\mathbf{A}_x$, and $\mathbf{A}_y$. The estimation of the matrix $\mathbf{S}$ can be conducted by minimizing the following quadratic cost function [21]:

$$\min_{\mathbf{S}\wedge^{(n)}} \left\| \begin{bmatrix} \tilde{\mathbf{X}}_1 \\ \tilde{\mathbf{X}}_2 \\ \vdots \\ \tilde{\mathbf{X}}_M \end{bmatrix} - \begin{bmatrix} \mathbf{A}\wedge_x^{(n-1)}D_1\left(\mathbf{A}\wedge_y^{(n-1)}\right) \\ \mathbf{A}\wedge_x^{(n-1)}D_2\left(\mathbf{A}\wedge_y^{(n-1)}\right) \\ \vdots \\ \mathbf{A}\wedge_x^{(n-1)}D_M\left(\mathbf{A}\wedge_y^{(n-1)}\right) \end{bmatrix} \mathbf{S}\wedge^{(n)} \right\|_F, \tag{14}$$

where $\tilde{\mathbf{X}}_m$ denotes the data matrix $\mathbf{X}_m$ with noise, $m = 1, 2, \cdots, M$; $\mathbf{A}\wedge_x^{(n-1)}$ and $\mathbf{A}\wedge_y^{(n-1)}$ denote the estimates of $\mathbf{A}_x$ and $\mathbf{A}_y$ obtained from $(n-1)$-th iteration.

Then, the LS estimate of $\mathbf{S}$ can be obtained as [21]

$$\mathbf{S}\wedge^{(n)} = \begin{bmatrix} \mathbf{A}\wedge_x^{(n-1)}D_1\left(\mathbf{A}\wedge_y^{(n-1)}\right) \\ \mathbf{A}\wedge_x^{(n-1)}D_2\left(\mathbf{A}\wedge_y^{(n-1)}\right) \\ \vdots \\ \mathbf{A}\wedge_x^{(n-1)}D_M\left(\mathbf{A}\wedge_y^{(n-1)}\right) \end{bmatrix}^{+} \begin{bmatrix} \tilde{\mathbf{X}}_1 \\ \tilde{\mathbf{X}}_2 \\ \vdots \\ \tilde{\mathbf{X}}_M \end{bmatrix}. \tag{15}$$

The LS fitting for $\mathbf{A}_y$ is similar to $\mathbf{S}$.

$$\min_{\mathbf{A}\wedge_y^{(n)}} \left\| \begin{bmatrix} \tilde{\mathbf{Y}}_1 \\ \tilde{\mathbf{Y}}_2 \\ \vdots \\ \tilde{\mathbf{Y}}_N \end{bmatrix} - \begin{bmatrix} \mathbf{S}\wedge^{T(n)}D_1\left(\mathbf{A}\wedge_x^{(n-1)}\right) \\ \mathbf{S}\wedge^{T(n)}D_2\left(\mathbf{A}\wedge_x^{(n-1)}\right) \\ \vdots \\ \mathbf{S}\wedge^{T(n)}D_N\left(\mathbf{A}\wedge_x^{(n-1)}\right) \end{bmatrix} \mathbf{A}\wedge^{T(n)}_y \right\|_F, \tag{16}$$

where $\tilde{\mathbf{Y}}_n$ denotes the data matrix $\mathbf{Y}_n$ with noise, $n = 1, 2, \cdots, N$; $\mathbf{S}\wedge^{(n)}$ denotes the estimate of $\mathbf{S}$ according to (15).

Then, the LS estimate of $\mathbf{A}_y$ can be represented as

$$\hat{\mathbf{A}}_y^{T(n)} = \begin{bmatrix} \mathbf{S}\wedge^{T(n)}D_1\left(\mathbf{A}\wedge_x^{(n-1)}\right) \\ \mathbf{S}\wedge^{T(n)}D_2\left(\mathbf{A}\wedge_x^{(n-1)}\right) \\ \vdots \\ \mathbf{S}\wedge^{T(n)}D_N\left(\mathbf{A}\wedge_x^{(n-1)}\right) \end{bmatrix}^{+} \begin{bmatrix} \tilde{\mathbf{Y}}_1 \\ \tilde{\mathbf{Y}}_2 \\ \vdots \\ \tilde{\mathbf{Y}}_N \end{bmatrix}. \tag{17}$$

Similarly, the LS fitting for $\mathbf{A}_x$ is

$$\min_{\mathbf{A}\wedge_x^{(n)}} \left\| \begin{bmatrix} \tilde{\mathbf{Z}}_1 \\ \tilde{\mathbf{Z}}_2 \\ \vdots \\ \tilde{\mathbf{Z}}_L \end{bmatrix} - \begin{bmatrix} \mathbf{A}\wedge_y^{(n)} D_1\left(\mathbf{S}\wedge^{T(n)}\right) \\ \mathbf{A}\wedge_y^{(n)} D_2\left(\mathbf{S}\wedge^{T(n)}\right) \\ \vdots \\ \mathbf{A}\wedge_y^{(n)} D_L\left(\mathbf{S}\wedge^{T(n)}\right) \end{bmatrix} \mathbf{A}\wedge_x^{T(n)} \right\|_F , \quad (18)$$

where $\tilde{\mathbf{Z}}_l$ denotes the data matrix $\mathbf{Z}_l$ with noise, $l = 1, 2, \cdots, L$, and $\mathbf{A}\wedge_y^{(n)}$ is the estimate of $\hat{\mathbf{A}}_y$ according to (17).

The estimate of $\mathbf{A}_x$ can be expressed as

$$\mathbf{A}\wedge_x^{T(n)} = \begin{bmatrix} \mathbf{A}\wedge_y^{(n)} D_1\left(\mathbf{S}\wedge^{T(n)}\right) \\ \mathbf{A}\wedge_y^{(n)} D_2\left(\mathbf{S}\wedge^{T(n)}\right) \\ \vdots \\ \mathbf{A}\wedge_y^{(n)} D_L\left(\mathbf{S}\wedge^{T(n)}\right) \end{bmatrix}^+ \begin{bmatrix} \tilde{\mathbf{Z}}_1 \\ \tilde{\mathbf{Z}}_2 \\ \vdots \\ \tilde{\mathbf{Z}}_L \end{bmatrix} . \quad (19)$$

According to (15), (17), and (19), we can repeatedly update the estimates of $\mathbf{S}$, $\mathbf{A}_y$, and $\mathbf{A}_x$ until convergence. Because of the utilization of PM, the proposed algorithm fast converges to the final estimates of $\mathbf{S}$, $\mathbf{A}_x$, and $\mathbf{A}_y$, noted as $\hat{\mathbf{S}}_f$, $\hat{\mathbf{A}}_{fx}$, and $\hat{\mathbf{A}}_{fy}$. At this point, the task of signal separation is complete. The last part is to perform DOA estimation.

It is worth noting that the TALS algorithm outlined above contains only the simplest steps. Some techniques, like line search [34, 35], can be coupling with the basic TALS algorithm, which may improve the rate of convergence further. There is no universally accepted most efficient TALS algorithm for all of the problems. We use the basic implementation of TALS for our issues and compare the complexity of our method with line search schemes [34, 35] in Section 5.

3.3. DOA Estimation. First, we normalize the column vectors of $\hat{\mathbf{A}}_{fx}$ and $\hat{\mathbf{A}}_{fy}$ and make the first element of the column to equal one. Then, compute the phase vector $\mathbf{r}_x$ by

$$\mathbf{r}_x = -\text{angle}(\mathbf{a}_{xk}) = \left[0, \frac{2\pi d_x}{\lambda}, \cdots, \frac{2\pi(N-1)d_x}{\lambda} \right]^T u_k, = \mathbf{B}_x u_k, \quad (20)$$

where $\mathbf{a}_{xk}$ denotes the k-th column vector of $\hat{\mathbf{A}}_{fx}$ after normalization.

According to LS criterion, calculate the estimates of u_k by

$$\hat{u}_k = \mathbf{B}_x^+ \mathbf{r}_x. \quad (21)$$

In a similar way, we can also get the estimates $\hat{v}_k$ of v_k by the following expressions:

$$\hat{v}_k = \mathbf{B}_y^+ \mathbf{r}_y, \quad (22)$$

where $\mathbf{r}_y$ is another phase vector defined as

$$\mathbf{r}_y = -\text{angle}(\mathbf{a}_{yk}) = \left[0, \frac{2\pi d_y}{\lambda}, \cdots, \frac{2\pi(M-1)d_y}{\lambda} \right]^T v_k, = \mathbf{B}_y v_k, \quad (23)$$

where $\mathbf{a}_{yk}$ denotes the k-th column vector of $\hat{\mathbf{A}}_{fy}$ after normalization.

Finally, the estimates of θ_k and ϕ_k can be calculated by

$$\hat{\theta}_k = \arcsin\left(|\hat{u}_k + j\hat{v}_k|\right), \quad (24)$$

$$\hat{\phi}_k = \text{angle}(\hat{u}_k + j\hat{v}_k), \quad (25)$$

where $|\cdot|$ denotes the modulus of the complex number and $\hat{\theta}_k$ and $\hat{\phi}_k$ are the estimates of the elevation and azimuth angles of the k-th signal.

Note that there are the same permutation effects for the estimation of $\hat{\mathbf{S}}_f$, $\hat{\mathbf{A}}_{fx}$, and $\hat{\mathbf{A}}_{fy}$ during the TALS decomposition, so the final estimates, $\hat{\theta}_k$ and $\hat{\phi}_k$, are automatically paired.

3.4. The Procedure of the Proposed Algorithm. We summarize the major steps of our algorithm as follows:

Step 1. Exploit the propagator method to calculate the initial estimates $\hat{u}_{k0}$ and $\hat{v}_{k0}$ of u_k, v_k.

Step 2. According to the PARAFAC models (6)–(8), reshape the received signal data to acquire the data matrices $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$, and $\tilde{\mathbf{Z}}$.

Step 3. Construct the direction matrices $\hat{\mathbf{A}}_x$ and $\hat{\mathbf{A}}_y$ with $\hat{u}_{k0}$ and $\hat{v}_{k0}$ and use them as initial matrices.

Step 4. According to (15), (17), and (19), update the estimates of $\mathbf{S}$, $\mathbf{A}_y$, and $\mathbf{A}_x$ alternately from the data matrices $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$, and $\tilde{\mathbf{Z}}$ until convergence.

Step 5. According to (20)-(25), calculate the DOA estimates of separated signals, $\hat{\theta}_k$ and $\hat{\phi}_k$.

4. Performance Analysis

4.1. Complexity Analysis. Since complex multiplication requires the most computation time and resources, we use the time of complex multiplication to evaluate the complexity of the algorithm. The algorithm proposed in this paper adopts PM for initial estimation, whose complexity is $O(5K^3 + 2K^2L + 3K^2N(M-1) + 3K^2M(N-1) + (NM-K)KL)$. The complexity of the TALS method is related to the number of iterations and the complexity of a single iteration, and the complexity of each iteration is easily obtained as $O(3K^3 + 3NMKL + 2K^2(NM + NL + ML))$. The number of iterations is affected by many factors such as array
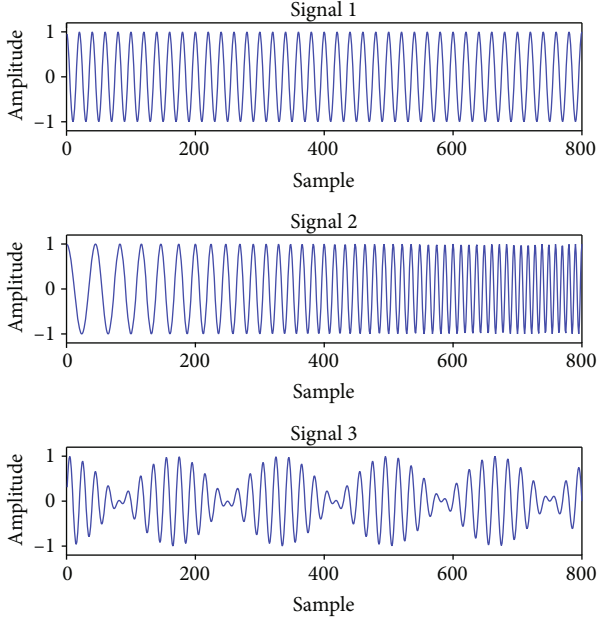
FIGURE 2: The source signal waveform. ($L = 800$).

FIGURE 3: Analysis of mean CPU time and mean number of iterations versus SNR.

size, iteration accuracy, and signal type. Define the number of iterations is T, and the total computational complexity is $O(T$ $(3K^3 + 3NMKL + 2K^2(NM + NL + ML)))$. Therefore, the complexity of the proposed method is $O(5K^3 + 2K^2L + 3K^2N$ $(M - 1) + 3K^2M(N - 1) + (NM - K)KL + T(3K^3 + 3NMKL$ $+ 2K^2(NM + NL + ML))$. Due to PM initialization, the iterations of the proposed method are greatly reduced compared with the standard PARAFAC method, which we can see in the next section.

4.2. Advantages. The advantages of the proposed algorithm are as follows:

(1) The proposed method has lower computational cost than the standard PARAFAC method due to introducing PM

(2) The proposed method outperforms 2D-ESPRIT and 2D-PM in the aspect of angle estimation performance for planar array

(3) The proposed method can obtain separated signals and corresponding DOA estimation without an additional pairing procedure

5. Simulation Results

In this section, we employ a URA equipped with 8×8 sensors to illustrate the improvement of the performance of 2D DOA estimation and signal separation of the proposed algorithm.

Suppose there are $K = 3$ typical modulated signals impinging on the array simultaneously, which are single-frequency signal $s_1(t) = \cos(2\pi \times 5 \times 10^6 t)$, linear frequency modulated signal $s_2(t) = \cos(\pi \times 10^{12}t^2 + 2\pi \times 2 \times 10^6 t)$,

FIGURE 4: Analysis of convergence rate.

and amplitude modulated signal $s_3(t) = \cos(2\pi \times 3 \times 10^5 t)$ $\sin(2\pi \times 5 \times 10^6 t)$. The DOAs of the signals are $(\theta_1, \phi_1) = (10°, 15°)(\theta_2, \phi_2) = (20°, 25°)$, and $(\theta_3, \phi_3) = (30°, 35°)$, and the sampling frequency is 100 MHz. The noiseless source signal waveforms are demonstrated in Figure 2.

To assess the performance of DOA estimation, root mean square error (RMSE) is used,

$$\text{RMSE} = \sqrt{\frac{1}{CK} \sum_{c=1}^{C} \sum_{k=1}^{K} (\alpha \wedge_{k,c} - \alpha_k)^2}, \qquad (26)$$

(a) Performance of elevation angle estimation

(b) Performance of azimuth angle estimation
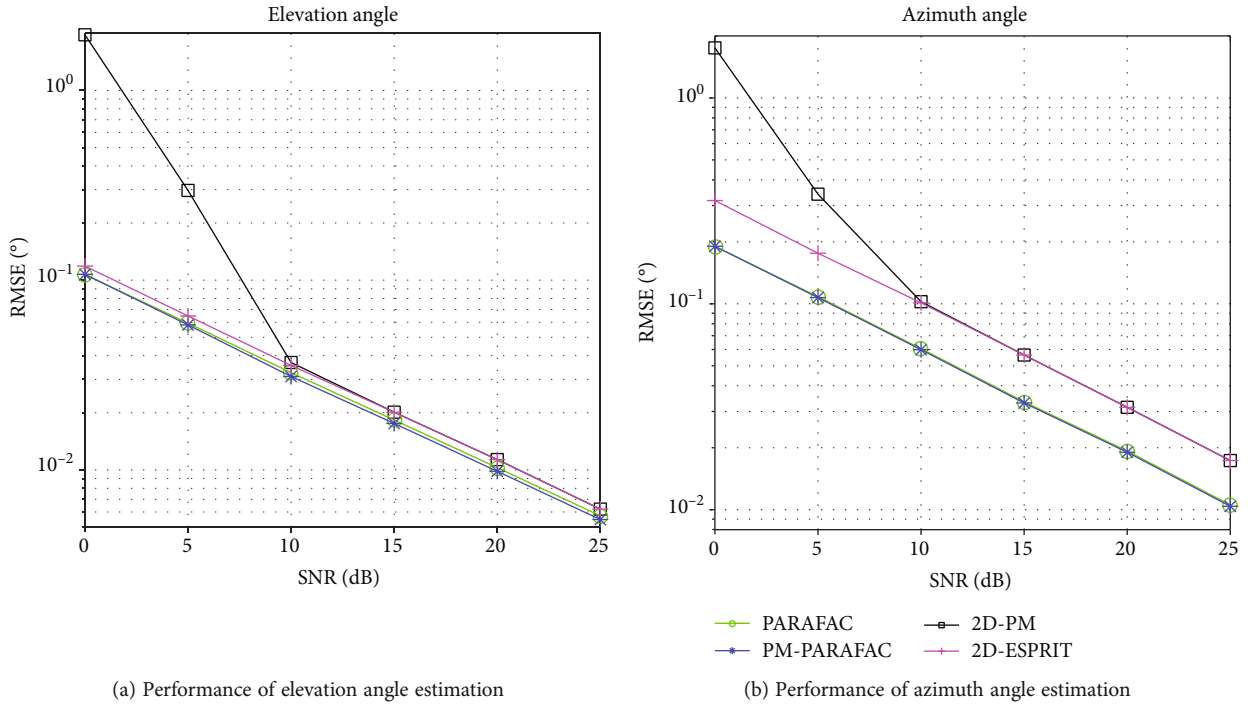
FIGURE 5: RMSE performance of different algorithms versus SNR ($L = 800$).
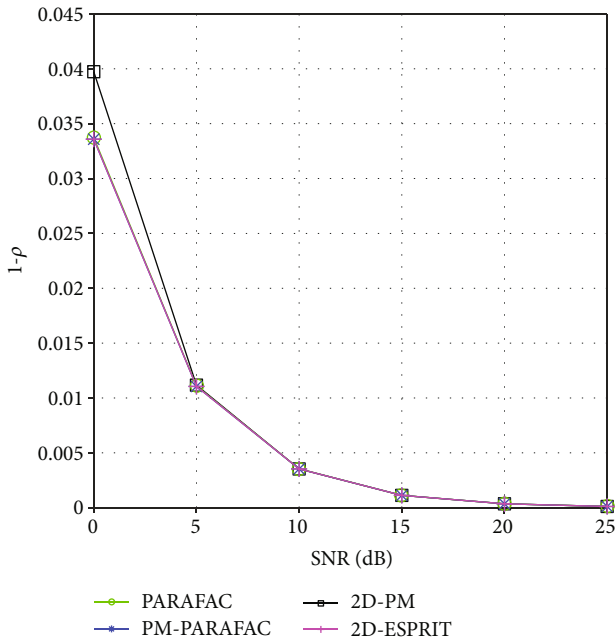
FIGURE 6: Performance of signal separation of different algorithms versus SNR ($L = 800$).
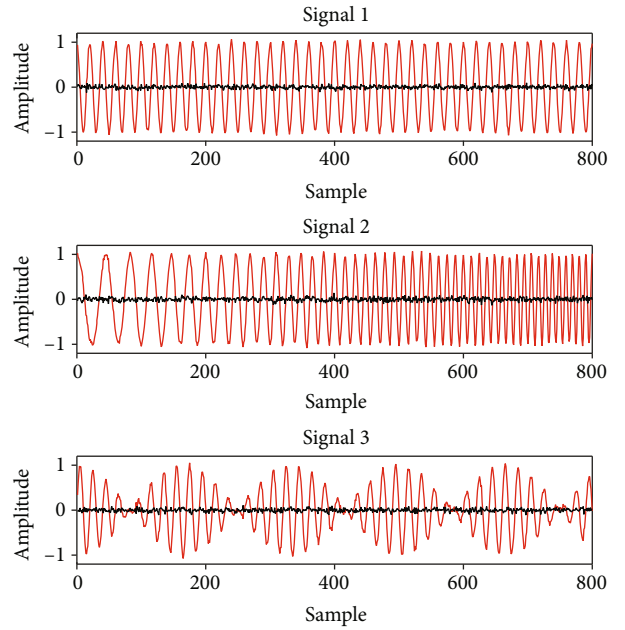
FIGURE 7: Separated signals by the proposed algorithm ($L = 800$, SNR = 5 dB). Red line denotes the separated signal, and black line denotes the estimation error.

where C is the total number of *Monte-Carlo* trials, α_k is the true value of the elevation or azimuth angle of k-th signal, and $\widehat{\alpha}_{k,c}$ is the estimate of the angle α_k in the c-th trial. For each simulation, we set $C = 1000$. The signal to noise ratio is defined by SNR $= 10 \log_{10}(\|\mathbf{X}\|_F^2/\|\mathbf{N}\|_F^2)$, where $\mathbf{X}$ is noise-

less received data matrix and $\mathbf{N}$ is zero-mean white Gaussian noise matrix.

Besides, to qualify the performance of signal separation, the average similar coefficient between the source signal and the separated signal is adopted,

(a) Performance of elevation angle estimation

(b) Performance of azimuth angle estimation
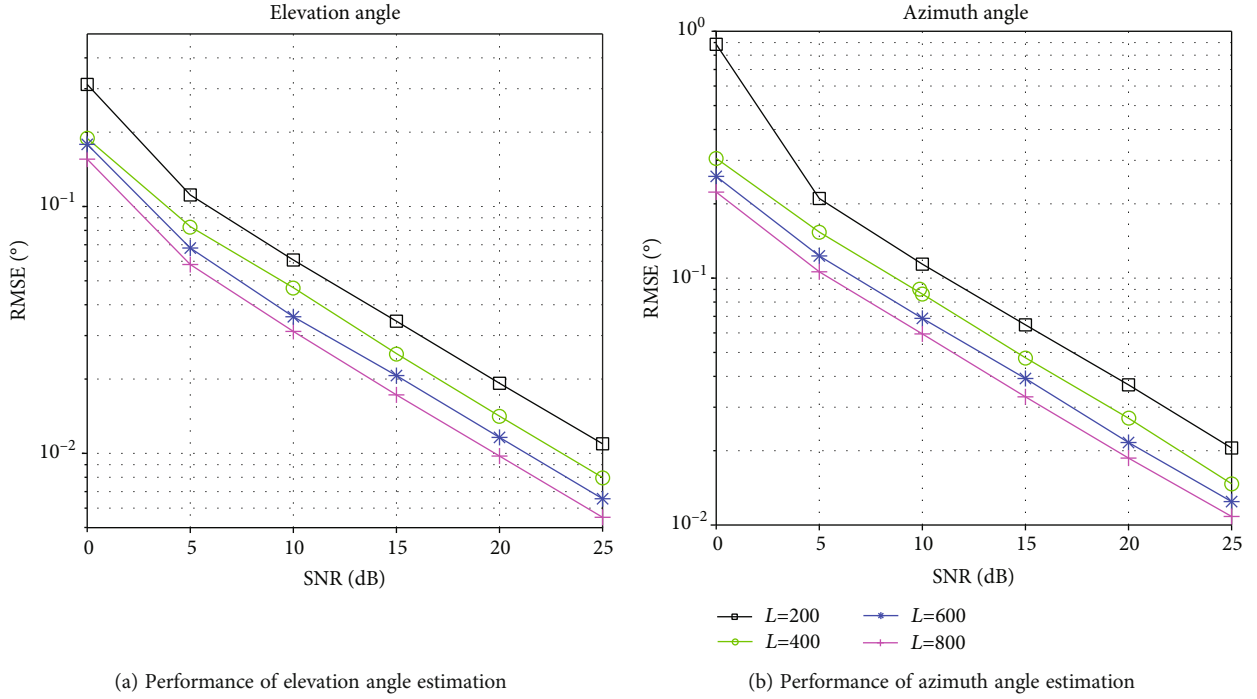
FIGURE 8: RMSE performance with different values of L versus SNR.

$$\rho = \frac{1}{CK} \sum_{c=1}^{C} \sum_{k=1}^{K} \frac{\left| \mathbf{s}_k \widehat{\mathbf{s}}_{k,c}^H \right|}{\sqrt{\left\| \mathbf{s}_k \right\|_2^2 \left\| \widehat{\mathbf{s}}_{k,c} \right\|_2^2}}, \tag{27}$$

where $\mathbf{s}_k$ is the k-th source signal and $\widehat{\mathbf{s}}_{k,c}$ is the corresponding estimate in the c-th trial.

5.1. Convergence Analysis. In Figure 3, we give the result of the mean CPU time and mean number of iterations based on our algorithm and other PARAFAC algorithm, where $L = 800$. PARAFAC and PM-PARAFAC denote the standard PARAFAC without modification and our algorithm, respectively. Besides, we also compare our method with line search schemes as "LS-PARAFAC" [34] and "ELSCS-PARAFAC" [35]. Owing to the initialization with PM, the mean CPU time and the mean number of iterations required are reduced considerably. The standard PARAFAC requires 153.2 iterations on average, while the proposed algorithm requires 10.8 iterations, which means that the proposed algorithm is ten times faster than the standard PARAFAC. It can also be seen that LS-PARAFAC and ELSCS-PARAFAC are faster than the standard PARAFAC but slower than our algorithm.

Define the sum of squared residuals $\text{SSR} = \sum_{n=1}^{N} \sum_{m=1}^{M}$ $\sum_{l=1}^{L} [\tilde{\mathbf{X}}_{n,m,l} - \sum_{k=1}^{K} \mathbf{A} \wedge_x (n,k) \mathbf{A} \wedge_y (m,k) \mathbf{S} \wedge (k,l)]^2$ and $\text{DSSR} = |\text{SSR}_i - \text{SSR}_{i-1}|/\text{SSR}_{i-1}$, where SSR_i is the SSR after the i-th iteration. Figure 4 shows typical curves of the evolution of DSSR, where $\text{SNR} = 10$ and $L = 800$. We can also observe that the proposed method has faster convergence in Figure 4.

As mentioned in Section 4, the complexity of the TALS method is related to many factors, like the scale of the array, the signal-to-noise ratio, and the types of signals. Although a

slight change in configures can cause a significant difference in the execution time, the proposed method can always reduce the computational cost compared with the standard TALS method.

5.2. Comparison of Performance. To verify the improvement of the proposed algorithm, the proposed method is compared with 2D-PM, 2D-ESPRIT, and the standard PARAFAC method. Note that the original 2D-PM and 2D-ESPRIT do not have the capability of signal separation, so we use their results of DOA estimation to construct the matrix $\mathbf{A}$ in (1) and compute the LS estimate of $\mathbf{S}$ by $\widehat{\mathbf{S}} = \mathbf{A}^+ \tilde{\mathbf{X}}$.

As shown in Figure 5, it is evident that our method has the same angle estimation performance as the standard PARAFAC method, which surpasses 2D-ESPRIT and 2D-PM.

Figure 6 shows the performance of signal separation of different algorithms with different SNR. From Figure 6, the signal separation performance of our algorithm, standard PARAFAC method, and 2D-ESPRIT are approximately the same with different SNR, while the 2D-PM algorithm has a slightly weaker signal separation performance with low SNR. Figure 7 is the separated signal diagram by the proposed algorithm. From Figure 7, it can be seen that the error between the source signal and the separated signal is very small, which is consistent with the high average similarity coefficient observed in Figure 6.

Figure 8 illustrates the DOA estimation performance as a function of SNR with different numbers of snapshots. It is seen from Figure 8 that when the number of snapshots L increases, angle estimation performance can be improved.
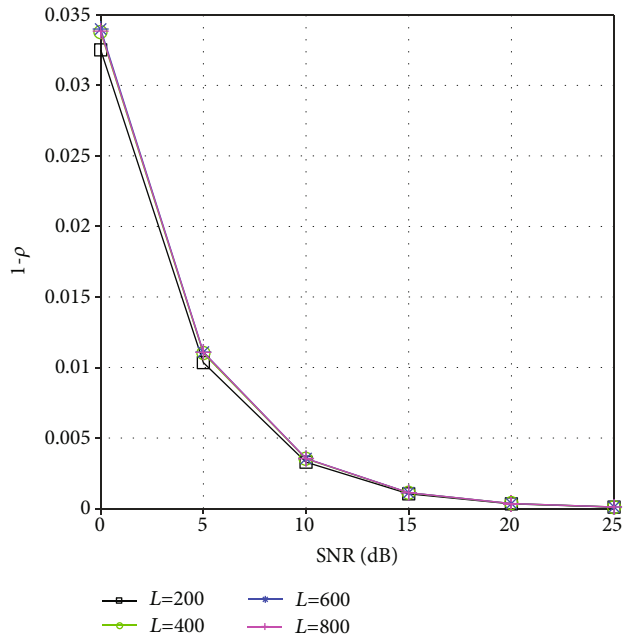
FIGURE 9: Performance of signal separation with different values of L versus SNR.

Figure 9 presents the performance of signal separation as a function of SNR with different numbers of snapshots. As seen from Figure 9, the average similar coefficients with different numbers of snapshots are close and increase with increasing of SNR.

6. Conclusions

In this paper, a joint processing of direction of arrival estimation and signal separation for planar array based on the fast-PARAFAC model is proposed. We model the output of planar array as the PARAFAC model and combine PM with the TALS method for DOA estimation and signal separation. The angle estimates by PM are used for the initialization of the TALS method. Then, the TALS method is used to separate the source signal and accurately estimate DOA. The proposed method not only inherits the advantages of the TALS method in signal separation but also takes advantage of the low complexity of the PM algorithm, which greatly reduces the total iterations of the standard PARAFAC algorithm. The results show that, as compared with the conventional DOA estimation approaches such as the 2D-PM and 2D-ESPRIT algorithm, the proposed method has better performance in signal separation and DOA estimation for planar array.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[2] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, 1998.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[4] X. Zhang, L. Xu, L. Xu, and D. Xu, "Direction of departure (DOD) and direction of arrival (DOA) estimation in MIMO radar with reduced-dimension MUSIC," *IEEE Communications Letters*, vol. 14, no. 12, pp. 1161–1163, 2010.

[5] F. Wen, J. Shi, and Z. Zhang, "Closed-form estimation algorithm for EMVS-MIMO radar with arbitrary sensor geometry," *Signal Processing*, vol. 186, pp. 108117–108119, 2021.

[6] A. S. Kayhan and M. G. Amin, "Spatial evolutionary spectrum for DOA estimation and blind signal separation," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 791–798, 2000.

[7] J. Li, D. Li, D. Jiang, and X. Zhang, "Extended-aperture unitary root MUSIC-based DOA estimation for coprime array," *IEEE Communications Letters*, vol. 22, no. 4, pp. 752–755, 2018.

[8] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, "An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 996–1011, 2014.

[9] J. Li, P. Ma, X. Zhang, and G. Zhao, "Improved DFT algorithm for 2D DOA estimation based on 1D nested array motion," *IEEE Communications Letters*, vol. 24, no. 9, pp. 1953–1956, 2020.

[10] K. Aghababaiyan, V. Shah-Mansouri, and B. Maham, "High-precision OMP-based direction of arrival estimation scheme for hybrid non-uniform array," *IEEE Communications Letters*, vol. 24, no. 2, pp. 354–357, 2020.

[11] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, 2013.

[12] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[13] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, 1993.

[14] C. Y. Kang, X. H. Zhang, and D. Han, "DOA estimation and signal recovery combined blind source separation with high resolution," *Acta Automatica Sinica*, vol. 36, no. 3, pp. 442–445, 2010.

[15] Z. Jia, J. Yang, and J. Jin, "DOA estimation based on JADE algorithm," *Journal on Communications*, vol. 31, no. 8, pp. 91–97, 2010.

[16] W. Luo, H. Jin, H. Li, and R. Hua, "Blind source separation of radar signals based on chaotic adaptive firework algorithm," *Systems Engineering and Electronics*, vol. 42, no. 11, pp. 2497–2505, 2020.

[17] K. Xiong, Z. Liu, Z. Liu, W. Jiang, and H. Wang, "Broadband DOA estimation and blind source separation based on EM algorithm," *Acta Electronica Sinica*, vol. 43, no. 10, pp. 2028–2033, 2015.

[18] K. Xiong and A. Yang, "Blind array signal separation and DOA estimation method based on eigenvalue decomposition," *Signal, Image and Video Processing*, vol. 15, no. 6, pp. 1107–1113, 2021.

[19] A. Cichocki, D. Mandic, L. de Lathauwer et al., "Tensor decompositions for signal processing applications: from two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.

[20] R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an explanatory multimodal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[21] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and its Applications*, vol. 18, no. 2, pp. 95–138, 1977.

[22] N. Sidiropoulos, R. Bro, and G. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2377–2388, 2000.

[23] N. Sidiropoulos, G. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 810–823, 2000.

[24] B. Xu, Y. Zhao, Z. Cheng, and H. Li, "A novel unitary PARAFAC method for DOD and DOA estimation in bistatic MIMO radar," *Signal Processing*, vol. 138, pp. 273–279, 2017.

[25] X. Zhang, Z. Xu, L. Xu, and D. Xu, "Trilinear decomposition-based transmit angle and receive angle estimation for multiple-input multiple-output radar," *IET Radar, Sonar & Navigation*, vol. 5, no. 6, pp. 626–631, 2011.

[26] J. Li and M. Zhou, "Improved trilinear decomposition-based method for angle estimation in multiple-input multiple-output radar," *IET Radar, Sonar & Navigation*, vol. 7, no. 9, pp. 1019–1026, 2013.

[27] F. Wen and G. Zhang, "Two-dimensional direction-of-arrival estimation for trilinear decomposition- based monostatic cross MIMO radar," *Mathematical Problems in Engineering*, vol. 2013, Article ID 427980, 8 pages, 2013.

[28] L. Xu, F. Wen, and X. Zhang, "A novel unitary PARAFAC algorithm for joint DOA and frequency estimation," *IEEE Communications Letters*, vol. 23, no. 4, pp. 660–663, 2019.

[29] Y. M. Chen, J. H. Lee, and C. C. Yeh, "Two-dimensional angle-of-arrival estimation for uniform planar arrays with sensor position errors," *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 1, pp. 37–42, 1993.

[30] Q. Wu, F. Sun, P. Lan, G. Ding, and X. Zhang, "Two-dimensional direction-of-arrival estimation for co-prime planar arrays: a partial spectral search approach," *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5660–5670, 2016.

[31] P. Ma, J. Li, G. Zhao, and X. Zhang, "Computation-efficient 2-D DOA estimation algorithm with array motion strategy," *Digital Signal Processing*, vol. 112, article 103013, 2021.

[32] S. Marcos, A. Marsal, and M. Benidir, "The propagator method for source bearing estimation," *Signal Processing*, vol. 42, no. 2, pp. 121–138, 1995.

[33] J. Li, X. Zhang, and H. Chen, "Improved two-dimensional DOA estimation algorithm for two-parallel uniform linear arrays using propagator method," *Signal Processing*, vol. 92, no. 12, pp. 3032–3038, 2012.

[34] R. Bro, *"Multi-Way Analysis in the Food Industry: Models, Algorithms, and Applications,"* [Ph.D. thesis], University of Amsterdam, 1998.

[35] D. Nion and L. De Lathauwer, "An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA," *Signal Processing*, vol. 88, no. 3, pp. 749–755, 2008.

WILEY | Hindawi

Research Article

Gridless Multiple Measurements Method for One-Bit DOA Estimation with a Nested Cross-Dipole Array

Haining Long [iD], Ting Su [iD], Xianpeng Wang [iD], and Mengxing Huang [iD]

Information and Communication Engineering, Hainan University, Hai Kou 570228, China

Correspondence should be addressed to Ting Su; suting4190@hainanu.edu.cn and Xianpeng Wang; wxpeng1986@126.com

The gridless one-bit direction of arrival (DOA) estimator is proposed to estimate electromagnetic (EM) sources on a nested cross-dipole array, and the multiple measurement vectors (MMV) mode is introduced to improve the reliability of parameter estimation. The gridless method is based on atomic norm minimization, solved by alternating direction multiplier method (ADMM). With gridless method used, sign inconsistency caused by one-bit measurements and basis mismatches by traditional grid-based algorithms can be avoided. Furthermore, the reconstructed denoising measurements with fast convergence and stable recovery accuracy are obtained by ADMM. Finally, spatial smoothing root multiple signal classification (SSRMUSIC) and dual polynomial (DP) methods are used, respectively, to estimate the DOAs on the reconstructed denoising measurements. Numerical results show that our method one-bit ADMM-SSRMUSIC has a better performance than that of one-bit SSRMUSIC used directly. At low signal to noise ratio (SNR) and low snapshot, the one-bit ADMM-DP has an excellent performance which is even better than that of unquantized MUSIC. In addition, the proposed methods are also suitable for both completely polarized (CP) signals and partially polarized (PP) signals.

1. Introduction

Since electromagnetic (EM) source signal carries the complete information buried in the fields, the direction finding of the EM source signal has a wide application. For instance, the direction of arrival (DOA) estimation of EM source signals was solved, respectively, by rotational invariance technique (ESPRIT) algorithm [1–3], maximum likelihood (ML) algorithm [4], and multiple signal classification-(MUSIC-) based solution [5]. The performance of parameter estimation is closely related to the array structures [6]. For a uniform linear array (ULA), N sensors can only restore $N-1$ sources at most [7, 8]. As a result, a large-scale ULA is needed for estimating multiple source signals, and the power consumption of analog-to-digital converters (ADC) increases exponentially.

A sparse array is a viable alternative, which can generate virtual sensors by using the position relation among physical sensors and obtain more degree of freedoms (DOFs). Sparse arrays mainly include minimum redundancy arrays (MRA) [9], nested arrays [10], coprime arrays [11], and other sparse

arrays based on them [12, 13]. The reversed and shift sparse array based on the difference and sum coarray is proposed for longer consecutive virtual array [14]. A novel sparse array with displaced multistage cascade subarrays is proposed to obtain high DOF [15]. [16] increases the DOF by using synthetic aperture technology and the concept of difference coarray. Moreover, the nested array is widely concerned for its unique and continuous virtual array elements and simple construction [17]. In [18], a nested cross-dipole array is used to estimate the DOA of EM source signals with spatial smoothing MUSIC (SS-MUSIC). Nevertheless, the hardware cost of ADC is still at a high level for large-scale nested cross-dipole arrays. One-bit measurements are popular to decline the consumption [19–21]. In [22], one-bit SS-MUSIC is proposed for DOA estimation of EM source signals.

In [23], the authors present that sign inconsistency are induced by one-bit measurements with additive noises and an atomic norm minimization with a linear loss function can overcome the shortage. Meanwhile, as a gridless method, atomic norm minimization can avoid the basis mismatches

caused by discretization of the signal parameter space [24, 25]. The alternating direction multiplier method (ADMM) is proposed to solve the atomic norm minimization with one-bit quantization in the single measurement vector (SMV) model [23, 26, 27]. ADMM has lower computational complexity and higher convergence speed compared with the semidefinite programming (SDP) [28].

In this paper, the atomic norm minimization with ADMM is extended for reconstructing the one-bit EM source signals using a nested cross-dipole array. The multiple measurement vectors (MMV) model is used instead of the SMV model. The MMV model has a higher success rate to reconstruct unknown signals compared with that of the SMV model. And then, two DOA estimation methods based on the constructed denoising measurements are proposed, spatial smoothing root-MUSIC (SSRMUSIC) algorithm [22] and dual polynomial (DP) method [23, 29]. In order to highlight the advantage of ADMM in signal reconstruction, SSRMUSIC algorithm with constructed denoising measurements will be compared with ordinary one-bit SSRMUSIC algorithm. And SSRMUSIC is robust in high SNR and multiple snapshots, but with the decrease of SNR and snapshot number, its resolution decreases rapidly. Therefore, the DP method is introduced for high recovery accuracy in low SNR and low snapshot. The main contributions of this paper are as follows: (1) The formulas of one-bit ADMM are derived to estimate DOAs of EM source signals using a nested cross-dipole array, and the MMV model is used instead of SMV to improve estimation accuracy. (2) With the gridless method based on atomic norm used, the sign inconsistency caused by one-bit quantization and basis mismatches by grid discretization are both avoid. (3) Two methods, SSRMUSIC and DP method, are used to solve the proposed problem, respectively. Simulation results show that one-bit ADMM SSRMUSIC has a better performance than that of one-bit SSRMUSIC used directly, while one-bit ADMM DP has an excellent performance in the situation of low signal to noise ratio (SNR) and low snapshot. (4) Most of the algorithms [30–32] for estimating the space parameters of both completely polarized (CP) signals and partially polarized (PP) signals are not universal. Our proposed algorithms are suitable for both CP signals and PP signals.

The rest of this paper is structured as follows. Section 2 describes the signal model of a nested cross-dipole array. The one-bit quantization model and difference coarray are introduced in Section 3. Section 4 presents the atomic norm minimization for the MMV model. The one-bit ADMM formulas and two methods for DOA estimation are also derived in Section 4. Simulation results are given and analysed to verify the efficiency and accuracy of the proposed algorithms in Section 5. Section 6 summarizes this paper.

Notations are as follows: $\Re\{\mathbf{X}\}$ and $\Im\{\mathbf{X}\}$ are real part and imaginary part of $\mathbf{X}$, respectively. $\mathbf{I}_M$ is the identity matrix by $M \times M$ and $\mathbf{T}(\mathbf{u})$ denotes the Toeplitz matrix whose first column is $\mathbf{u}$. $\mathrm{Tr}(\mathbf{X})$, $\det(\mathbf{X})$, $\langle\mathbf{X}\rangle_R$, $\mathrm{vec}(\mathbf{X})$, and $E(\mathbf{X})$ present the trace, the determinant, the real inner product, the vectorization, and the expectation of $\mathbf{X}$, respectively. $\mathbf{X}^\dagger$, $\mathbf{X}^H$, $\mathbf{X}^T$, and $\mathbf{X}^*$ denote the Moore-Penrose pseudoin-

verse, the conjugate transpose matrix, the transpose, and the conjugation of $\mathbf{X}$, respectively. $\textbf{\textit{conj}}(\cdot)$ means the conjugate operation on each entry of a vector or a matrix. $\mathbf{e}_1$ is a vector where the first element is 1, and the rest are 0. $\inf(\cdot)$ means the infimum operator. $\mathrm{conv}(\mathscr{A})$ is the convex hull of $\mathscr{A}$. $\|\cdot\|_{\mathscr{A}}^*$ is the dual atomic norm.

2. Signal Model

Suppose that K narrowband EM source signals impinge onto a nested cross-dipole array, the sensors with locations $\{d_1 d, d_2 d, \cdots, d_M d\}, d_m \in \mathbb{S}, m = 1, \cdots, M$. M is the number of elements, $M = M_1 + M_2$. The unit interelement spacing d is usually set as half-wavelength. The signal model is shown in Figure 1. Each cross-dipole consists of two dipoles parallel to the x-axis and y-axis, respectively [18]. Consequently, the received signals for x-axis and y-axis at sensor m can be given by

$$x_m^{[l]}(t) = \sum_{k=1}^{K} B_k^{[l]} s_k^{[l]}(t) a_m\left(\overline{\theta}_k\right) + n_m^{[l]}(t) \quad l = x, y, \tag{1}$$

where θ_k and $\overline{\theta}_k$ denote the DOA and the normalized DOA of the kth source, respectively. $\theta_k \in [-\pi/2, \pi/2], \overline{\theta}_k = \sin\theta_k/2$, and $\overline{\theta}_k \in [-1/2, 1/2]$. $a_m(\overline{\theta}_k) = e^{j2\pi\overline{\theta}_k d_m}$ is the spatial response of the mth dipole for the kth source. $B_k^{[l]}$ and $s_k^{[l]}(t)$ present the cross-dipole response and signal for l-axis of the kth source. Moreover, $B_k^{[x]} = -1$ and $B_k^{[y]} = \cos\left(\arcsin 2\overline{\theta}_k\right)$. $n_m^{[l]}(t)$ is the noise component for l-axis at the mth sensor. When x-axis and y-axis are considered simultaneously, $x_m^{[l]}(t)$, $n_m^{[l]}(t)$, and $s_k^{[l]}(t)$ can be written as vectors: $\mathbf{x}_m(t) = [x_m^{[x]}(t), x_m^{[y]}(t)]^T$, $\mathbf{n}_m(t) = [n_m^{[x]}(t), n_m^{[y]}(t)]^T$, and $\mathbf{s}_k(t) = [s_k^{[x]}(t), s_k^{[y]}(t)]^T$. The covariance of $\mathbf{s}_k(t)$ is given by [4].

$$\mathbf{R}_{\mathbf{s}_k} = E\left[\mathbf{s}_k(t)\mathbf{s}_k^H(t)\right] = \frac{p_{k_{UP}}^2}{2}\mathbf{I}_2 + p_{k_{CP}}^2 \boldsymbol{\Phi}(\alpha_k)\mathbf{w}(\beta_k)\mathbf{w}^H(\beta_k)\boldsymbol{\Phi}^H(\alpha_k), \tag{2}$$

where

$$\boldsymbol{\Phi}(\alpha_k) = \begin{bmatrix} \cos(\alpha_k) & \sin(\alpha_k) \\ -\sin(\alpha_k) & \cos(\alpha_k) \end{bmatrix}, \tag{3}$$

$$\mathbf{w}(\beta_k) = [\cos(\beta_k)\ j\sin(\beta_k)]^T, \tag{4}$$

with $\alpha_k \in (-\pi/2, \pi/2)$ and $\beta_k \in (-\pi/4, \pi/4)$ being the polarization orientation angle and polarization ellipticity angle, respectively. $p_{k_{CP}}^2$ and $p_{k_{UP}}^2$ denote the power of the kth source
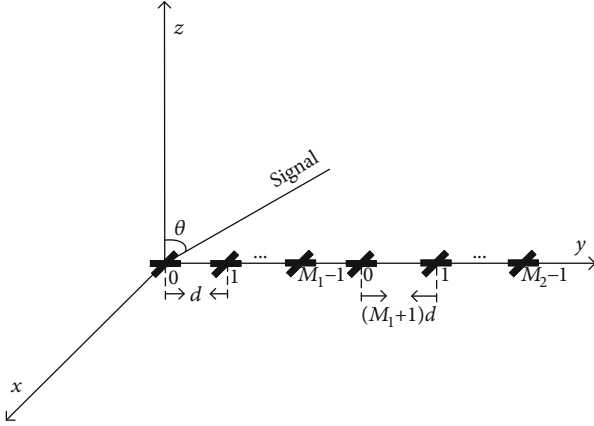
FIGURE 1: Nested array of cross-dipoles [22].

of CP signals and unpolarized (UP) signals, respectively. The degree of polarization (DOP) of $\mathbf{s}_k(t)$ can be calculated from the $\mathbf{R}_{s_k}$.

$$\eta_k = \left[1 - \frac{4 \det\left(\mathbf{R}_{s_k}\right)}{\left[\mathrm{Tr}\left(\mathbf{R}_{s_k}\right)\right]^2} \right]^{1/2}. \qquad (5)$$

Moreover, η_k is defined as the ratio of CP component power to total signal power [4].

$$\eta_k = \frac{p_{k_{CP}}^2}{p_{k_{CP}}^2 + p_{k_{UP}}^2}. \qquad (6)$$

The DOPs of CP signals and UP signals are $\eta_k = 1$ and $\eta_k = 0$, respectively, and that of PP signals is $\eta_k \in (0,1)$. CP signal of the kth source is expressed as

$$\mathbf{s}_k(t) = \begin{bmatrix} \cos \varphi_k \\ e^{j\psi_k} \sin \varphi_k \end{bmatrix} \tilde{s}_k(t), \qquad (7)$$

where $\varphi_k \in [0, \pi/2]$ and $\psi_k \in (-\pi, \pi]$ are polarization parameters, representing auxiliary polarization angle and the auxiliary polarization phase difference, respectively [20]. The covariance matrix $\mathbf{R}_{s_k}$ of CP signals and UP signals are all rank 2. Since PP signals can be expressed as the superposition of CP signals and UP signals, PP signals of the kth source can be expressed as

$$\mathbf{s}_k(t) = \left\{ \sqrt{\eta_k} \begin{bmatrix} \cos \varphi_k \\ e^{j\psi_k} \sin \varphi_k \end{bmatrix} + \sqrt{\frac{1-\eta_k}{2}} \begin{bmatrix} \beta_1(t) \\ \beta_2(t) \end{bmatrix} \right\} \tilde{s}_k(t). \qquad (8)$$

Both $\beta_1(t)$ and $\beta_2(t)$ are Gaussian random processes with zero mean. The covariance matrix $\mathbf{R}_{s_k}$ of PP signals is rank deficient.

Consider the dipoles measurements of x-axis and y-axis, Equation (1) can be written as

$$\mathbf{x}_{\mathbb{S}}^{[l]}(t) = \sum_{k=1}^{K} B_k^{[l]} s_k^{[l]}(t) \mathbf{a}_{\mathbb{S}}\left(\overline{\theta}_k\right) + \mathbf{n}_{\mathbb{S}}^{[l]}(t) \ \ l = x, y, \qquad (9)$$

where $\mathbf{x}_{\mathbb{S}}^{[l]}(t) = [x_1^{[l]}(t), \cdots, x_M^{[l]}(t)]$ is the array response of l-axis, $\mathbf{a}_{\mathbb{S}}(\overline{\theta}_k) = [a_1(\overline{\theta}_k), \cdots, a_M(\overline{\theta}_k)]^T$ is the steering vector, and $\mathbf{n}_{\mathbb{S}}^{[l]}(t) = [n_1^{[l]}(t), \cdots, n_M^{[l]}(t)]^T$ denotes the additive noise of l-axis. Moreover, the source signals $\{s_1^{[l]}(t), \cdots, s_K^{[l]}(t)\}$ are assumed to be independent random Gaussian processes. The additive noise $\mathbf{n}_{\mathbb{S}}^{[l]}(t)$ is an independent additive complex white Gaussian process with distribution $\mathscr{CN}(0, (\sigma^{[l]})^2 \mathbf{I}_M)$. The covariance of $\mathbf{x}_{\mathbb{S}}^{[l]}(t)$ is

$$\begin{aligned} \mathbf{R}_{\mathbf{x}_{\mathbb{S}}^{[l]}} &= E\left[\mathbf{x}_{\mathbb{S}}^{[l]}(t) \left(\mathbf{x}_{\mathbb{S}}^{[l]}(t) \right)^H \right] \\ &= \sum_{k=1}^{K} \left(p_k^{[l]} \right)^2 \mathbf{a}_{\mathbb{S}}^{[l]}\left(\overline{\theta}_k\right) \left(\mathbf{a}_{\mathbb{S}}^{[l]}\left(\overline{\theta}_k\right) \right)^H + \left(\sigma^{[l]} \right)^2 \mathbf{I}_M, \end{aligned} \qquad (10)$$

where $\left(p_k^{[l]}\right)^2$ and $(\sigma^{[l]})^2$ denote the source power and noise power of the kth source of l-axis, respectively.

3. One-Bit Quantization

The one-bit quantizer can be implemented by a sign operation, which is defined as

$$\mathbf{sign}(x) = \begin{cases} -1, x \le 0, \\ 1, x > 0. \end{cases} \qquad (11)$$

After one-bit quantizer, only the symbol information is left. Although the one-bit quantizer does not reduce the computational complexity, it greatly reduces the hardware consumption compared with the unquantized algorithm. Equation (9) is modified as

$$\mathbf{y}_{\mathbb{S}}^{[l]}(t) = \frac{1}{\sqrt{2}} \left[\mathrm{sign}\left(\Re\left\{ \mathbf{x}_{\mathbb{S}}^{[l]}(t) \right\} \right) + j \, \mathrm{sign}\left(\Im\left\{ \mathbf{x}_{\mathbb{S}}^{[l]}(t) \right\} \right) \right]. \qquad (12)$$

The factor $1/\sqrt{2}$ normalizes the power of $\mathbf{y}_{\mathbb{S}}^{[l]}(t)$ [19]. $\mathbf{y}_{\mathbb{S}}^{[x]}$ and $\mathbf{y}_{\mathbb{S}}^{[y]}$ are rewritten as $\mathbf{Y}_{\mathbb{S}}^{[x]}$ and $\mathbf{Y}_{\mathbb{S}}^{[y]}$ in multiple snapshots. The covariance matrices corresponding to $\mathbf{Y}_{\mathbb{S}}^{[x]}$ and $\mathbf{Y}_{\mathbb{S}}^{[y]}$ are approximated by the sample covariance, expressed as $\widehat{\mathbf{R}}_{\mathbf{Y}_{\mathbb{S}}^{[x]}}$ and $\widehat{\mathbf{R}}_{\mathbf{Y}_{\mathbb{S}}^{[y]}}$, respectively.

$$\widehat{\mathbf{R}}_{\mathbf{Y}_{\mathbb{S}}^{[l]}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{Y}_{\mathbb{S}}^{[l]}(n) \left(\mathbf{Y}_{\mathbb{S}}^{[l]}(n) \right)^H. \qquad (13)$$

We combine the DOA estimation with the difference coarray instead of the original array. The number of

estimable sources is closely related to the structure of the array. The difference coarray corresponding to $\mathbb{S}$ is defined as $\mathbb{D} = \{d_i - d_j | \forall d_i, d_j \in \mathbb{S}\}$. Specifically, the difference coarray of the nested array can be expressed as $\mathbb{D} = \{1 - M_2(M_1 + 1), \cdots, 0, \cdots, M_2(M_1 + 1) - 1\}$. $D_{\max} = M_2(M_1 + 1) - 1$, where $D_{\max}$ is the maximum number of DOFs [10]. The difference coarray thus occurs naturally in problems involving second-order statistics of the received signals [10]. The one-bit autocorrelation vectors of the difference coarray $\mathbb{D}$ for the x-axis and y-axis are defined as follows:

$$\mathbf{y}_{\mathbb{D}}^{[l]} = \mathbf{J}^{\dagger} \mathrm{vec}\left(\widehat{\mathbf{R}}_{\mathbf{Y}_{\mathbb{S}}^{[l]}}\right). \tag{14}$$

The definition of binary matrix $\mathbf{J}$ is the same as [33], with size $|\mathbb{S}|^2 - by - |\mathbb{D}|$. The columns of $\mathbf{J}$ satisfy $\langle \mathbf{J} \rangle_{:,\tilde{d}} = [\mathrm{vec}(\tilde{\mathbf{J}}(\tilde{d}))]^T, \tilde{d} \in \mathbb{D}$, where $\tilde{\mathbf{J}}(\tilde{d}) \in \{0, 1\}^{|\mathbb{S}| \times |\mathbb{S}|}$ is given by

$$\left\langle \tilde{\mathbf{J}}\left(\tilde{d}\right)\right\rangle_{d_i, d_j} = \begin{cases} 1, & \text{if } d_i - d_j = \tilde{d}, \\ 0, & \text{otherwise.} \end{cases} \forall d_i, d_j \in \mathbb{S} \tag{15}$$

4. The Proposed Methods

4.1. One-Bit Multiple Snapshots Model with Atomic Norm Minimization. The atomic norm is used to find the minimum number of atoms in the continuous parameter space; then, the basis mismatches by grid discretization will be avoided. And the linear loss function with characteristics of one-bit quantization will overcome the sign inconsistency caused by one-bit quantization. In order to make use of the joint sparsity between $\mathbf{y}_{\mathbb{D}}^{[x]}$ and $\mathbf{y}_{\mathbb{D}}^{[y]}$, the following definition is given:

$$\mathbf{Y} = \left[\mathbf{y}_{\mathbb{D}}^{[x]}, \mathbf{y}_{\mathbb{D}}^{[y]}\right]. \tag{16}$$

In Equation (16), $\mathbf{Y}$ is defined as the multiple snapshots (two snapshots) measurements. Let $\mathbf{X}$ be the denoising signal of $\mathbf{Y}$ and defines an atom to represent $\mathbf{X}$ as

$$\mathbf{A}(f, \mathbf{b}) = \mathbf{a}_{\mathbb{D}}(f)\mathbf{b}^H \in \mathscr{A}, \tag{17}$$

where $f \in [0, 1), \mathbf{b} \in \mathbb{C}^L, \|\mathbf{b}\|_2 = 1$, and $\mathscr{A}$ is the set of atoms. We assume that no element $\mathbf{A} \in \mathscr{A}$ lies in the convex hull of the other elements $\mathrm{conv}(\mathbf{A} \setminus \mathscr{A})$, i.e., the elements of $\mathscr{A}$ are the extreme points of $\mathrm{conv}(\mathscr{A})$ [34]. $\mathscr{A}$ is defined as

$$\mathscr{A} = \left\{\mathbf{A}(f, \mathbf{b}) \mid f \in [0, 1), \|\mathbf{b}\|_2 = 1\right\}, \tag{18}$$

where $\mathscr{A}$ is regarded as an infinite dictionary to describe the continuous changing parameters, and the atom of $\mathscr{A}$ is the basic unit to construct $\mathbf{X}$. $\|\mathbf{X}\|_{\mathscr{A},0}$ is defined to represent

the minimum number of atoms describing $\mathbf{X}$ [29].

$$\|\mathbf{X}\|_{\mathscr{A},0} = \inf_K \left\{\mathbf{X} = \sum_{k=1}^K c_k \mathbf{A}(f_k, \mathbf{b}_k), c_k \geq 0\right\}. \tag{19}$$

Since the minimization of Equation (19) is a NP problem, we consider the convex relaxation of $\|\mathbf{X}\|_{\mathscr{A},0}$, denoted by $\|\mathbf{X}\|_{\mathscr{A}}$. $\|\mathbf{X}\|_{\mathscr{A}}$ denotes the gauge of $\mathscr{A}$, and the gauge function can be defined as [35, 36].

$$\begin{aligned} \|\mathbf{X}\|_{\mathscr{A}} &= \inf \left\{t > 0 : \mathbf{X} \in t \, \mathrm{conv}(\mathscr{A})\right\} \\ &= \inf \left\{\sum_k c_k \middle| \mathbf{X} = \sum_k c_k \mathbf{A}(f_k, \mathbf{b}_k), c_k \geq 0\right\}, \end{aligned} \tag{20}$$

where $\|\mathbf{X}\|_{\mathscr{A}}$ is called the atomic norm of $\mathbf{X}$, which actually adds a sparse constraint to $\mathscr{A}$ but without discretization.

Considering the characteristics of one-bit quantization, and to ensure consistent recovery and constrain the signals to the unit ball, a linear loss function is proposed to recover the signals.

$$\begin{aligned} \arg\min_{\mathbf{X}_R, \mathbf{X}_S} \quad & -\frac{1}{2|\mathbb{D}|}\left(\left(\mathbf{Y}_R^{\breve{l}}\right)^T \mathbf{X}_R^{\breve{l}} + \left(\mathbf{Y}_S^{\breve{l}}\right)^T \mathbf{X}_S^{\breve{l}}\right) + \tau\|\mathbf{X}\|_{\mathscr{A}}, \\ s.t. \quad & \left\|\mathbf{X}_R^{\breve{l}}\right\|_1 + \left\|\mathbf{X}_S^{\breve{l}}\right\|_1 \leq 1, \end{aligned} \tag{21}$$

where $l \in L, \tau$ is regularization parameter, which is defined as

$$\tau = c\sqrt{L \log |\mathbb{D}|}, \tag{22}$$

where scale factor $c = 0.358$. It can be seen from Equation (16) that Equation (21) can be transformed into a matrix equation with dimension of 2×2 when $L = 2$. It is easy to know that only diagonal elements play an important role. Then, Equation (21) can be transformed into the formula about trace.

$$\begin{aligned} \arg\min_{\mathbf{X}_R, \mathbf{X}_S} \quad & -\frac{1}{2|\mathbb{D}|}\mathrm{Tr}\left(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\right) + \tau\|\mathbf{X}\|_{\mathscr{A}}, \\ s.t. \quad & \mathrm{Tr}\left(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\right) \leq \mathrm{Tr}(\mathbf{I}_2). \end{aligned} \tag{23}$$

Equation (23) can be solved by the SDP problem.

$$\begin{aligned} \arg\min_{\mathbf{X}_R, \mathbf{X}_S} \quad & -\frac{1}{2|\mathbb{D}|}\mathrm{Tr}\left(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\right) + \frac{\tau}{2}[\mathrm{Tr}(\mathbf{T}(\mathbf{u})) + \mathrm{Tr}(\mathbf{W})], \\ s.t. \quad & \begin{bmatrix} \mathbf{T}(\mathbf{u}) & \mathbf{X}_R + j\mathbf{X}_S \\ (\mathbf{X}_R + j\mathbf{X}_S)^H & \mathbf{W} \end{bmatrix} \geq 0, \\ & \mathrm{Tr}\left(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\right) \leq \mathrm{Tr}(\mathbf{I}_2). \end{aligned} \tag{24}$$

The SDP problem can be solved by CVX [35]. However, the convergence speed of SDP is very slow.

4.2. ADMM for One-Bit Multiple Snapshot Model.
In order to speed up the convergence speed and protect the accuracy of signal reconstructed, Equation (23) can also be solved by ADMM.

$$
\underset{\mathbf{X}_R, \mathbf{X}_S}{\arg\min} \quad -\frac{1}{2|\mathbb{D}|} \mathrm{Tr}\big(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\big) + \frac{\tau}{2}\big[\mathrm{Tr}(\mathbf{T}(\mathbf{u})) + \mathrm{Tr}(\mathbf{W})\big],
$$

$$
\textbf{s.t.} \qquad \mathbf{Z} = \begin{bmatrix} \mathbf{T}(\mathbf{u}) & \mathbf{X}_R + j\mathbf{X}_S \\ (\mathbf{X}_R + j\mathbf{X}_S)^H & \mathbf{W} \end{bmatrix},
$$

$$
\mathbf{Z} \geq 0,
$$

$$
\mathrm{Tr}(\mathbf{G}) = \mathrm{Tr}(\mathbf{I}_2) - \mathrm{Tr}\big(\mathbf{Y}_R^T \mathbf{X}_R\big) - \mathrm{Tr}\big(\mathbf{Y}_S^T \mathbf{X}_S\big),
$$

$$
\mathrm{Tr}(\mathbf{G}) \geq 0.
$$

$$(25)$$

$\mathbf{Z}$, $\boldsymbol{\Lambda}$, and $\mathbf{W}$ are Hermitian matrices, $\mathrm{Tr}(\mathbf{G})$ is the dual variable, and the augmented Lagrangian corresponding to the above formula is

$$
\begin{aligned}
&L(\mathbf{X}_R, \mathbf{X}_S, \mathbf{W}, \boldsymbol{\Lambda}, \mathbf{Z}, \mathrm{Tr}(\mathbf{G}), \mathbf{u}, \gamma) \\
&= -\frac{1}{2|\mathbb{D}|} \mathrm{Tr}\big(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\big) \\
&\quad + \frac{\tau}{2}\big[\mathrm{Tr}(\mathbf{T}(\mathbf{u})) + \mathrm{Tr}(\mathbf{W})\big] \\
&= \left\langle \boldsymbol{\Lambda}, \mathbf{Z} - \begin{bmatrix} \mathbf{T}(u) & \mathbf{X}_R + j\mathbf{X}_S \\ (\mathbf{X}_R + j\mathbf{X}_S)^H & \mathbf{W} \end{bmatrix} \right\rangle \\
&= \frac{\rho}{2} \left\| \mathbf{Z} - \begin{bmatrix} \mathbf{T}(u) & \mathbf{X}_R + j\mathbf{X}_S \\ (\mathbf{X}_R + j\mathbf{X}_S)^H & \mathbf{W} \end{bmatrix} \right\|_F^2 \\
&= \gamma \mathrm{Tr}\big(\mathbf{G} - \mathbf{I}_2 + \mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\big) \\
&= \frac{\rho}{2} \big\| \mathbf{G} - \mathbf{I}_2 + \mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S \big\|_F^2,
\end{aligned}
$$

$$(26)$$

where γ is the dual variable and ρ is the penalty parameter. Note that

$$
\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_0 & \mathbf{Z}_1 \\ \mathbf{Z}_2 & \mathbf{Z}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_0 & \mathbf{Z}_{1_R} + j\mathbf{Z}_{1_S} \\ (\mathbf{Z}_{1_R} + j\mathbf{Z}_{1_S})^H & \mathbf{Z}_3 \end{bmatrix}, \quad (27)
$$

$$
\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_0 & \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Lambda}_2 & \boldsymbol{\Lambda}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_0 & \boldsymbol{\Lambda}_{1_R} + j\boldsymbol{\Lambda}_{1_S} \\ (\boldsymbol{\Lambda}_{1_R} + j\boldsymbol{\Lambda}_{1_S})^H & \boldsymbol{\Lambda}_3 \end{bmatrix}. \quad (28)
$$

ADMM has the following update steps:

$$
\big(\mathbf{X}_R^{t+1}, \mathbf{X}_S^{t+1}, \mathbf{W}^{t+1}, \mathbf{u}^{t+1}\big) = \underset{\mathbf{X}, \mathbf{W}, \mathbf{u}}{\arg\min}\, L\big(\mathbf{X}_R, \mathbf{X}_S, \mathbf{W}, \mathbf{u}, \boldsymbol{\Lambda}^t, \mathbf{Z}^t, \mathrm{Tr}(\mathbf{G})^t, \gamma^t\big),
$$

$$(29)$$

$$
\big(\mathbf{Z}^{t+1}, \mathrm{Tr}(\mathbf{G})^{t+1}\big) = \underset{\mathbf{Z}, \mathrm{Tr}(\mathbf{G})}{\arg\min}\, L\big(\mathbf{X}_R^{t+1}, \mathbf{X}_S^{t+1}, \mathbf{W}^{t+1}, \mathbf{u}^{t+1}, \boldsymbol{\Lambda}^t, \mathbf{Z}^t, \mathrm{Tr}(\mathbf{G}), \gamma^t\big),
$$

$$(30)$$

$$
\boldsymbol{\Lambda}^{t+1} = \boldsymbol{\Lambda}^t + \rho\left(\mathbf{Z}^{t+1} - \begin{bmatrix} \mathbf{T}(\mathbf{u}^{t+1}) & \mathbf{X}_R^{t+1} + j\mathbf{X}_S^{t+1} \\ (\mathbf{X}_R^{t+1} + j\mathbf{X}_S^{t+1})^H & \mathbf{W}^{t+1} \end{bmatrix}\right),
$$

$$(31)$$

$$
\gamma^{t+1} = \gamma^t + \rho\mathrm{Tr}\big(\mathbf{G}^{t+1} - \mathbf{I}_2 + \mathbf{Y}_R^T \mathbf{X}_R^{t+1} + \mathbf{Y}_S^T \mathbf{X}_S^{t+1}\big). \quad (32)
$$

The closed form expressions updated with the iteration times t can be obtained.

$$
\mathbf{X}_R^{t+1} = \left(\rho\mathbf{I}_{|\mathbb{D}|} + \frac{\rho}{2}\mathbf{Y}_R\mathbf{Y}_R^T\right)^{-1}\left(\left(\frac{1}{4|\mathbb{D}|} + \frac{\tau}{2}\right)\mathbf{Y}_R + \rho\mathbf{Z}_R^t + \boldsymbol{\Lambda}_{1_R}^t + \frac{\rho}{2}\mathbf{Y}_R\mathbf{G}^t\right),
$$

$$(33)$$

$$
\mathbf{X}_S^{t+1} = \left(\rho\mathbf{I}_{|\mathbb{D}|} + \frac{\rho}{2}\mathbf{Y}_S\mathbf{Y}_S^T\right)^{-1}\left(\left(\frac{1}{4|\mathbb{D}|} + \frac{\tau}{2}\right)\mathbf{Y}_S + \rho\mathbf{Z}_S^t + \boldsymbol{\Lambda}_{1_S}^t + \frac{\rho}{2}\mathbf{Y}_S\mathbf{G}^t\right),
$$

$$(34)$$

$$
\mathbf{W}^{t+1} = \frac{1}{2}\mathbf{Z}_3^t + \frac{1}{2}\big(\mathbf{Z}_3^t\big)^H + \frac{1}{\rho}\left(\boldsymbol{\Lambda}_3^t - \frac{\tau}{2}\mathbf{I}_2\right), \quad (35)
$$

$$
\mathbf{u}^{t+1} = \frac{1}{\rho} \cdot \boldsymbol{\Psi} \cdot \mathrm{conj}\left(\mathscr{G}\big(\boldsymbol{\Lambda}_0^t\big) + \rho\mathscr{G}\big(\mathbf{Z}_0^t\big) - \frac{\tau}{2}|\mathbb{D}|\mathbf{e}_1\right), \quad (36)
$$

where $\mathbf{a} = \mathscr{G}(\mathbf{A})$ is the mapping of a matrix to a vector, where the *ith* entry in $\mathbf{a}$ is the sum of all entries $\mathbf{A}_{j,\tilde{j}}$'s of $\mathbf{A}$, satisfying $j - \tilde{j} + 1 = \tilde{d}$ [29]. $\boldsymbol{\Psi}$ is a diagonal matrix with diagonal elements $\boldsymbol{\Psi}_{\tilde{d},\tilde{d}} = 1/(|\mathbb{D}| - \tilde{d} + 1)$, $\tilde{d} = 1, \cdots, |\mathbb{D}|$. The update of $\mathrm{Tr}(G)$ is expressed as

$$
\mathrm{Tr}(\mathbf{G})^{t+1} = \left[\mathrm{Tr}\left[\left(1 - \frac{\tau}{\rho}\right)\mathbf{I}_2 - \mathbf{Y}_R^T\mathbf{X}_R^{t+1}\right]\right]_+. \quad (37)
$$

Let

$$
\boldsymbol{\Psi}^t = \begin{bmatrix} \mathbf{T}(\mathbf{u}^{t+1}) & \mathbf{X}_R^{t+1} + j\mathbf{X}_S^{t+1} \\ (\mathbf{X}_R^{t+1} + j\mathbf{X}_S^{t+1})^H & \mathbf{W}^{t+1} \end{bmatrix} - \frac{1}{\rho}\boldsymbol{\Lambda}^t, \quad (38)
$$

and $\mathbf{U}^t \mathrm{diag}\,(\{\sigma_{\tilde{d}}^t\})(\mathbf{U}^t)^H$ is the eigenvalue decomposition of $\boldsymbol{\Psi}^t$. Then, the updated closed form expression of $\mathbf{Z}$ can be given as

$$
\mathbf{Z}^{t+1} = \mathbf{U}^t \mathrm{diag}\,\big(\{\sigma_{\tilde{d}}^t\}\big)\big(\mathbf{U}^t\big)^H. \quad (39)
$$

The reconstructed denoising measurements $\hat{\mathbf{X}}$ can be obtained through the aforementioned closed form expressions. Next, two specific DOA estimation methods for the $\hat{\mathbf{X}}$ will be introduced: SSRMUSIC algorithm and DP method.

4.3. One-Bit ADMM-SSRMUSIC Algorithm.
One-Bit ADMM-SSRMUSIC first utilizes one-bit ADMM to reconstruct the measurements $\hat{\mathbf{X}}$ and then estimates DOAs by SSRMUSIC. SSRMUSIC is the algorithm that obtains the

covariance matrix with full rank by spatial smoothing [37] firstly and then finds the directions by Root-MUSIC [38]. In order to find directions by Root-MUSIC, $\widehat{\mathbf{X}}$ will be divided into $\tilde{L}$ subarrays, where $\tilde{L} = |\mathbb{D}| - D_{\max}$. The subarray can be expressed as $\widehat{\mathbf{X}}_{\tilde{l}}, \tilde{l} = 1, \cdots, \tilde{L}$. The full-rank covariance matrix can be obtained by the above operations.

$$\widehat{\mathbf{R}}_{\widehat{\mathbf{X}}_{\tilde{l}}} = \frac{1}{\tilde{L}} \sum_{\tilde{l}}^{\tilde{L}} E\left(\widehat{\mathbf{X}}_{\tilde{l}} \widehat{\mathbf{X}}_{\tilde{l}}^H\right). \tag{40}$$

Then, the DOAs can be estimated by Root-MUSIC based on Equation (40). The subspace algorithm is a very classical and well-known DOA estimation method, so we will not introduce the subspace algorithm in detail but focus on the DP method.

4.4. One-Bit ADMM-DP Algorithm. The first step of one-bit ADMM-DP is the same as one-bit ADMM-SSRMUSIC and then estimates DOAs by the DP method. Each norm has a corresponding dual norm. Compared with the original norms, dual norms have several useful properties and are widely used in many problems. The DP method transforms the optimal solution of the original problem to that of the dual problem. We can obtain the frequency support set and estimate the DOA [23, 29].

The Lagrangian corresponding to Equation (23) is

$$L(\mathbf{X}_R, \mathbf{X}_I, \text{Tr}(\mathbf{G})) = \tau\|\mathbf{X}\|_{\mathscr{A}} - \frac{1}{2|\mathbb{D}|} \text{Tr}\left(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\right)$$
$$+ \text{Tr}(\mathbf{G})\left[\text{Tr}\left(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\right) - \text{Tr}(\mathbf{I}_2)\right]. \tag{41}$$

The dual function of Equation (41) is as follows:

$$g(\text{Tr}(\mathbf{G})) = \inf_{\mathbf{X}_R, \mathbf{X}_S} L(\mathbf{X}_R, \mathbf{X}_S, \text{Tr}(\mathbf{G}))$$
$$= -2\text{Tr}(\mathbf{G}) + \inf_{\mathbf{X}_R, \mathbf{X}_S}\left[\tau\|\mathbf{X}\|_{\mathscr{A}} + \text{Tr}(\mathbf{G})\text{Tr}\left(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\right)\right.$$
$$\left. - \frac{1}{2|\mathbb{D}|} \text{Tr}\left(\mathbf{Y}_R^T \mathbf{X}_R + \mathbf{Y}_S^T \mathbf{X}_S\right)\right]$$
$$= -2\text{Tr}(\mathbf{G}) + \inf_{\mathbf{X}_R, \mathbf{X}_S}\left[\tau\|\mathbf{X}\|_{\mathscr{A}} - \left(\frac{1}{2|\mathbb{D}|} - \text{Tr}(\mathbf{G})\right)\text{Tr}[\langle \mathbf{Y}, \mathbf{X}\rangle_R]\right]$$
$$= -2\text{Tr}(\mathbf{G}) + \inf_{\mathbf{X}_R, \mathbf{X}_S}\left[\tau\|\mathbf{X}\|_{\mathscr{A}} - \left\langle\left(\frac{1}{2|\mathbb{D}|} - \text{Tr}(\mathbf{G})\right)\mathbf{Y}^*, \mathbf{X}\right\rangle_R\right]$$
$$= -2\text{Tr}(\mathbf{G}) + I_{\left\{\varpi:\|\varpi\|_A^* \leq \tau\right\}}(\tilde{\mathbf{X}}), \tag{42}$$

where $\tilde{\mathbf{X}} = (1/2|\mathbb{D}| - \text{Tr}(\mathbf{G}))\mathbf{Y}^*$. $I_{\mathscr{A}}(\cdot)$ is an indicator function; the following formula can also be expressed as

$$I_{\left\{\varpi:\|\varpi\|_A^* \leq \tau\right\}}(\tilde{\mathbf{X}}) = \begin{cases} 0, & \|\tilde{\mathbf{X}}\|_{\mathscr{A}}^* \leq \tau, \\ -\infty, & \text{otherwise.} \end{cases} \tag{43}$$

From Equations (42) and (43), we have $\tau\|\mathbf{X}\|_{\mathscr{A}} - \langle\tilde{\mathbf{X}}, \mathbf{X}\rangle_R \leq 0$,

$$\tau\|\mathbf{X}\|_{\mathscr{A}} \leq \langle\tilde{\mathbf{X}}, \mathbf{X}\rangle_R. \tag{44}$$

The dual atomic norm is defined as

$$\|\tilde{\mathbf{X}}\|_{\mathscr{A}}^* = \sup_{\|\mathbf{X}\|_{\mathscr{A}} \leq 1} \langle\tilde{\mathbf{X}}, \mathbf{X}\rangle_R = \sup_{\mathbf{A} \in \mathscr{A}} \langle\tilde{\mathbf{X}}, \mathbf{A}\rangle_R. \tag{45}$$

From the definition of dual atomic norm and $\|\tilde{\mathbf{X}}\|_{\mathscr{A}}^* \leq \tau$, we can obtain

$$\langle\tilde{\mathbf{X}}, \mathbf{X}\rangle_R \leq \|\tilde{\mathbf{X}}\|_{\mathscr{A}}^* \|\mathbf{X}\|_{\mathscr{A}} \leq \tau\|\mathbf{X}\|_{\mathscr{A}}. \tag{46}$$

From Equations (44) and (46), we have $\langle\widehat{\tilde{\mathbf{X}}}, \widehat{\mathbf{X}}\rangle_R = \tau \|\widehat{\mathbf{X}}\|_{\mathscr{A}}$. It can be seen from Equations (17), (19), and (20), that $\widehat{\mathbf{A}}(f, \mathbf{b}) = \widehat{\mathbf{a}}_{\mathbb{D}}(f)\mathbf{b}^H$, $\widehat{\mathbf{X}} = \sum_{k=1}^K c_k \widehat{\mathbf{A}}(f_k, \mathbf{b}_k)$, $\|\widehat{\mathbf{X}}\|_{\mathscr{A}} = \sum c_k$. Finally,

$$\left\langle\widehat{\tilde{\mathbf{X}}}, \widehat{\mathbf{a}}_{\mathbb{D}}(f)\right\rangle_R = \tau. \tag{47}$$

As a result, the signal frequencies can be recovered and DOAs can be estimated from Equation (47).

5. Simulation Results

Considering the nested array $\mathbb{S} = \{1, 2, 3, 4, 5, 6, 12, 18, 24, 30\}$, the difference array corresponding to $\mathbb{S}$ is $\mathbb{D} = \{0, \pm 1, \pm 2, \cdots, \pm 29\}$. $\varphi_k = [0, \pi/2]$ and $\psi_k = [-\pi, \pi]$ are obtained randomly. $p_k^2 = 1$ and $SNR = 10 \log\left(\sum_{k=1}^K p_k^2/2k\sigma^2\right) = 10 \log\left(1/2\sigma^2\right)$. The performance of the DOA estimation is measured by the mean-square error (MSE) as

$$\text{MSE} = \frac{1}{RK} \sum_r^R \sum_k^K \left(\overline{\theta}\wedge_{r,k} - \overline{\theta}_k\right)^2, \tag{48}$$

where R means Monte Carlo runs, $R = 500$.

In order to show the advantages of the combination of the nested cross-dipole array and one-bit ADMM, the spatial spectrums are used to show the maximum number of sources that can be estimated. Figure 2 shows two spatial spectrums obtained by one-bit ADMM-SS-MUSIC and one-bit SS-MUSIC, respectively. One-bit ADMM-SS-MUSIC first reconstructs denoising measurements with one-bit ADMM and then estimates DOAs by SS-MUSIC, while one-bit SS-MUSIC estimates DOAs directly by SS-MUSIC [22]. The number of snapshots $N = 100$ and SNR $= 0$ dB. The number of sources $K = D_{\max} = 29$. The sources are uniformly distributed on $[-0.49, 0.49]$. Assuming that there are five CP signals, six UP sources, and eighteen PP signals, let the DOP of six PP signals are $\eta_k = 0.25$ and that of six PP signals are $\eta_k = 0.5$; the rest of PP signals are $\eta_k = 0.75$. We can see that 29 sources are completely estimated by one-bit ADMM-SS-MUSIC. Obviously, $K = 29 > |\mathbb{S}| = 10$
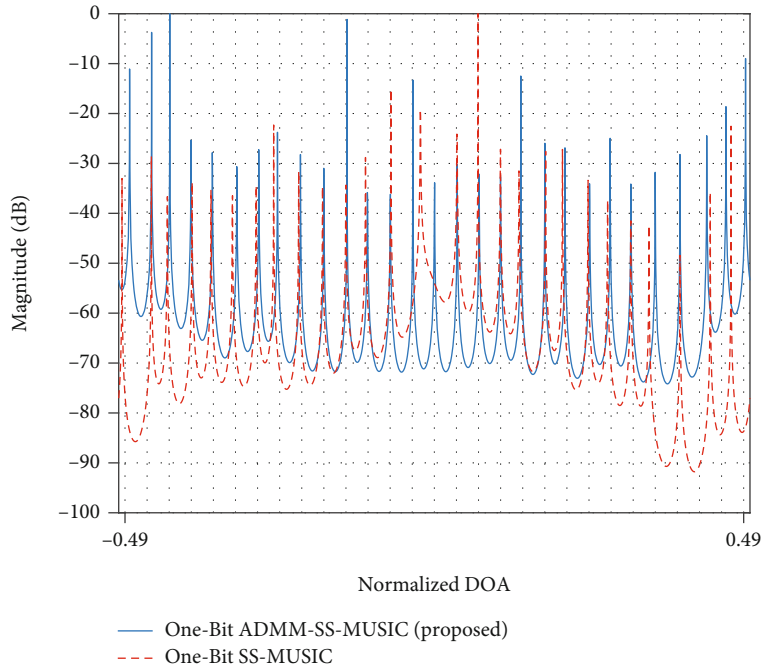
FIGURE 2: Spatial spectrum.
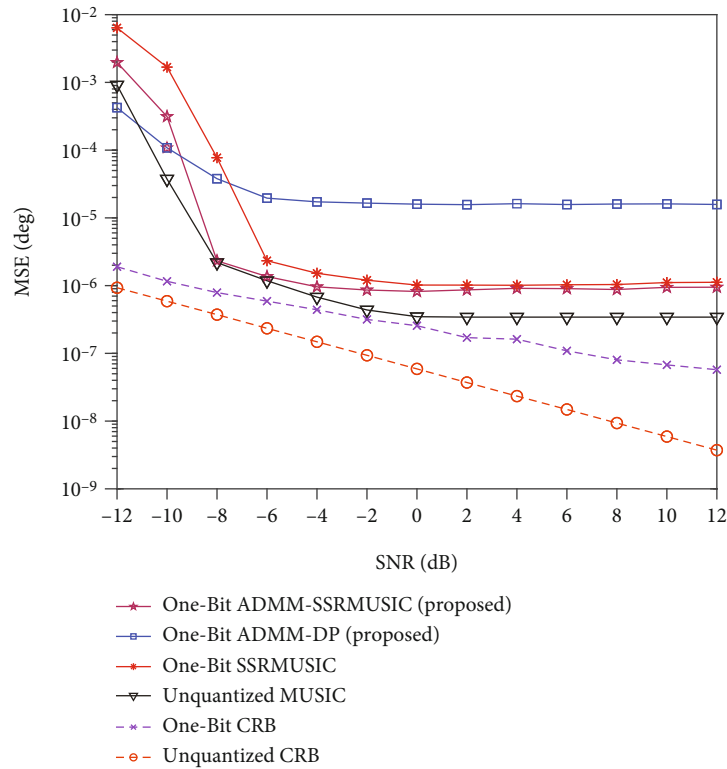
FIGURE 3: MSE versus the SNR.

. One-bit SS-MUSIC can only roughly estimate less than 29 sources, and the angle deviation is larger than one-bit ADMM-SS-MUSIC. It can infer that the algorithm will estimate more sources and improve the reliability if the signal is reconstructed by one-bit ADMM firstly.

Next, we consider the following DOA estimation algorithms:

(1) One-bit ADMM-SSRMUSIC: the gridless algorithm proposes in this paper
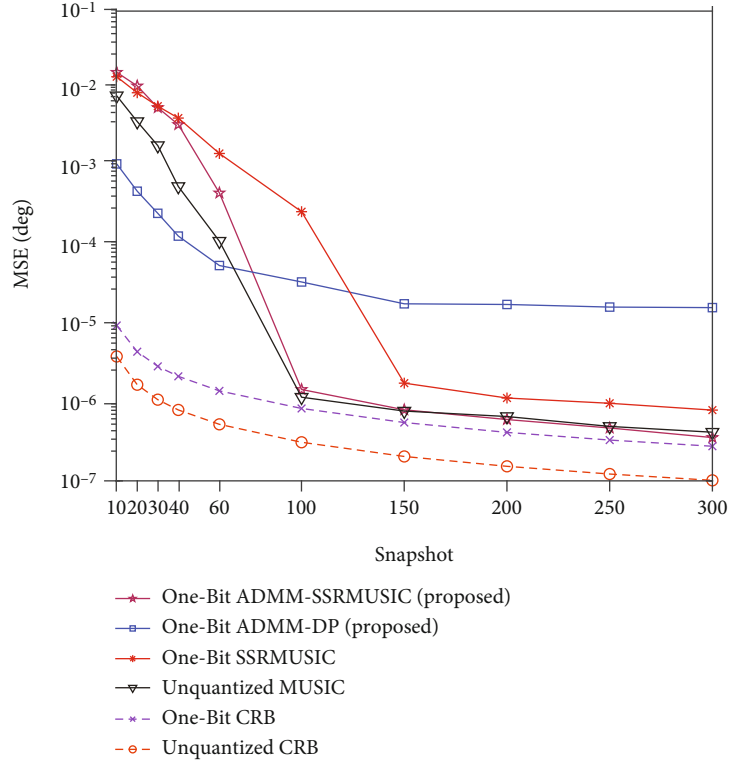
FIGURE 4: MSE versus the number of snapshots N.
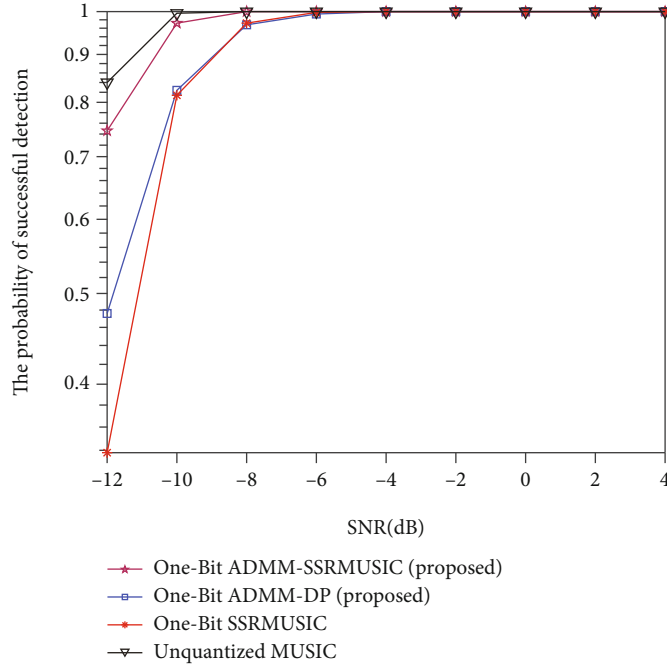
FIGURE 5: PSD versus the SNR.

(2) One-bit ADMM-DP: the gridless algorithm proposes in this paper

(3) One-bit SSRMUSIC: the method describes in [22], which finds directions directly by one-bit SS-Root-MUSIC

(4) Unquantized MUSIC: the method describes in [18], which is an unquantized method, and estimates DOAs by SS-MUSIC

Figure 3 illustrates the MSEs and Cramer-Rao bounds (CRBs) [22] versus the SNR with $N = 100$, $K = 5$, and $\overline{\theta}_k \in$

$\{-0.4, -0.2, 0, 0.2, 0.4\}$. The sources are divided into one CP signal, one UP signal, and three PP signals with $\eta_k = \{0, 0.25, 0.5, 0.75, 1\}$. If there is no special explanation, $\bar{\theta}_k$ and η_k of the sources will not change in the next simulation. It can be seen from Figure 3 that one-bit ADMM-DP has higher estimation accuracy than one-bit ADMM-SSRMUSIC and one-bit SSRMUSIC at low SNR. The accuracy of one-bit ADMM-DP even exceeds the unquantized MUSIC when SNR is near -12 dB. When SNR is less than about 9 dB, one-bit ADMM-DP has better performance than one-bit ADMM-SSRMUSIC. The accuracy of one-bit ADMM-SSRMUSIC is much higher than that of one-bit SSRMUSIC at low SNR, and the accuracy gap between them gradually decreases with SNR increasing. Obviously, ADMM is very effective for improving the accuracy of DOA estimation, especially at low SNR.

Figure 4 illustrates the MSEs and CRBs versus the number of snapshots with SNR $= -6$ dB, $K = 5$. In the range of $N = 10$ to $N = 60$, the estimation accuracy of one-bit ADMM-DP is higher than that of the other three algorithms. In addition, when the number of snapshots is less than about 80, one-bit ADMM-DP has better performance than one-bit ADMM-SSRMUSIC. It can be seen that the DP method is robust in the low snapshot. Besides, the accuracy of one-bit ADMM-SSRMUSIC and unquantized MUSIC is very close between $N = 100$ and $N = 300$. It reveals that the one-bit ADMM-SSRMUSIC and unquantized MUSIC have comparable performance in certain SNR and snapshot ranges. One-bit ADMM-SSRMUSIC and one-bit SSRMUSIC have almost the same recovery accuracy between $N = 10$ and $N = 40$, but the estimation accuracy of one-bit ADMM-SSRMUSIC is higher than that of one-bit SSRMUSIC with the increase of the number of snapshots. It can be recognized that the introduction of ADMM greatly improves performance.

Figure 5 shows the probability of a successful detection (PSD) versus the SNR with $N = 100$, $K = 5$. The PSD of the algorithms proposed in this paper (one-bit ADMM-SSRMUSIC and one-bit ADMM-DP) is higher than one-bit SSRMUSIC. In particular, one-bit ADMM-SSRMUSIC can achieve a 100% success rate at low SNR.

6. Conclusion

The one-bit ADMM on a nested cross-dipole array with the MMV model is used to estimate the DOAs of EM sources in this paper. Based on the properties of the cross-dipole array, one-bit ADMM is applied to the x-axis and y-axis dipole array and extended to the MMV model for solving the atomic norm minimization, the sign inconsistency will be solved and the basis mismatch will be avoided finally. The reconstructed signal denoising measurements will be obtained via one-bit ADMM. Finally, the SSRMUSIC algorithm and DP method are derived to estimate DOAs by the reconstructed denoising measurements. The simulation results show that the proposed algorithms are robust to DOA estimation of CP signals and PP signals, and ADMM is an effective method to improve the accuracy of DOA esti-

mation. The one-bit ADMM-SSRMUSIC has a better performance than the one-bit SSRMUSIC in DOA estimation. Moreover, the accuracy of one-bit ADMM-DP is even surpassing that of unquantized MUSIC at low SNR and low snapshot.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] J. Li and R. T. Compton, "Angle and polarization estimation using ESPRIT with a polarization sensitive array," *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 9, pp. 1376–1383, 1991.

[2] J. Li and R. T. Compton, *Angle estimation using a polarization sensitive array*, vol. 1, Antennas and Propagation Society Symposium 1991 Digest, London, Ontario, Canada, 1991.

[3] J. Li, "On polarization estimation using a polarization sensitive array," in *IEEE Sixth SP Workshop on Statistical Signal and Array Processing*, pp. 465–468, Victoria, BC, Canada, 1992.

[4] J. He, M. O. Ahmad, and M. N. S. Swamy, "Near-field localization of partially polarized sources with a cross-dipole array," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 2, pp. 857–870, 2011.

[5] K. Wang, J. He, T. Shu, and Z. Liu, "Localization of mixed completely and partially polarized signals with crossed-dipole sensor arrays," *Sensors*, vol. 15, no. 12, pp. 31859–31868, 2015.

[6] C. L. Liu and P. P. Vaidyanathan, "One-bit sparse array DOA estimation," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. , 20173126–3130, 2017.

[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[8] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[9] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Transactions on Antennas and Propagation*, vol. 16, no. 2, pp. 172–175, 2003.

[10] P. Pal and P. P. Vaidyanathan, "Nested arrays: a novel approach to array processing with enhanced degrees of freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, 2010.

[11] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.

[12] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the music algorithm," in *2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)}*, pp. 289–294, 2011.

[13] S. Qin, Y. D. Zhang, and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1377–1390, 2015.

[14] A. Cui, T. Xu, and W. Yu, "An array interpolation based compressive sensing DOA method for sparse array," in *2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC)*, pp. 24–27, 2019.

[15] Y.-k. Zhang, H.-y. Xu, R. Zong, B. Ba, and D.-m. Wang, "A novel high degree of freedom sparse array with displaced multistage cascade subarrays," *Digital Signal Processing*, vol. 90, pp. 36–45, 2019.

[16] S. Li and X. P. Zhang, "A new approach to construct virtual array with increased degrees of freedom for moving sparse arrays," *IEEE Signal Processing Letters*, vol. 27, pp. 805–809, 2020.

[17] J. Shi, F. Wen, and T. Liu, *Nested MIMO Radar: Coarrays, Tensor Modeling and Angle Estimation*, IEEE Transactions on Aerospace and Electronic Systems, 2020.

[18] J. He, Z. Zhang, T. Shu, and W. Yu, "Direction finding of multiple partially polarized signals with a nested cross-dipole array," *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 1679–1682, 2017.

[19] L. Pan, S. Xiao, and B. Li, "Continuous-time signal recovery from 1-bit multiple measurement vectors," *AEU-International Journal of Electronics and Communications*, vol. 76, pp. 132–136, 2017.

[20] T. Chen, M. Guo, and X. Huang, "Direction finding using compressive one-bit measurements," *IEEE Access*, vol. 6, pp. 41201–41211, 2018.

[21] Y. Gao, D. Hu, Y. Chen, and Y. Ma, "Gridless 1-b DOA estimation exploiting SVM approach," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2210–2213, 2017.

[22] Z. Cheng, S. Chen, and Q. Shen, "Direction finding of electromagnetic sources on a sparse cross-dipole array using one-bit measurements," *IEEE Access*, vol. 8, pp. 83131–83143, 2020.

[23] Z. Wei, W. Wang, and F. Dong, "Gridless one-bit direction-of-arrival estimation via atomic norm denoising," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2177–2181, 2020.

[24] G. Tang, B. N. Bhaskar, and P. Shah, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[25] L. Wan, X. Kong, and F. Xia, "Joint range-Doppler-angle estimation for intelligent tracking of moving aerial targets," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1625–1636, 2018.

[26] F. Wen and J. Shi, "Fast direction finding for bistatic EMVS-MIMO radar without pairing," *Signal Process*, vol. 173, p. 107512, 2020.

[27] X. Wang, M. Huang, and L. Wan, *Joint 2D-DOD and 2D-DOA Estimation for Coprime EMVS-MIMO Radar*, Circuits, Systems, and Signal Processing, accept, 2020.

[28] S. Boyd, N. Parikh, and E. Chu, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[29] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, 2016.

[30] Wong, L. Li, and Zoltowski, "Root-MUSIC-based direction-finding and polarization estimation using diversely polarized possibly collocated antennas," *IEEE Antennas and Wireless Propagation Letters*, vol. 3, pp. 129–132, 2004.

[31] Y. Merah, S. Miron, and D. Brie, "A generalized acquisition scheme for vector cross-product direction finding with spatially spread vector-sensor components," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 3977–3980, Vancouver, BC, 2013.

[32] D. Rahamim, J. Tabrikian, and R. Shavit, "Source localization using vector sensor array in a multipath environment," *IEEE Transactions on Signal Processing*, vol. 52, no. 11, pp. 3096–3103, 2004.

[33] C. Liu and P. P. Vaidyanathan, "New Cramer-Rao bounds for coprime and other sparse arrays," *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. , 20161–5, 2016.

[34] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[35] R. T. Rockafellar, *Convex Analysis*, Princeton University Pre, 1970.

[36] F. F. BONSALL, "A general atomic decomposition theorem and Banach's closed range theorem," *Quarterly Journal of Mathematics*, vol. 42, no. 1, pp. 9–14, 1991.

[37] M. L. McCloud and L. L. Scharf, "A new subspace identification algorithm for high-resolution DOA estimation," *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 10, pp. 1382–1390, 2002.

[38] B. D. Rao and K. V. S. Hari, "Performance analysis of root-music," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, 1989.

WILEY | Hindawi

Research Article

Compressed Sensing-Based Range-Doppler Processing Method for Passive Radar

Xia Bai [ID], Hejing Guo, Juan Zhao, and Tao Shan

School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Xia Bai; bai@bit.edu.cn

Passive radar (PR) systems use the existing transmitters of opportunity in the environment to perform tasks such as detection, tracking, and imaging. The classical cross-correlation based methods to obtain the range-Doppler map have the problems of high sidelobe and limited resolution due to the influence of signal bandwidth. In this paper, we propose a novel range-Doppler processing method based on compressed sensing (CS), which performs sparse reconstruction in range and Doppler dimensions to achieve high resolution and reduces sidelobe without excessive computational burden. Results from numerical simulations and experimental measurements recorded with the Chinese standard digital television terrestrial broadcasting (DTTB) based PR show that the proposed method successfully handles the range-Doppler map formatting problem for PR and outperforms the existing CS-based PR processing methods.

1. Introduction

Passive radar [1] (PR) is a kind of radar system which uses the existing transmitters of opportunity (such as FM [2, 3], GSM [4], and DVB-T [5, 6]) in space to achieve target detection and tracking and other tasks without special deployment or installation of transmitters. In recent years, PR has been widely concerned in the military and commercial fields because of advantages in terms of low-cost implementation, confidentiality, strong antijamming, and reduced electromagnetic pollution to the environment.

In this study, we mainly focus on the generation of a passive radar range-Doppler (RD) map. Based on the matched filtering theory, the classical method uses the cross-ambiguity function (CAF) and fast Fourier transform (FFT) to calculate RD response. However, the classical method faces some challenges. On the one hand, the generated RD map has a high sidelobe level, which may get false target position and Doppler frequency. On the other hand, the signal bandwidth of passive radar is usually narrow compared to that of active radar. It limits range resolution, and this limitation leads to undesirable performance in various applications [7]. In order to improve the range resolution of PR, the mul-

tiple broadcast channels from a single transmitter were exploited [8–10], which can be implemented only when the multiband system is used. Besides, a longer integration time can improve the Doppler resolution. But that would cause migration phenomenon during integration and require complex compensation [11, 12].

In recent years, the application of compressed sensing (CS) has been considered in passive radar. CS is a sparse signal processing technology [13, 14], which can reduce the amount of sampling data and use a small number of measurements to achieve excellent signal reconstruction. CS has been widely used in the field of radar signal processing; it has great potential in improving resolution [15, 16]. With the increasing attention in the field of PR, the RD map of targets for PR can be formulated as a sparse recovery problem, and the works related to CS-based PR processing have been published [17–19]. In [17, 18], a normal CS-based PR processing scheme was proposed to achieve better range and Doppler resolutions. Due to the fact that the dictionary is composed of the template signals with discrete delays and Doppler shifts, this method has high reconstruction accuracy, but it needs a large amount of storage and calculations. [19] proposed a RD map generation algorithm for PR. The

extended orthogonal matching pursuit (EOMP) algorithm is employed to obtain the Doppler frequency with an improved resolution and a reduced sidelobe level. But one-dimensional cross-correlation is used to obtain the range compressed profile, which is the same as the classical method.

In this paper, we propose a novel range-Doppler processing method for PR. CS is applied to the range domain and Doppler domain, respectively, which no longer requires huge storage space occupied by the dictionary. In addition, our contributions are the following. (1) In the Doppler dimension, considering the sparse characteristics of the signal, we present a modified OMP algorithm based on the multiple measurement vector (MMV) [20] model to improve the processing ability. (2) In the range dimension, we use the Fourier dictionary in the range-frequency domain, which can easily achieve high-resolution range estimation. Furthermore, a global search is to find the most relevant atom of the dictionary matrix for sparse reconstruction of the RD map, which can be treated as another way to use EOMP.

The rest of this paper is organised as follows. A brief signal model of PR is introduced in Section 2. In Section 3, a novel theoretical derivation of CS-based RD map generation for PR is presented. The comparison with other processing methods is also discussed in this section. Section 4 demonstrates experimental results using the simulated data and the real data. Finally, Section 5 concludes this work.

2. Passive Radar Geometry and Signal Model

Figure 1 schematically illustrates a typical passive radar geometry, where the system is composed of transmitting station T and receiving station R. It is assumed that there is a moving target P in the scene. R_T, R_R, and R_L represent the transmitter-target, receiver-target, and transmitter-receiver distances. v is the velocity vector of the moving target. β is the bistatic angle. ϕ is the angle between the vector v and the bistatic angle. θ_T is the emission angle, and θ_R is the target observation angle.

The PR receiver consists of two channels, the reference channel and surveillance channel. The reference channel gathers a time-delayed version of the transmitted waveform, and the surveillance channel records the signals scattered from targets. Let the transmitted waveform be represented by $x_T(t)$, then the signal collected by the reference channel can be written as

$$x(t) = A_r x_T(t - \tau_L) + n_r(t), \qquad (1)$$

where A_r is the complex amplitude, $\tau_L = R_L/c$ represents the time delay, and $n_r(t)$ represents the thermal noise in the reference channel.

Admittedly, the surveillance channel also contains direct signal and multipath in practice. Here, we assume that disturbance has been removed [21–24], and then, the response of the moving target P can be expressed as

$$y(t) = A x_T(t - \tau_p) \exp\left(j2\pi f_p t\right) + n(t), \qquad (2)$$
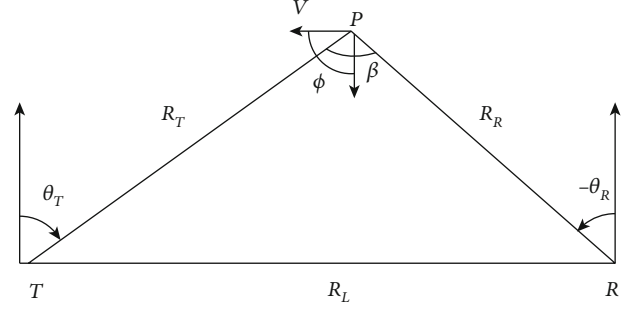
FIGURE 1: Simplified bistatic geometry for passive radar.

where A is the complex amplitude, τ_p is the bistatic time delay corresponding to the target location, f_p is the Doppler frequency shift related to the target velocity, and $n(t)$ is the thermal noise in the surveillance channel. According to the geometric relationship between the target and the bistatic radar system, the time delay and instantaneous Doppler frequency of the target can be expressed as [25]

$$\tau_p = \frac{\sqrt{R_R^2 + R_L^2 + 2R_R R_L \sin \theta_R} + R_R}{c},$$

$$f_p = \frac{2v}{\lambda} \cos(\phi) \cos\left(\frac{\beta}{2}\right). \qquad (3)$$

Considering an observation scene consisting of K scattering points, a generalized expression of the surveillance signal can be written as

$$y(t) = \sum_{k=1}^{K} A_k x_T(t - \tau_k) \exp(j2\pi f_k t) + n(t), \qquad (4)$$

where A_k, f_k, and τ_k are the complex amplitude, the bistatic time delay, and the Doppler frequency shift of the kth scattering point. To simplify the analysis, the thermal fluctuations $n_r(t)$ and $n(t)$ are neglected, A_r is set to 1, and τ_L is set to 0. Then, the surveillance signal can be rewritten as

$$y(t) = \sum_{k=1}^{K} A_k x(t - \tau_k) \exp(j2\pi f_k t). \qquad (5)$$

It should be noted that this simplification has no significant impact. For example, the condition $\tau_L = 0$ can be satisfied as long as the reference signal is added a corresponding time delay.

3. CS-Based Processing for PR

In this section, we present our investigation of the following CS-based method to form a range-Doppler map with passive radar data.

3.1. Signal Preprocessing. The surveillance signal and the reference signal first need to be divided into multiple short segments [26], as shown in Figure 2. It means that the Doppler frequency change within a segment interval is ignored.
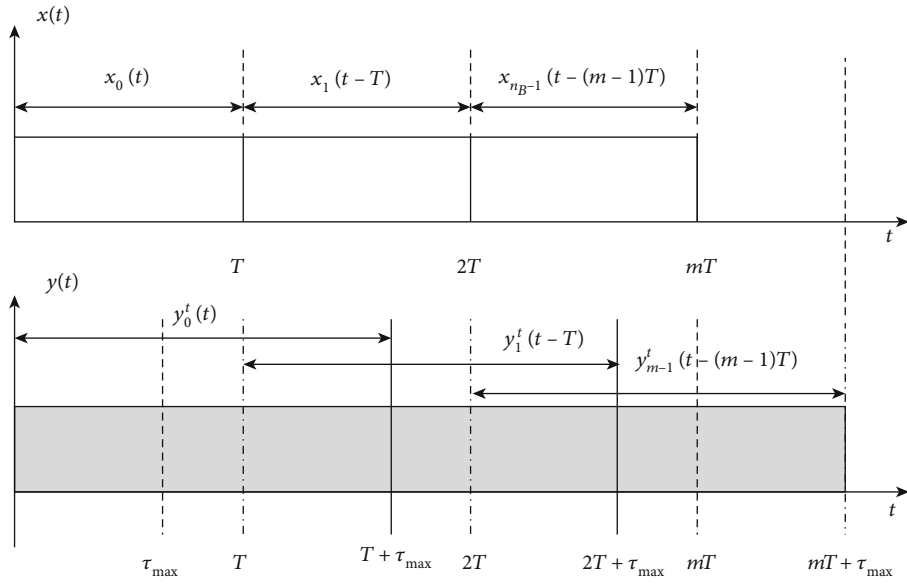
Figure 2: Signal segmentation.
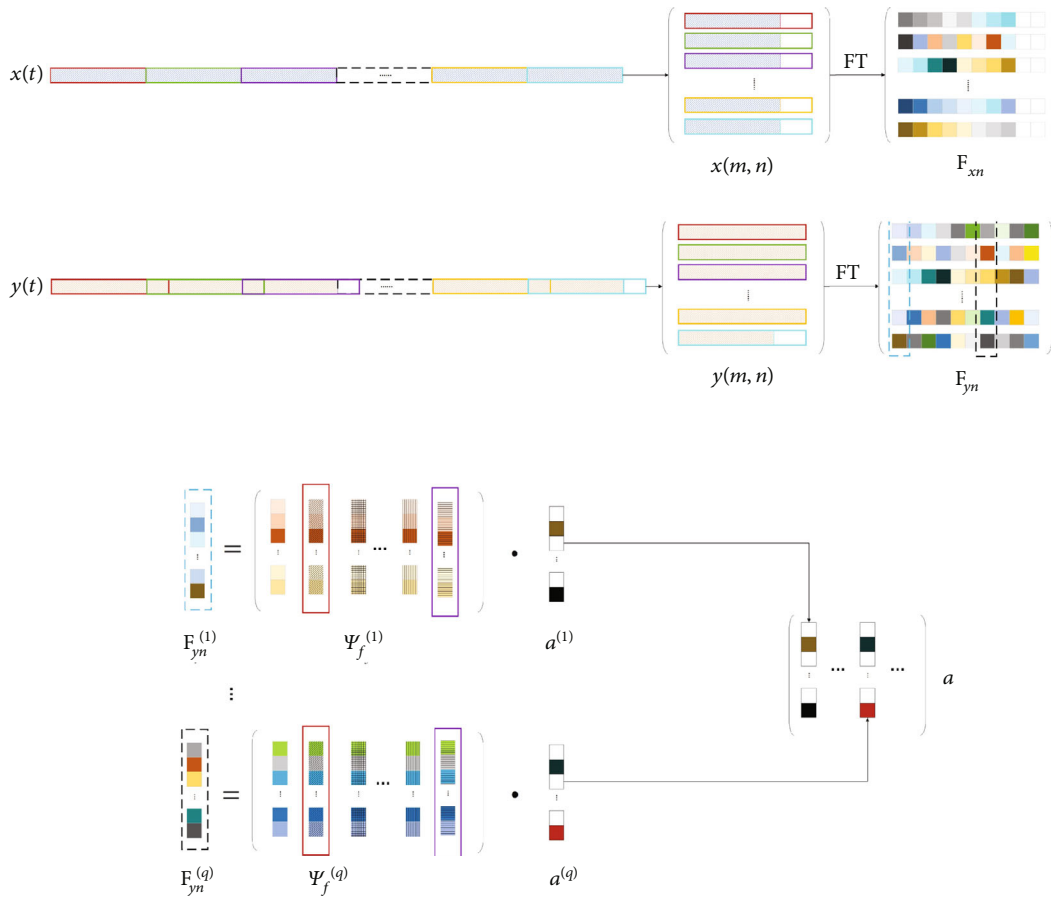
Figure 3: MMV model for passive radar.

Input: surveillance signal $F_{yn\_sub} \in \mathbb{C}^{M_{sub} \times N}$
        dictionary $\Psi_{f\_sub} =$
        $[\Psi_{f_{sub}}^{(1)}, \cdots, \Psi_{f_{sub}}^{(l)}, \cdots, \Psi_{f_{sub}}^{(N)}] \in \mathbb{C}^{M_{sub} \times M_{grid} N}$
        sparsity K
        residual threshold Th_1

Initialize: iteration count $k = 0$
        residual matrix $R = F_{yn\_sub}$
        estimate support collection $\Lambda = \varnothing$
        coefficient matrix $a = 0_{M_{grid} \times N}$

While $k \le K$ or $\|R^k\|_F > Th_1$ do $k = k + 1$
        (Identification)
        $v_k \longleftarrow \text{argmax}(\sum_{l=1}^{N} |(\Psi_{f\_sub}^{(l)})^H F_{yn\_sub}^{(l)}|)$, where v_k is
        the column index of the largest element in
        $\Psi_{f_{sub}}^{(l)}{}^H F_{yn\_sub}^{(l)}$
        (Update Index Support) $\Lambda^k = \Lambda^{k-1} \cup v_k$;
        (Estimation) $\hat{w}_{\Lambda^k} = \Psi_{f\_sub}^{(l)}[\Lambda^k]^\dagger F_{yn\_sub}^{(l)}$, where $(\cdot)^\dagger$
        represents the pseudo-inverse of the matrix, $a(\Lambda^k, n) =$
        $\hat{w}_{\Lambda^k}$,
        (Update Residual)
        $(R^k)^{(l)} = F_{yn\_sub}^{(l)} - \Psi_{f_{sub}}^{(l)}[\Lambda^k]\hat{w}_{\Lambda^k}$.

Output   coefficient matrix a
End

ALGORITHM 1: Pseudocode of modified MMV-OMP algorithm for Doppler reconstruction.

Input: surveillance signal after Doppler processing
        $(F_{ym\_sub})^T \in \mathbb{C}^{N_{sub} \times M_{grid}}$
        dictionary $\Psi_{\tau\_sub} \in \mathbb{C}^{N_{sub} \times N_{grid}}$
        sparsity Q
        residual threshold Th_2

Initialize: iteration count $q = 0$
        residual matrix $r = (F_{ym\_sub})^T$
        estimate support collection $L = 0_{Q \times M_{grid}}$
        coefficient matrix $\alpha' = 0_{N_{grid} \times M_{grid}}$

While $q \le Q$ or $\|r^k\|_F > Th_2$ do $q = q + 1$
        (Identification) $[u_q, b_q] \longleftarrow \text{arg max } |\Psi_{\tau\_sub}{}^H r|$, where u_q
        and b_q are the row index and column index of the largest
        element in $|\Psi_{\tau\_sub}{}^H r|$;
        (Update Index Support) $L(q, b_q) = u_q$;
        $P = nonzeros\{L[b_q]\}$, where $nonzeros\{\bullet\}$ represents
        non-zero elements in the vector;
        (Estimation) $\gamma = \Psi_{\tau\_sub}[P]^\dagger F_{ym\_sub}[b_q]$, where $(\cdot)^\dagger$
        represents the pseudo-inverse of the matrix, $\alpha'[P, b_q] = \gamma$;
        (Update Residual) $r[b_q] = F_{ym\_sub}[b_q] - \Psi_{\tau\_sub}[P]\gamma$.

Output   coefficient matrix $\alpha = (\alpha')^T$
End

ALGORITHM 2: Pseudocode of EOMP algorithm for range reconstruction.
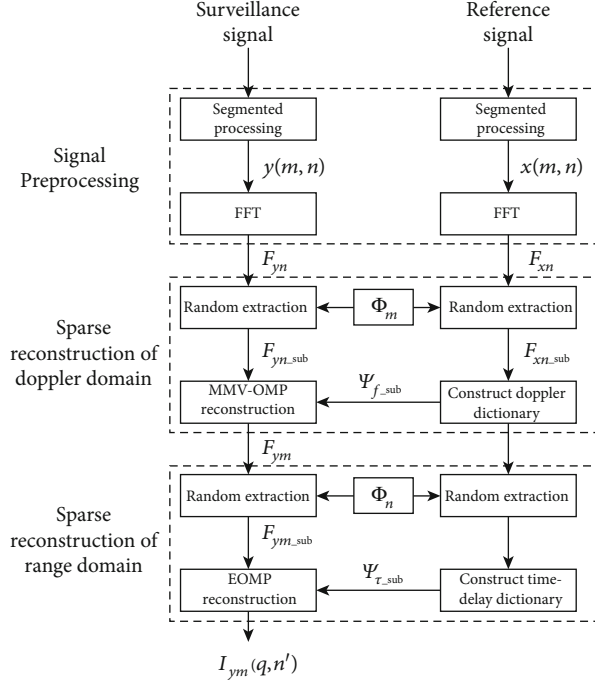
FIGURE 4: Flowchart of the proposed method.

TABLE 1: CS-PR method comparison.

Methods	Original method	Feng et al.'s method	Proposed method
Signal model	$rD : \mathbf{y}_{\text{vec}} = \mathbf{\Psi}_{rD}\mathbf{\alpha}_{\text{vec}}$	$D : \mathbf{y}_r = \mathbf{\Psi}_D \mathbf{\alpha}$	$D : \mathbf{F}_{yn\_\text{sub}}{}^{(l)} = \mathbf{\Psi}_{f\_\text{sub}}^{(l)}\mathbf{a}^{(l)}$ $r : \mathbf{F}_{ym\_\text{sub}} = \mathbf{\Psi}_{\tau\_\text{sub}}\mathbf{\alpha}$
Dictionary size	$\mathbf{\Psi}_{\mathbf{rD}}$ $\left(M_{\text{sub}}N_{\text{sub}} \times M_{\text{grid}}N_{\text{grid}}\right)$	$\mathbf{\Psi}_D \left(M_{\text{sub}} \times M_{\text{grid}}\right)$	$D : \mathbf{\Psi}_{f\_\text{sub}}^{(l)} \left(M_{\text{sub}} \times M_{\text{grid}}\right)$ $r : \mathbf{\Psi}_{\tau\_\text{sub}} \left(N_{\text{sub}} \times N_{\text{grid}}\right)$

Let $t = \tau + mT$ (where T is the segment interval); t can be called slow time. τ represents time delay, which is called fast time. Now, the surveillance signal is expressed as follows:

$$
\begin{aligned}
y(\tau) &= \sum_{k=1}^{K} A_k x(\tau - \tau_k + mT) \exp\left(j2\pi f_k \tau\right) \exp\left(j2\pi f_k mT\right) \\
&\approx \sum_{k=1}^{K} A_k x(\tau - \tau_k + mT) \exp\left(j2\pi f_k mT\right),
\end{aligned}
$$
(6)

where $m = 0, 1, \cdots, M - 1$ is the slow time index and M denotes the number of segments. The two-dimensional discrete form of (6) can be expressed as

$$
y(m, n) = \sum_{k=1}^{K} A_k x(n\Delta_\tau - n_k\Delta_\tau + mT) \exp\left(j2\pi f_k mT\right), \quad (7)
$$

where Δ_τ is the sampling time interval and $\tau_k = n_k\Delta_\tau$. $n = 0, 1, \cdots, N - 1$ is the fast time index, and N denotes the sample number of each segment. Similarly, the discrete representation of the reference signal is expressed as

$$
x(m, n) = x(n\Delta_\tau + mT). \tag{8}
$$

3.2. Sparse Reconstruction of Doppler Domain.
By taking an FFT of $x(m, n)$ with respect to n, we have

$$
F_{xn}(m, l) = \sum_{n=0}^{N-1} x(m, n) \exp\left(-j2\pi l\Delta f_n n\Delta_\tau\right) = \sum_{n=0}^{N-1} x(m, n) \exp\left(-jl\frac{2\pi}{N}n\right),
$$
(9)

where $\Delta f_n = 1/(N\Delta_\tau)$ is the range-frequency bin size and l is the range-frequency index. Similarly, after performing an FFT of $y(m, n)$ with respect to n, we can obtain

$$
\begin{aligned}
F_{yn}(m, l) &= \sum_{n=0}^{N-1} y(m, n) \exp\left(-j2\pi l\Delta f_n n\Delta_\tau\right) \\
&= \sum_{k=1}^{K} A_k F_{xn}(m, l) \exp\left(-j2\pi l\Delta f_n n_k\Delta_\tau\right) \exp\left(j2\pi f_k mT\right) \\
&= \sum_{k=1}^{K} A_k F_{xn}(m, l) \exp\left(-jl\frac{2\pi}{N}n_k\right) \exp\left(j2\pi f_k mT\right).
\end{aligned}
$$
(10)
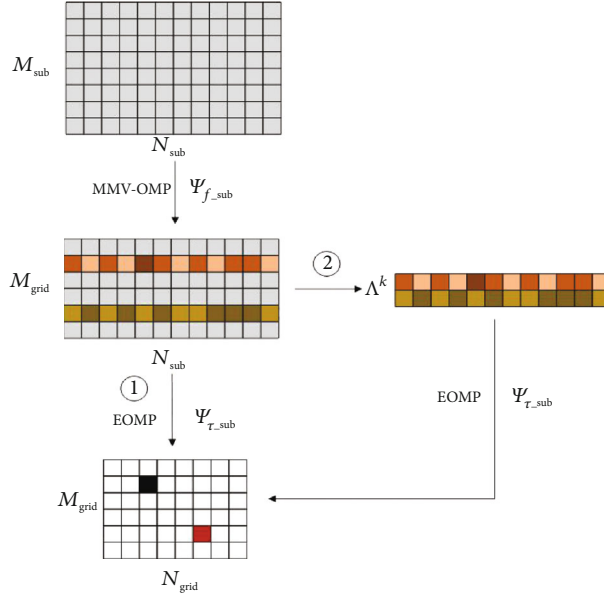
FIGURE 5: The diagram of signal reconstruction.

In order to make a sparse representation of the range-frequency bins, a Doppler dictionary $\mathbf{\Psi}_f^{(l)} \in \mathbb{C}^{M \times M_{\text{grid}}}$ is constructed as

$$\mathbf{\Psi}_f^{(l)} = \left[\psi_1^{(l)} \cdots \psi_q^{(l)} \cdots \psi_{M_{\text{grid}}}^{(l)} \right],$$

$$\psi_q^{(l)} = F_{xn}(m, l) \exp \left(j \frac{2\pi}{M_{\text{grid}}} qm \right), \quad m = 0, 1, \cdots, M - 1,$$

(11)

where M_{grid} denotes the number of the Doppler grid. And then, (10) can be expressed as

$$F_{yn}(m, l) = \sum_{k=1}^{K} A_k F_{xn}(m, l) \exp \left(-jl \frac{2\pi}{N} n_k \right) \exp \left(j \frac{2\pi}{M_{\text{grid}}} q_k m \right),$$

(12)

where $q_k = f_k / \Delta f_q$ and $\Delta f_q = 1/(M_{\text{grid}} T)$ represent the size of the Doppler grid. From this, we may know that $M_{\text{grid}} = M$ makes the same resolution level as the classical method, $\Delta f_q = \Delta f = 1/(MT)$, and the larger M_{grid} can generate higher Doppler resolution.

For each range-frequency bin l, (12) can be rewritten as

$$\mathbf{F}_{yn}^{(l)} = \mathbf{\Psi}_f^{(l)} \mathbf{a}^{(l)},$$

(13)

where

$$\mathbf{F}_{yn}^{(l)} = \left[F_{yn}(0, l) \quad \cdots \quad \cdots \quad F_{yn}(M - 1, l) \right]^T,$$

$$\mathbf{a}^{(l)} = \left[a_1 \quad \cdots \quad a_q \quad \cdots \quad a_{M_{\text{grid}}} \right]^T.$$

(14)

TABLE 2: Frame structure of DTTB signal.

	Frame head	System information	Data
DTTB	945 symbols	36 symbols	3744 symbols

TABLE 3: System parameter.

Parameters	Symbol	Value
Carrier frequency	f_c	674 MHz
Bandwidth	B	7.56 MHz
Sample frequency	f_s	10 MHz
Number of segments	M	512
Sample number	N	256
Number of range grid	M_{grid}	50
Number of Doppler grid	N_{grid}	20

There are K nonzero elements in $\mathbf{a}^{(l)}$, when the scattering points make different Doppler frequency shifts. And then, the positions and amplitude values are, respectively,

$$q_k = \frac{f_k}{\Delta f_q}, \quad k = 1, 2, \cdots, K,$$

$$a_{q_k} = A_k \exp \left(-jl \frac{2\pi}{N} n_k \right), \quad k = 1, 2, \cdots, K.$$

(15)

If there are the same Doppler frequencies, then the number of nonzero elements becomes smaller. In turn, the complex amplitude is the sum of all coefficients related to the same Doppler.

According to the CS theory, the sparsity of the signal ensures the feasibility of reducing the amount of data. The

TABLE 4: Target scenario parameters.

Parameters	Target 1	Target 2		
		Scatterer point 1	Scatterer point 2	Scatterer point 3
Target location index	10	26	28	30
Doppler bin	4	-4	-4	-4
Amplitude	1	0.7	0.9	1
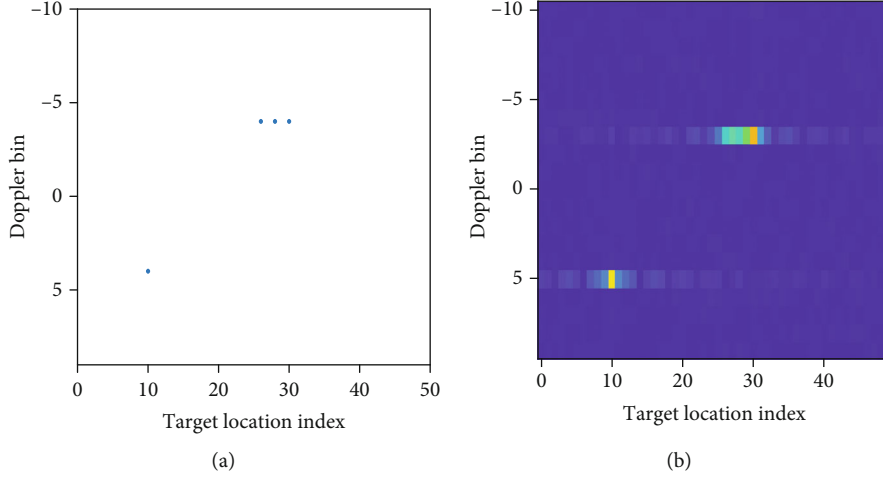
FIGURE 6: Result obtained by CAF: (a) the original scene; (b) RD map.

sensing matrix obtained by multiplying the measurement matrix $\mathbf{\Phi}$ and the Doppler dictionary (sparse basis matrix) $\mathbf{\Psi}_f^{(l)}$ needs to satisfy the restricted isometric property (RIP). The commonly used measurement matrices are the random Gaussian matrix and partial random unit matrix. In order to facilitate the practical operation, we adopt the method of random extraction of data, which can be expressed as

$$\mathbf{F}_{yn\_\mathrm{sub}}^{(l)} = \mathbf{\Phi}_m \mathbf{F}_{yn}^{(l)} = \mathbf{\Phi}_m \mathbf{\Psi}_f^{(l)} \mathbf{a}^{(l)} = \mathbf{\Psi}_{f\_\mathrm{sub}}^{(l)} \mathbf{a}^{(l)}, \qquad (16)$$

where $\mathbf{\Phi}_m \in \mathbb{R}^{M_\mathrm{sub} \times M}$ is a partial random unit matrix. In order to obtain the coefficient vector $\mathbf{a}^{(l)}$, (16) needs to be solved. Due to the sparseness of the coefficient vector, it is equivalent to solving the following minimum norm problem:

$$\mathbf{a}\wedge^{(l)} = \arg\min \left\| \mathbf{a}^{(l)} \right\|_0 \mathrm{s.t.} \mathbf{F}_{yn\_\mathrm{sub}}^{(l)} = \mathbf{\Psi}_{f\_\mathrm{sub}}^{(l)} \mathbf{a}^{(l)}. \qquad (17)$$

There are many methods to solve (17). The greedy algorithm is widely used in practical application because of its excellent geometric interpretation, good reconstruction effect, and fast reconstruction speed. The most representative greedy algorithm is the OMP algorithm. Considering the consistency of signal models of multiple range-frequency bins, the same operation can be carried out for each range-frequency bin. That is to say, it can solve $\mathbf{a}^{(l)}$ through OMP under the single measurement vector (SMV) model according to $\mathbf{F}_{yn\_\mathrm{sub}}^{(l)}$ and $\mathbf{\Psi}_{f\_\mathrm{sub}}^{(l)}$, respectively.

We note that 13 is independent of l. It is found that the positions of nonzero elements of $\mathbf{a}^{(l)}$ in multiple range-frequency bins are the same; that is, the support set of each sparse coefficient vector is the same. This feature means that it can be considered an MMV model (see Figure 3), and we can use the joint sparsity to improve reconstruction performance. However, the existing OMP algorithm under the MMV model cannot be directly applied. Considering that $\mathbf{\Psi}_{f\_\mathrm{sub}}^{(l)}$ is different for passive radar data, a modified version of MMV-OMP is proposed here. The pseudocode is shown in Algorithm 1. In order to facilitate the derivation, the matrix $\mathbf{a} = [\mathbf{a}\wedge^{(1)}, \cdots, \mathbf{a}\wedge^{(l)}, \cdots, \mathbf{a}\wedge^{(N)}]$ which is obtained by the modified MMV-OMP algorithm will be rewritten as $F_{ym}(q, l)$.

3.3. Sparse Reconstruction of Range Domain. After Doppler reconstruction, the two-dimensional data can be expressed as

$$F_{ym}(q, l) = \begin{cases} \sum_k A_k \exp\left(-jl\dfrac{2\pi}{N} n_k\right), & q = q_k, \\ 0, & q \neq q_k. \end{cases} \qquad (18)$$

It can be seen from (18) that the signal is independent of Doppler frequency bin q. Therefore, the same time-delay dictionary $\mathbf{\Psi}_\tau \in \mathbb{C}^{N \times N_\mathrm{grid}}$ can be constructed for each Doppler bin, which can be expressed as

$$\mathbf{\Psi}_\tau = \begin{bmatrix} \psi_1 & \cdots & \psi_n' & \cdots & \psi_{N_\mathrm{grid}} \end{bmatrix},$$

$$\psi_n' = \exp\left(-j\dfrac{2\pi}{N_{grid}} n'l\right), \quad l = 0, 1, \cdots, N-1, \qquad (19)$$
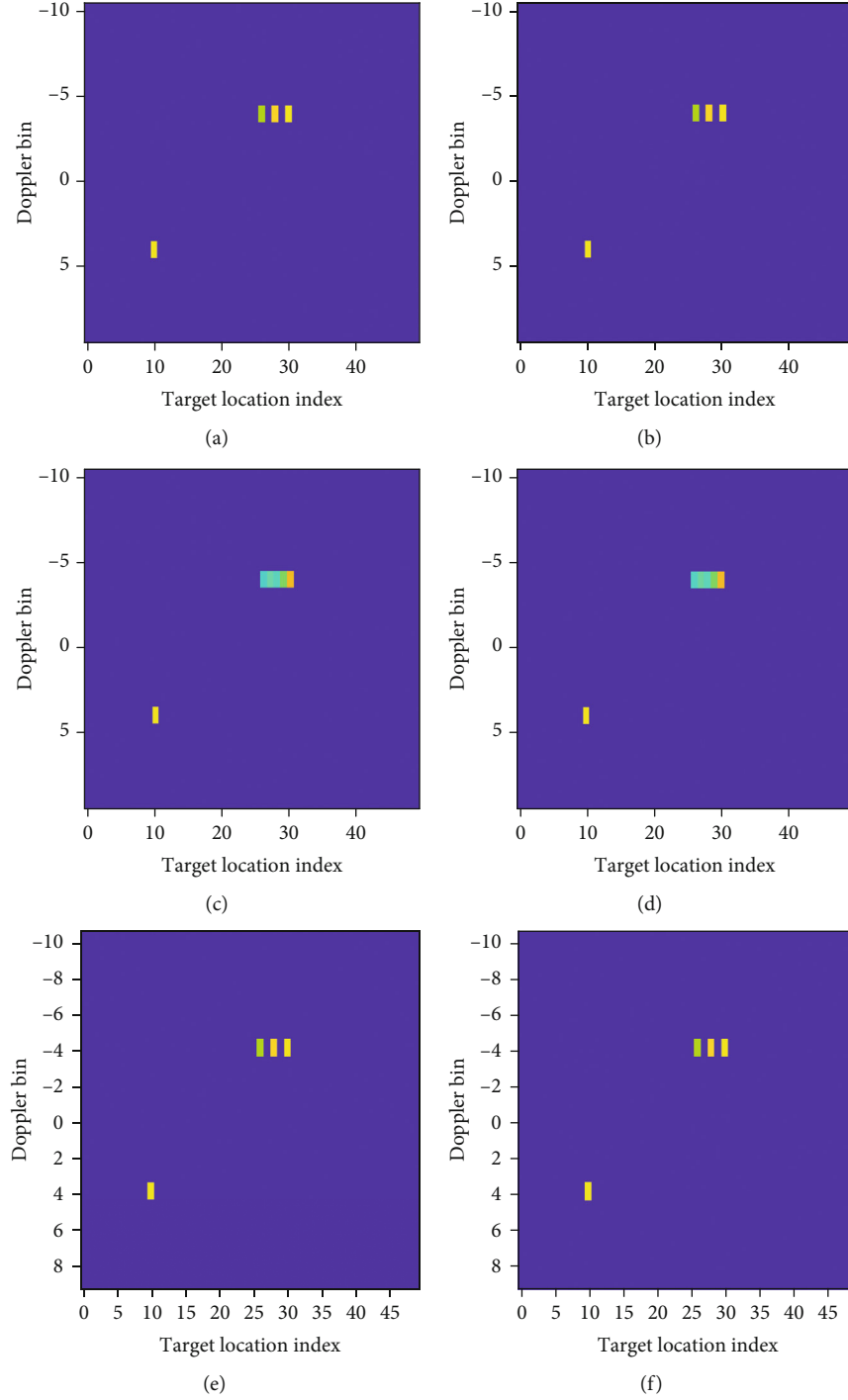
FIGURE 7: RD maps obtained by (a, b) original CS-PR method, $K = 4$; (c, d) Feng et al.'s method, $K = 4$; (e, f) proposed method, $K = Q = 4$; (a, c, e) with full samples; (b, d, f) with partial samples (1/4).

where N_{grid} denotes the number of time-delay (range) grids. So, the signal of qth Doppler bin can be expressed as

$$\mathbf{F}_{ym}{}^{(q)} = \boldsymbol{\Psi}_\tau \boldsymbol{\alpha}^{(q)}, \tag{20}$$

where

$$\mathbf{F}_{ym}{}^{(q)} = \left[F_{ym}(q,0) \quad \cdots \quad \cdots \quad F_{ym}(q,N-1) \right]^T,$$
$$\boldsymbol{\alpha}^{(q)} = \left[\alpha_1 \quad \cdots \quad \alpha_{n'} \quad \cdots \quad \alpha_{N_{grid}} \right]^T. \tag{21}$$

In whole range-Doppler plane, there are K nonzero elements ($K \ll M_{grid} N_{grid}$), and the positions and amplitude values are, respectively,

$$n'_k = \frac{\tau_k}{\Delta_\tau{}'}, \quad k = 1, 2, \cdots, K, \tag{22}$$
$$\alpha_{n'_k} = A_k, \quad k = 1, 2, \cdots, K,$$

where $\Delta_\tau{}' = \Delta_\tau(N/N_{\text{grid}})$. It was obvious that $N_{\text{grid}} = N$ makes the same or better resolution level as the classical method; the larger N_{grid} can generate finer range resolution.

Similarly, the signal is randomly extracted, and the observation equation can be expressed as

$$\mathbf{F}_{ym\_\text{sub}}{}^{(q)} = \mathbf{\Phi}_n \mathbf{F}_{ym}{}^{(q)} = \mathbf{\Phi}_n \mathbf{\Psi}_\tau \boldsymbol{\alpha}^{(q)} = \mathbf{\Psi}_{\tau\_\text{sub}} \boldsymbol{\alpha}^{(q)}, \tag{23}$$

where $\mathbf{\Phi}_n \in \mathbb{R}^{N_{\text{sub}} \times N}$ is a partial random unit matrix. It is important to note that the randomness has a constraint. For range-frequency bins, the frequency range is determined by the sampling rate f_s, which usually satisfies $f_s \geq B$ (B is the signal bandwidth). This means that some range-frequency bins contain invalid information. Therefore, the random extraction only considers the effective part of the signal bandwidth.

The coefficient vector $\boldsymbol{\alpha}^{(q)}$ can be solved as follows:

$$\alpha \wedge^{(q)} = \arg \min \left\| \boldsymbol{\alpha}^{(q)} \right\|_0 \text{ s.t. } \mathbf{F}_{ym\_\text{sub}}{}^{(q)} = \mathbf{\Psi}_{\tau\_\text{sub}} \boldsymbol{\alpha}^{(q)}. \tag{24}$$

To effectively solve the problem, we utilize a sparse matrix recovery algorithm. Its pseudocode is described in Algorithm 2, which can be seen as another way of using EOMP. Similarly, the matrix $\boldsymbol{\alpha} = [\boldsymbol{\alpha}\wedge^{(1)}, \cdots, \boldsymbol{\alpha}\wedge^{(q)}, \cdots, \boldsymbol{\alpha}\wedge^{(M_{\text{grid}})}]^T$ is rewritten as $\mathbf{I}_{ym}(q, n')$.

The signal processed by the reconstruction algorithm can be expressed as

$$\mathbf{I}_{ym}\left(q, n'\right) = \begin{cases} A_k, & q = q_k, n' = n'_k, \\ 0, & \text{others.} \end{cases} \tag{25}$$

Obviously, (25) is the distribution of the scattering coefficient in the time-delay Doppler grid. Therefore, based on the above analysis, we can see that the range-Doppler map for passive radar can be obtained by using the proposed CS-based processing method.

3.4. Summary of Method Flow.

In order to intuitively show the processing technique, the flowchart of the proposed method is shown in Figure 4. The steps are briefly summarized as follows.

Step 1. After segmentation of reference signal and surveillance signal, perform FFT on $x(m, n)$ and $y(m, n)$ in the fast time n direction, respectively.

Step 2. According to $\mathbf{\Phi}_m$, construct Doppler dictionary $\mathbf{\Psi}_{f\_\text{sub}}$ by using the reference signal of the range-frequency domain and observation vector $\mathbf{F}_{yn\_\text{sub}}$ by using the surveillance signal of the range-frequency domain, and then perform

TABLE 5: Run times for different CS-PR methods.

Methods	Full samples	Partial samples
Original method	3.365 s	0.615 s
Feng et al.'s method	0.069 s	0.054 s
Proposed method	0.161 s	0.123 s

Algorithm 1 (MMV-OMP) to reconstruct Doppler frequency distribution.

Step 3. According to $\mathbf{\Phi}_n$, construct time-delay dictionary $\mathbf{\Psi}_{\tau\_\text{sub}}$ and observation matrix $\mathbf{F}_{ym\_\text{sub}}$, and then perform Algorithm 2 (EOMP) to reconstruct range-Doppler map.

There are two points which remain to be explained. In Step 2, we actually only need to process N_{sub} range-frequency bins related to $\mathbf{\Phi}_n$. In addition, the grid sizes $(\Delta f_q, \Delta_\tau{}')$ and grid numbers $(M_{\text{grid}}, N_{\text{grid}})$ of dictionaries $(\mathbf{\Psi}_{f\_\text{sub}}, \mathbf{\Psi}_{\tau\_\text{sub}})$ can be set in accordance with the actual condition.

3.5. Comparative Analysis.

We compare the proposed method with the time-delay/Doppler combination dictionary-based CS-PR method (named as original method here) presented in [17, 18] and Feng et al.'s method presented in [19]. Table 1 demonstrates the signal model, dictionary size of the original method, Feng et al.'s method, and the proposed method, respectively.

The major difference among these three methods lies in the signal model and method implementation. In the original method, the reflectivity map matrix is reconstructed by a cascade of 1-D CS reconstruction. All measurements are stacked into a single observation vector $\mathbf{y}_{\text{vec}}$, the state of each time-delay/Doppler combination is stacked into the state vector $\boldsymbol{\alpha}_{\text{vec}}$, and the dictionary $\mathbf{\Psi}_{rD}$ by discretizing the delay Doppler plane on a grid takes up the most memory. The size is $M_{\text{sub}} N_{\text{sub}} \times M_{\text{grid}} N_{\text{grid}}$, which leads to the memory occupation being too large and the computational burden being huge.

In both Feng et al.'s method and the proposed algorithm, the range reconstruction and the Doppler reconstruction are separately completed, which means that the dictionaries $\mathbf{\Psi}_D$, $\mathbf{\Psi}_{f\_\text{sub}}^{(l)}$, $\mathbf{\Psi}_{\tau\_\text{sub}}$ have much smaller size. The difference between the two methods is that the former only uses CS to estimate Doppler frequency while the latter uses CS in both directions, even though the use of EOMP is sameness. We can obtain high-resolution capability in the range coordinate. This is just what Feng et al.'s method does not have.

In addition, we can consider the fact that the input of EOMP is the output of MMV-OMP, which records the support set information. Therefore, the input data size of EOMP can be adjusted according to the size of the support set. As shown in Figure 5, there are two execution modes to obtain the range-Doppler map. The second execution mode can further reduce the computational burden due to the small amount of data.
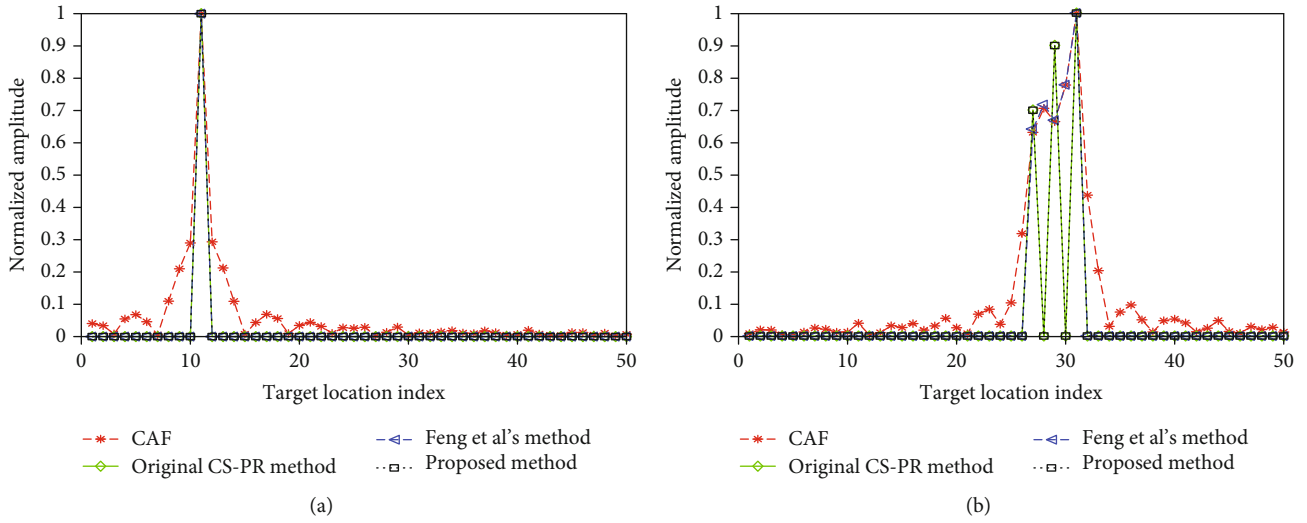
FIGURE 8: Range profiles obtained by different methods: (a) target 1; (b) target 2.
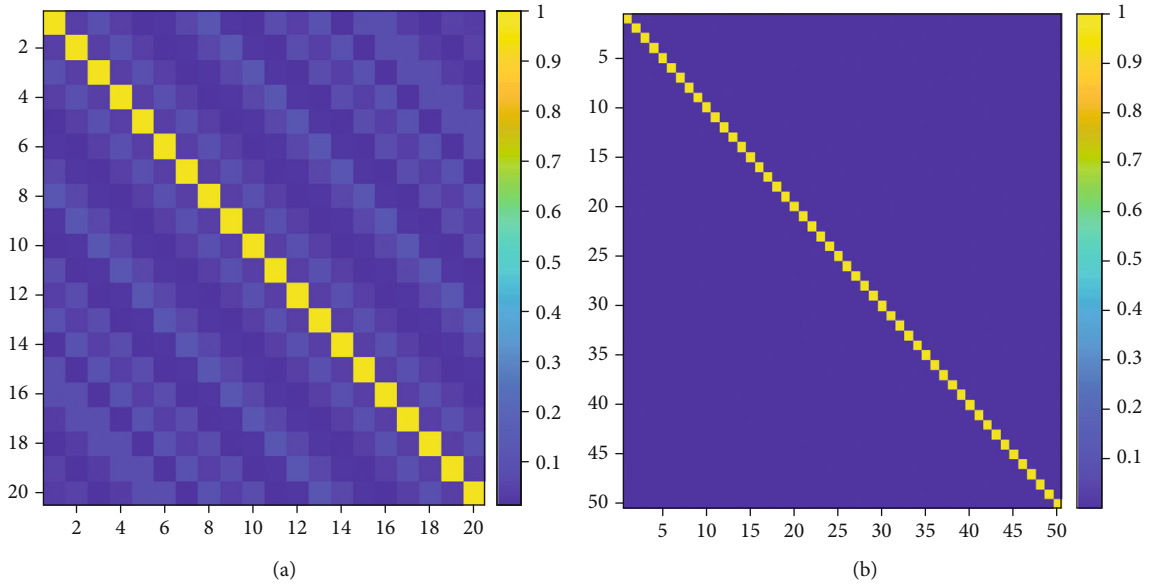
FIGURE 9: Gram matrices of original dictionaries ($\Delta f_q = \Delta f$, $\Delta_\tau' = \Delta_\tau$): (a) $\mathbf{G_f}$; (b) $\mathbf{G_\tau}$.
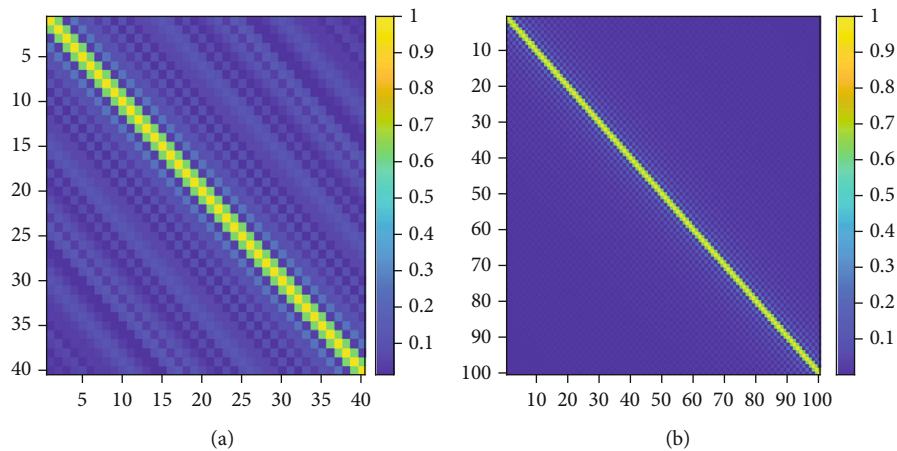
FIGURE 10: Gram matrices of refined dictionaries ($\Delta f_q = \Delta f/2$, $\Delta_\tau' = \Delta_\tau/2$): (a) $\mathbf{G_f}$; (b) $\mathbf{G_\tau}$.

(a) (b)
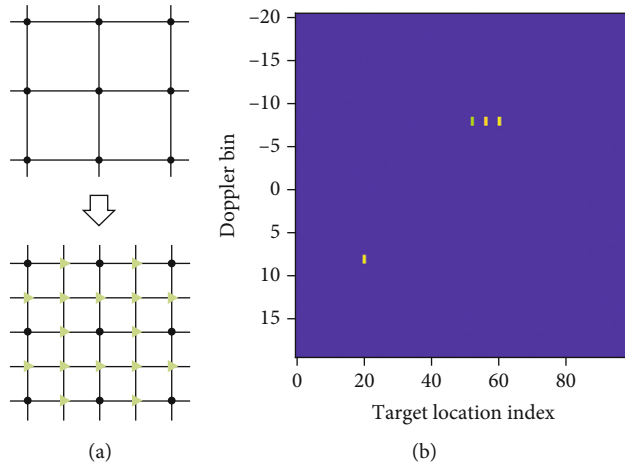
FIGURE 11: Result obtained by the proposed method with fine grid: (a) refining grid with 2×2; (b) RD map.

(a) (b)
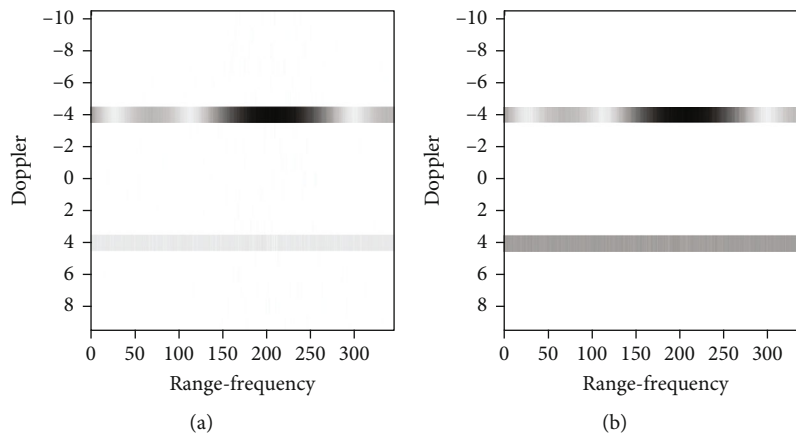
FIGURE 12: Results of sparse reconstruction of Doppler domain (SNR = 10 dB) by (a) SMV-OMP and (b) MMV-OMP.

4. Experimental Results

In this section, we present experimental results with simulated data and real data. The effectiveness of the proposed processing method is demonstrated.

4.1. Simulation Data. We have conducted numerical experiments to investigate the performance of the proposed processing method. A digital television terrestrial broadcasting (DTTB) signal is simulated. The frame structure of signal is shown in Table 2, which includes frame header (945 symbols) and frame body (3780 symbols).

The parameters used in the simulation experiment are shown in Tables 3 and 4. In the observation scene, the reflection mechanisms are assumed to be a point-like target (target 1) and a line-like target (target 2). Target 2 consists of three scatterer points, which are located in the same Doppler bin and become neighbors in the range direction.

At the beginning, the result obtained by the classical CAF is shown in Figure 6. As can be seen from Figures 6(a) and 6(b), the CAF can achieve the target scene recovery, but the reconstructed map is out of clarity due to the large sidelobes.

On the contrary, the CS-PR methods can be used to remove the sidelobes. The processing results by using the CS-PR methods listed in Table 1 are shown in Figure 7, which are the RD maps obtained by using full samples and partial samples from the same scenario. For the partial sample case, in order to ensure the same amount of data, the original method and the proposed method perform one-half data undersampling processing in the range direction and Doppler direction, respectively, while Feng et al.'s method only performs one-quarter data undersampling processing in the Doppler direction.

Figure 7 demonstrates that these processing methods based on CS can effectively suppress the sidelobe. The run times of different methods are summarized in Table 5. Apparently, the original CS-PR method uses much longer running time than the other two methods, due to the usage of the dictionary by the discretizing RD plane. In order to more intuitively observe the effect of different methods on the sidelobe suppression, the range profiles obtained by different methods are shown in Figure 8. From the figures, it shows that Feng et al.'s method cannot clearly identify the scatterer points in the range direction.

(a)

(b)
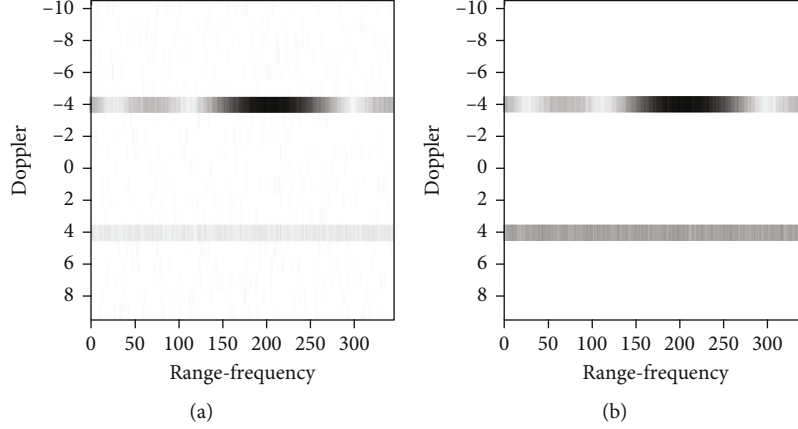
FIGURE 13: Results of sparse reconstruction of Doppler domain (SNR = 0 dB) by (a) SMV-OMP and (b) MMV-OMP.

(a)

(b)
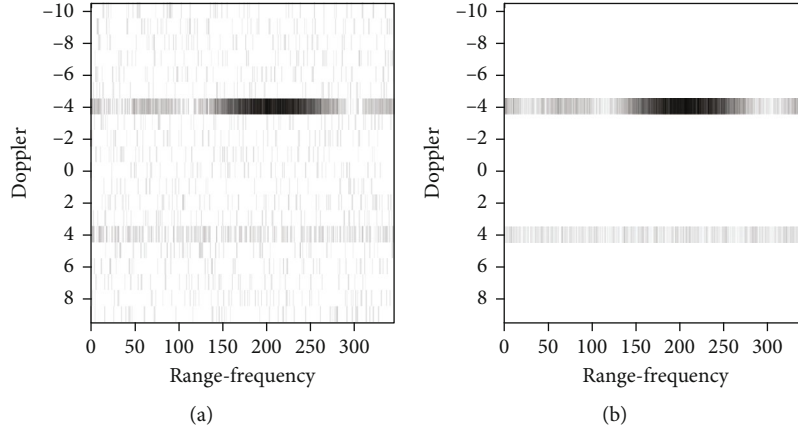
FIGURE 14: Results of sparse reconstruction of Doppler domain (SNR = −10 dB) by (a) SMV-OMP and (b) MMV-OMP.

It indicates that the range resolution of the method is limited, even though its run time is the least. Fortunately, the proposed processing method can reconstruct the target scene with a good performance both in accuracy and in computational efficiency.

It is known that the CS approach offers great potential for better resolution by using a finer dictionary. The grid sizes are set to be the same as CAF in previous experiments $(\Delta f_q = \Delta f, \Delta_\tau' = \Delta_\tau)$. Now, the Doppler grid and time-delay grid are only half the original size $(\Delta f_q = \Delta f/2, \Delta_\tau' = \Delta_\tau/2)$. As the grid is refined, the challenge is that the dictionary correlation will increase, which may lead to the performance degradation for CS. The Gram matrix is used to verify the dictionary coherence, and results are depicted in Figures 9 and 10. $\mathbf{G_f} = \mathbf{\Psi}_{f\_sub}^H \mathbf{\Psi}_{f\_sub}$ and $\mathbf{G_\tau} = \mathbf{\Psi}_{\tau\_sub}^H \mathbf{\Psi}_{\tau\_sub}$ represent Gram matrices in Doppler direction and range direction, respectively. As can be seen from Figures 9 and 10, each Gram matrix is close to the unit matrix. Due to this coherence characteristic, CS is able to produce superresolution radar images. Figure 11 shows the RD map obtained by the proposed method when the grid is refined. It is observed that the resolution improvement is achieved by utilizing a fine grid.

In order to demonstrate the advantages of OMP under the MMV model, we compare the Doppler reconstructions by SMV-OMP and MMV-OMP at different signal-to-noise ratios (SNR). The values of SNR are 10 dB, 0 dB, and −10 dB. The parameters of target 1 use the following settings. The location index is 10, the Doppler bin is 4, and the amplitude is 0.3. Other parameters remain unchanged. The results are shown in Figures 12–14. For SMV-OMP, there are a number of Doppler reconstruction errors, which will lead to insufficient energy accumulation in the range direction. Figure 15 shows the final RD maps when SNR is −10 dB. It can be found that target 1 is not visible in the RD map obtained by SMV-OMP. On the contrary, MMV-OMP has robust performance because it considers the information of multiple observations.

4.2. Real Data. We use real data from an DTV-based PR system to further verify the proposed method. The parameters used are shown in Table 6. We obtain results for CAF and the proposed method. The RD map obtained by the CAF approach is shown in Figure 16; Figures 17 and 18 show the results of CS reconstruction if 25% of the full data is used. In comparison to the CAF, the proposed method shows good performance owing to CS.
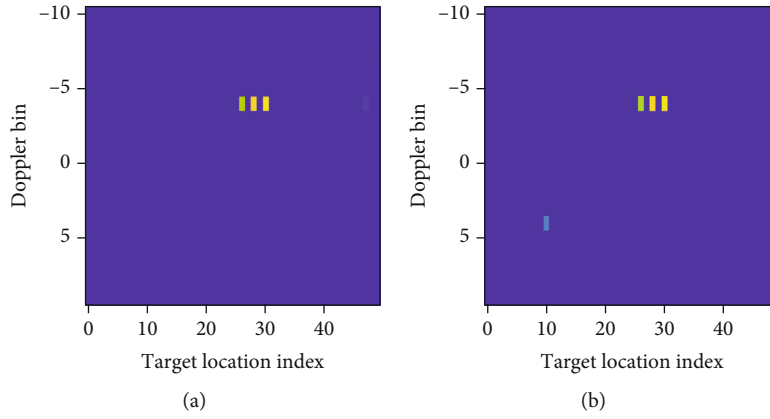
FIGURE 15: RD maps (SNR = −10 dB) obtained by (a) SMV-OMP and (b) MMV-OMP.

TABLE 6: System parameters.

Parameters	Symbol	Value
Carrier frequency	f_c	674 MHz
Bandwidth	B	7.56 MHz
Sample frequency	f_s	10 MHz
Number of segments	M	2048
Sample number	N	2048
Number of range grid	M_{grid}	50
Number of Doppler grid	N_{grid}	200
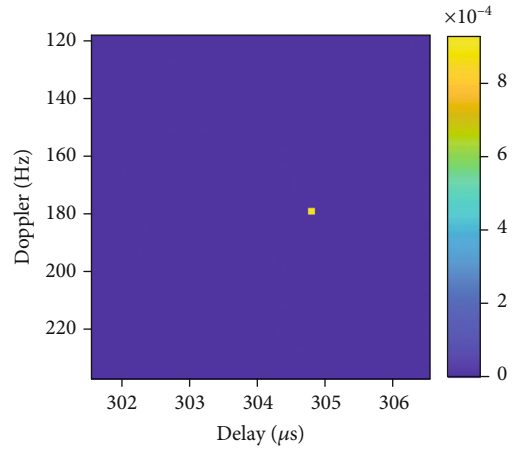
FIGURE 17: RD map obtained by the proposed method.
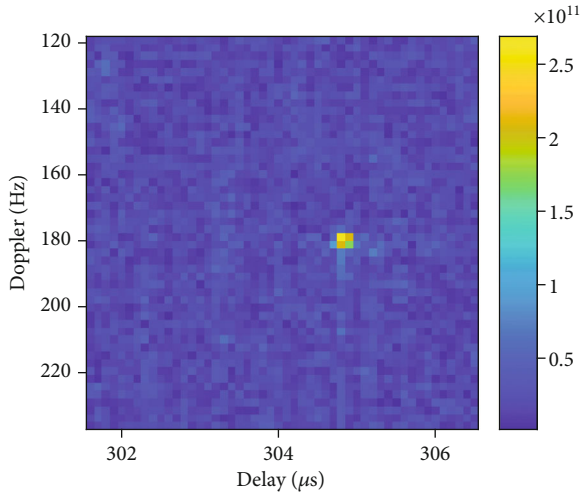
FIGURE 16: RD map obtained by CAF.
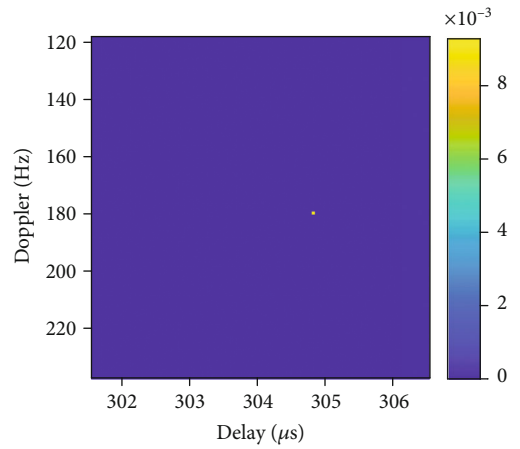
FIGURE 18: RD map obtained by the proposed method with fine grid.
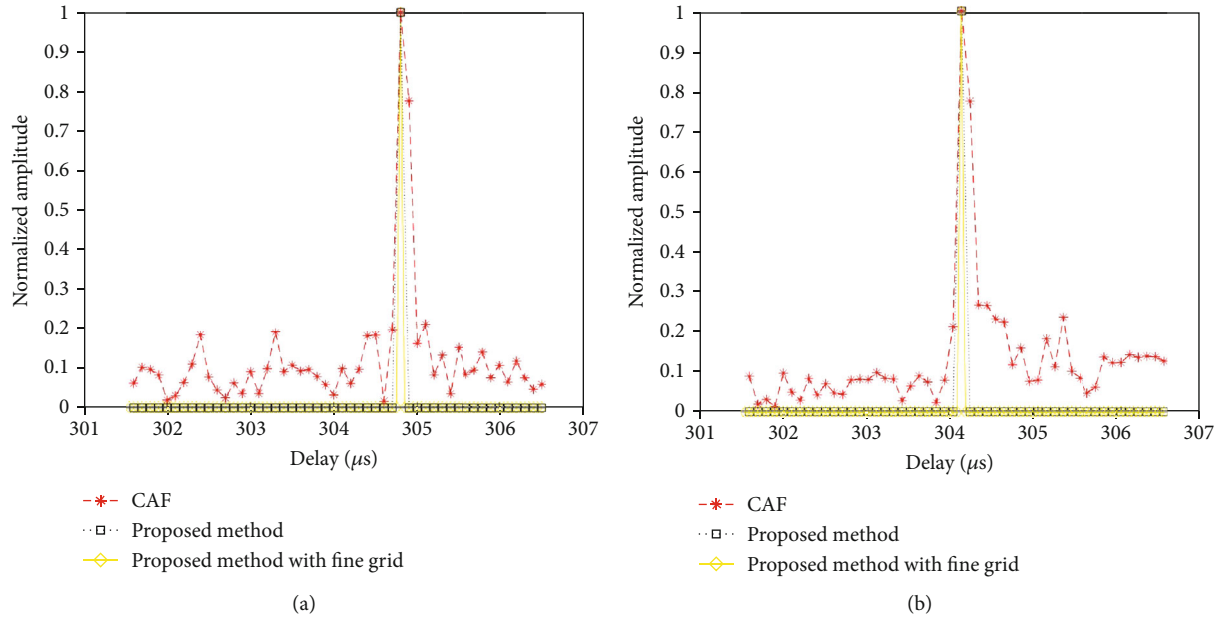
FIGURE 19: Profiles of target: (a) in range direction; (b) in Doppler direction.

For a more illustrative comparison, two cuts are made along the location of the target in the range and Doppler directions shown in Figures 19(a) and 19(b). The proposed method has a considerably lower sidelobe level than the CAF and indeed allows for improved range and Doppler resolutions by the fine gridding.

5. Conclusion

In this paper, the problem of compressed sensing-based range-Doppler processing for passive radar is investigated. In order to reduce the sidelobes and improve the resolution, we have proposed a novel CS-PR method, in which the modified MMV-OMP algorithm is used to perform sparse reconstruction of Doppler dimension, and then, the EOMP algorithm is used to perform sparse reconstruction of range dimension. Compared to previous CS-PR methods, we can achieve a high-quality reconstruction of the range-Doppler map of target scenario and do not suffer from the heavy computational burden. The effectiveness of the proposed method is verified by experiments with simulated data and real data. The improved resolution capability will be helpful to widen the extent of application.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] P. Howland, "Passive radar systems," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 105-106, 2005.

[2] A. Lauri, F. Colone, R. Cardinali, C. Bongioanni, and P. Lombardo, "Analysis and emulation of FM radio signals for passive radar," in *2007 IEEE Aerospace Conference*, pp. 1–10, Big Sky, MT, 2007.

[3] L. Wan, X. Kong, and F. Xia, "Joint range-Doppler-angle estimation for intelligent tracking of moving aerial targets," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1625–1636, 2018.

[4] D. K. P. Tan, H. Sun, Y. Lu, M. Lesturgie, and H. L. Chan, "Passive radar using global system for mobile communication signal: theory, implementation and measurements," *IEE Proceedings. Part F, Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 116–123, 2005.

[5] J. E. Palmer, H. A. Harms, S. J. Searle, and L. Davis, "DVB-T passive radar signal processing," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2116–2126, 2013.

[6] Z. Gao, R. Tao, Y. Ma, and T. Shao, "DVB-T signal cross-ambiguity functions improvement for passive radar," in *2006 CIE International Conference on Radar*, pp. 1–4, Shanghai, 2006.

[7] Y. D. Zhang, M. G. Amin, and B. Himed, "Structure-aware sparse reconstruction and applications to passive multi-static radar," *IEEE Aerospace and Electronic Systems Magazine*, vol. 32, no. 2, pp. 68–78, 2017.

[8] A. S. Tasdelen and H. Koymen, "Range resolution improvement in passive coherent location radar systems using multiple

FM radio channels," in *2006 IET Seminar on Waveform Diversity and Design in Communications, Radar and Sonar*, pp. 23–31, London, 2006.

[9] K. E. Olsen and K. Woodbridge, "Analysis of the performance of a multiband passive bistatic radar processing scheme," in *2010 International Waveform Diversity and Design Conference*, pp. 000142–000149, Niagara Falls, ON, 2010.

[10] K. E. Olsen and K. Woodbridge, "Multiband passive bistatic DVB-T radar range resolution improvements and implications," in *2012 13th International Radar Symposium*, pp. 28–31, Warsaw, 2012.

[11] Y. Feng, T. Shan, Z. Zhuo, and R. Tao, "The migration compensation methods for DTV based passive radar," in *2013 IEEE Radar Conference (RadarCon13)*, pp. 1–4, Ottawa, ON, 2013.

[12] K. E. Olsen and J. M. Christiansen, "Multiband passive bistatic radar coherent range and Doppler-walk compensation," in *2015 IEEE Radar Conference (RadarCon)*, pp. 0123–0126, Arlington, VA, 2015.

[13] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[14] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[15] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, 2009.

[16] J. Ender, "A brief review of compressive sensing applied to radar," in *2013 14th International Radar Symposium (IRS)*, pp. 3–16, Dresden, 2013.

[17] M. Wei, "Compressive sensing for passive surveillance radar using DAB signals," in *2014 International Radar Conference*, pp. 1–6, Lille, 2014.

[18] M. N. Tabassum, M. A. Hadi, and S. Alshebeili, "Compressive sensing based passive bistatic radar processing using time-domain complex data," in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 63–68, Abu Dhabi, United Arab, 2015.

[19] W. Feng, J. Friedt, G. Cherniak, and M. Sato, "Batch compressive sensing for passive radar range-Doppler map generation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 3090–3102, 2019.

[20] S. F. Cotter, B. D. Rao, Kjersti Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[21] R. Cardinali, F. Colone, C. Ferretti, and P. Lombardo, "Comparison of clutter and multipath cancellation techniques for passive radar," in *2007 IEEE Radar Conference*, pp. 469–474, Boston, MA, 2007.

[22] J. E. Palmer and S. J. Searle, "Evaluation of adaptive filter algorithms for clutter cancellation in passive bistatic radar," in *2012 IEEE Radar Conference*, pp. 0493–0498, Atlanta, GA, 2012.

[23] J. L. Garry, G. E. Smith, and C. J. Baker, "Direct signal suppression schemes for passive radar," in *2015 Signal Processing Symposium (SPSympo)*, pp. 1–5, Debe, 2015.

[24] T. Peto and R. Seller, "Time domain filter comparison in passive radar systems," in *2017 18th International Radar Symposium (IRS)*, pp. 1–10, Prague, 2017.

[25] A. Filip-Dhaubhadel and D. Shutin, "Long coherent integration in passive radar systems using super-resolution sparse Bayesian learning," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 1, pp. 554–572, 2020.

[26] C. Moscardini, D. Petri, A. Capria, M. Conti, M. Martorella, and F. Berizzi, "Batches algorithm for passive radar: a theoretical analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 2, pp. 1475–1487, 2015.

WILEY | Hindawi

Research Article

Coherent Target Direction-of-Arrival Estimation for Coprime Arrays: From Spatial Smoothing Perspective

Dongming Wu [ID], Fangzheng Liu, Zhihui Li, and Zhenzhong Han

National University of Defense Technology, Hefei 230000, China

Correspondence should be addressed to Dongming Wu; wudongming163@163.com

In this paper, we investigate the issue of direction-of-arrival (DOA) estimation of multiple signals in coprime arrays. An algorithm based on multiple signal classification (MUSIC) and forward and backward spatial smoothing (FBSS) is used for DOA estimation of this signal caused by multipath and interference. The large distance between adjacent elements of each subarray in the coprime arrays will bring phase ambiguity issues. According to the feature of the coprime number, the ambiguity problem can be eliminated. The correct DOA estimation can be obtained by searching for the common peak of the spatial spectrum and finding the overlapping peaks in the MUSIC spectrum of the two subarrays. For the rank deficit problem caused by the coherent signal, the FBSS algorithm is used for signal preprocessing before the MUSIC algorithm. Theoretical analysis and simulation results show that the algorithm can effectively solve the rank deficiency and phase ambiguity problems caused by coherent signals and sparse arrays in the coprime arrays.

1. Introduction

Array signal processing is a branch of the signal processing field and is widely used in radar, sonar, satellite, wireless communications, seismology, and other fields [1, 2]. Array signal processing is based on a group of spatially arranged array antennas to process the signal [3]. The purpose of array signal processing is to enhance useful target signals, suppress noise, and obtain signal spatial information. Compared with a single antenna, the use of an antenna array has outstanding advantages in terms of spatial resolution, receiving sensitivity, and anti-interference [4]. Thus, array signal processing has made rapid progress in research and engineering applications in the past 30 years [5].

DOA estimation of space signal is a basic problem in array signal processing. DOA estimation is to estimate the direction of arrival of the signal by receiving the target echo data through the array antenna in the noise or interference environment. And it is a kind of direction-finding technique [6, 7]. In wireless communication, accurate DOA estimation of the signal source can improve communication quality [8]. And it can improve physical layer security combined with beamforming technology [9]. In radar target detection, DOA estimation is the basis for achieving high-precision direction finding [10]. Therefore, it is of great significance to study how to improve the accuracy of DOA estimation. The performance of DOA estimation is determined by the resolution, accuracy, the number of distinguishable targets, etc. [11]. In response to these aspects, corresponding theoretical and applied research has been carried out at home and abroad, which has enabled the rapid development of DOA estimation theory [12–14].

The past researches have proposed a large number of DOA estimation algorithms for different array models, such as the uniform linear array, L-shaped linear array, and uniform circular array [15–17]. In the traditional array structure, the angle ambiguity is avoided by setting the spacing of array elements no more than half wavelength. However, when the frequency of the received signal is high, too small array element spacing will cause larger mutual coupling, and the physical array layout is difficult to achieve. At the same time, high resolution means a larger array aperture, and more physical array elements will further increase the system cost and complexity. Sparse arrays can overcome the structural

limitations of traditional arrays by increasing the array element spacing [18] and thus have been developed and widely used, such as the Minimum Redundancy Array (MRA) [19, 20], Nested Array (NA), and coprime array (CPA) [21–23].

The coprime formation is composed of two subarrays, and the spacing between the subarrays is mutually prime [24]. Compared with the traditional uniform array, the element spacing of the coprime array is greater than half a wavelength. The increase in the element spacing brings the advantages of an increase in the array aperture and a significant reduction in the mutual coupling effect between elements and significantly improves the estimation accuracy and resolution [25]. The DOA of two uniform subarrays of the coprime array is estimated, respectively. According to the relatively prime characteristics of the element spacing of the two subarrays, it is proved that the DOA estimation results of the two subarrays are unique [26, 27]. The coprime array which does not reduce the array aperture of the original array is simple to implement, and the estimation accuracy is greatly improved compared with the uniform array with the same number of antennas [28–30].

DOA estimation algorithms mainly include traditional beamforming, subspace algorithm, and maximum likelihood estimation [31–33]. Among them, the beamforming method has larger error and low resolution; the maximum likelihood algorithm uses the probability distribution of the signal and adopts the high-dimensional search method, which has a large amount of computation. The subspace algorithm uses the orthogonality of signal and noise subspace to realize angle estimation, which requires less computation but cannot process coherent signals [34]. Generally, the minimum resolution that can be achieved under a certain array length is called the Rayleigh Resolution Limit, and the method that exceeds the Rayleigh Resolution Limit is called the superresolution algorithm. Multiple signal classification (MUSIC) proposed in 1979 and estimating signal parameters via rotational invariance techniques (ESPRIT) proposed in 1986 belong to subspace algorithm and are also early classical superresolution methods [35, 36]. No matter the MUSIC algorithm or ESPRIT algorithm, it is necessary for the array element to receive the uncorrelated signal. At this time, the covariance matrix of the source is a full rank matrix, so that the covariance matrix of the signal can be eigendecomposed and the signal subspace and noise subspace can be distinguished.

Most signals are coherent signals in the actual application environment because of the multipath effect and complex transmission channel [37, 38]. For early DOA estimation algorithms such as MUSIC and ESPRIT, they are all based on subspace for DOA estimation. When the received signal is correlated, the eigenvector corresponding to the source signal cannot be obtained by decomposing the subspace eigenvalues. Therefore, DOA estimation of coherent source signals has always been a difficult problem, which is also the focus of spectral estimation. In order to distinguish coherent signals accurately, the spatial smoothing method, singular value decomposition method (SVD method), matrix decomposition method (MD method), and Toeplitz method are developed [39–41].

In this paper, the MUSIC algorithm and the FBSS algorithm are combined to estimate the DOA of coherent signals based on the coprime matrix model under the condition of multipath and interference, and the formulas to solve the signal coherence and angle ambiguity under the coprime matrix are given. Finally, the DOA estimation method for coherent signals is simulated, and the simulation results show the effectiveness of the method.

The remainder is given as follows: Section 2 outlines the basic array signal model of the coprime array. In Section 3, the proposed method for coherent target DOA estimation based on coprime arrays is presented, and the problem of phase ambiguity and rank deficiency is discussed together with its elimination method. Numerical simulations and conclusions are presented in Sections 4 and 5, respectively.

Notations. Throughout the paper, we use the lowercase (uppercase) boldface symbols to represent vectors (matrices). $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and the conjugate transpose, respectively. I_N denotes $N \times N$ identity matrix, diag $(\cdot)$ denotes the diagonal matrix operator, and $E[\cdot]$ denotes the expectation operator.

2. Array Signal Model

The coprime array is a sparse array constructed by using the property of a coprime number. It is composed of two uniform linear arrays. Assuming that the number of subarray elements is M and N, the spacing between two subarrays is $N\lambda/2$ and $M\lambda/2$, respectively, where M and N are coprime integers and λ represents the wavelength of the received signal. The first element of the two subarrays coincides, which is also called the reference element. The coprime array contains a total of elements, and the positions of $M + N - 1$ elements, and the positions of the elements are

$$d = \left\{ Nm\left(\frac{\lambda}{2}\right) \cup Mn\left(\frac{\lambda}{2}\right) \right\}, \tag{1}$$

where $0 \le m \le M - 1, 0 \le n \le N - 1$.

Figure 1(a) is a schematic diagram of the structure of a coprime array with $M + N - 1$ elements. For the convenience of analysis, the coprime matrix is divided into two subarrays, in which the black dot represents subarray 1 and the hollow dot represents subarray 2, as shown in Figure 1(b). In fact, the two subarrays are in a straight line and share the first element.

It is assumed that there are far-field narrow-band signals from different directions in the space, the incident angle is $\theta_k, k = 1, 2, 3, \cdots, K$, and the output noise of each element is a complex Gaussian distribution with zero mean value, which are independent of each other and have the same average power σ^2. The output of the mth element can be expressed as

$$x_m(t) = \sum_{k=1}^{K} a(\theta_k)s_k(t) + n_m(t). \tag{2}$$

(a) An example of coprime array

(b) Two coprime uniform linear subarrays
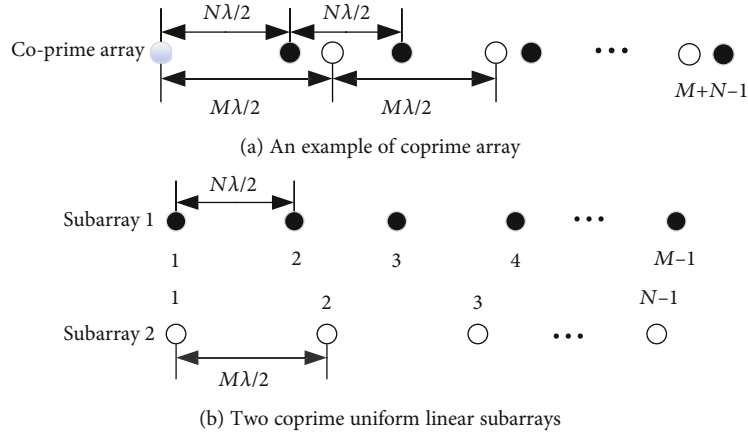
FIGURE 1: Basic structure of coprime linear array.

If the first element is selected as the reference element, the output of the subarray with M elements is

$$\mathbf{x}_M(t) = \mathbf{A}_M(\theta)\mathbf{s}(t) + \mathbf{n}_M(t), \tag{3}$$

where $\mathbf{A}_M(\theta) = [\mathbf{a}_M(\theta_1), \mathbf{a}_M(\theta_2), \cdots, \mathbf{a}_M(\theta_K)]$. The steering vector of the Kth source is expressed as

$$\mathbf{a}_M(\theta_K) = \left[1, e^{-jN\pi \sin \theta_K}, \cdots, e^{-jN(M-1)\pi \sin \theta_K}\right]^T. \tag{4}$$

Source vector $s(t) = [s_1(t), s_2(t), \cdots, s_K(t)]^T$. Similarly, the output of the subarray with N elements is

$$\mathbf{x}_N(t) = \mathbf{A}_N(\theta)\mathbf{s}(t) + \mathbf{n}_N(t), \tag{5}$$

where $\mathbf{A}_N(\theta) = [\mathbf{a}_N(\theta_1), \mathbf{a}_N(\theta_2), \cdots, \mathbf{a}_N(\theta_K)]$. The steering vector of the Kth source is expressed as

$$\mathbf{a}_N(\theta_K) = \left[1, e^{-jM\pi \sin \theta_K}, \cdots, e^{-jM(N-1)\pi \sin \theta_K}\right]^T. \tag{6}$$

Because the noise is independent of each other, the noise and the signal are independent of each other, the covariance matrix of the noise is $\sigma^2\mathbf{I}$, and the covariance matrix of the output of the two subarrays:

$$\mathbf{R}_M = E\left[\mathbf{x}_M\mathbf{x}_M{}^{\mathbf{H}}\right] = \mathbf{A}_M E\left[\mathbf{s}\mathbf{s}^{\mathbf{H}}\right]\mathbf{A}_M{}^{\mathbf{H}} + \sigma^2\mathbf{I}_M, \ \mathbf{A}_M\mathbf{R}_{ss}\mathbf{A}_M{}^{\mathbf{H}} + \sigma^2\mathbf{I}_M,$$

$$\mathbf{R}_N = \mathbf{A}_N\mathbf{R}_{ss}\mathbf{A}_N{}^{\mathbf{H}} + \sigma^2\mathbf{I}_N. \tag{7}$$

In the DOA estimation based on the coprime array, the array aperture is greatly expanded by the construction of a virtual array model. At the same time, the ranks of covariance matrices constructed by different methods are also different, but generally, the virtual array degree of freedom of the coprime array is far greater than that of the physical array. The degree of freedom is an important sign that the antenna array can estimate the number of targets or sources. The higher the degree of freedom is, the more sources the array can estimate. Besides, the degree of freedom is proportional to the estimation accuracy. Generally, the higher the degree of freedom is, the higher the positioning accuracy will be.

3. DOA Estimation of Coherent Signals

Due to the interference effect of coherent signals, the number of subspaces processed by the ordinary DOA estimation algorithm will be reduced and affect the direction-finding accuracy. However, the FBSS algorithm does not appear in such a situation. Based on this idea, we decompose the coprime array into two uniform subarrays. For each subarray, the FBSS algorithm and the traditional MUSIC algorithm are combined to process the coherent signal. By analyzing the DOA results of the two subarrays, the correct target angle can be obtained, and the problem of rank deficiency caused by phase ambiguity and coherent signal is solved.

3.1. Spatial Smoothing on Subarrays. Coherent signals are easily generated in signal transmission due to the complex space environment. The appearance of coherent sources may lead to serious degradation of DOA estimation performance. In the traditional MUSIC algorithm based on subspace, the covariance matrix of the received data needs to be full rank, but the covariance matrix of the coherent source is not full rank, the signal eigenvectors diverge into the noise subspace, and the singular value decomposition cannot completely distinguish the signal subspace from the noise subspace, which leads to deterioration of DOA estimation performance.

The basic idea of the spatial smoothing algorithm is to divide the array into several overlapping subarrays and use the covariance matrix of the received data of subarrays to replace the original covariance matrix. By sacrificing a certain effective array aperture, the covariance matrix of the received data is restored to full rank, so as to achieve the preprocessing operation of decoherence.

The covariance matrix uses the autocorrelation relationship between signals to extract information. In practical applications, the maximum likelihood function of the covariance matrix is usually calculated by selecting a large enough number of snapshots to approximate the ideal covariance matrix. In this case, the estimated covariance matrix of the output data can be expressed as

$$\mathbf{R}_M = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_M(l)\mathbf{x}_M^H(l),$$

$$\mathbf{R}_N \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_N(l)\mathbf{x}_N^H(l). \tag{8}$$

As shown in Figure 2, consider the subarray with the number of M elements and the spacing of $N\lambda/2$ in the coprime matrix. Under the forward space smoothing algorithm, the equidistant linear array is divided into L sub-arrays by sliding, and each subarray has n elements, where $n = M - L + 1$.

In this case, the output of the first forward subarray can be expressed as

$$\mathbf{x}_l^f(t) = [\mathbf{x}_l(t), \mathbf{x}_{l+1}(t), \cdots, \mathbf{x}_{l+n-1}(t)]^T = \mathbf{A}_M\mathbf{D}^{l-1}\mathbf{s}(t) + \mathbf{n}_l(t) \ (1 \le l \le L), \tag{9}$$

where $\mathbf{A}_M = [\mathbf{a}_M(\theta_1), \mathbf{a}_M(\theta_2), \cdots, \mathbf{a}_M(\theta_K)]$ is the $n * K$ dimension direction matrix and $\mathbf{a}_M(\theta)$ is the n-dimension guidance vector. $\mathbf{D} = \text{diag}\left(e^{j(2\pi Nd/\lambda)\sin\theta_1}, e^{j(2\pi Nd/\lambda)\sin\theta_2}, \cdots, e^{j(2\pi Nd/\lambda)\sin\theta_K}\right)$ is a rotation-invariant matrix between subarrays.

The covariance matrix of the lth forward submatrix can be expressed as

$$\mathbf{R}_l^f = E\left[\mathbf{x}_l^f(t)\mathbf{x}_l^f(t)^H\right] = \mathbf{A}_M\mathbf{D}^{l-1}\mathbf{R}_S\left(\mathbf{D}^{l-1}\right)^H\mathbf{A}_M^H + \sigma^2\mathbf{I}. \tag{10}$$

The forward spatial smoothing covariance matrix is defined as

$$\mathbf{R}_l = \frac{1}{L}\sum_{l=1}^{L}\mathbf{R}_l^f. \tag{11}$$

Similarly, if the subarray is divided from the last element of the array, the covariance matrix of backward spatial smoothing can be obtained as follows:

$$\mathbf{R}_b = \frac{1}{L}\sum_{l=1}^{L}\mathbf{R}_l^b. \tag{12}$$

Because the backward smooth array is the conjugate reverse order of the forward smooth array, the relation between $\mathbf{R}_b$ and $\mathbf{R}_l$ is the conjugate reverse order invariant. Although the one-way smoothing algorithm can solve the problem of coherent signals, it sacrifices more array aperture. The FBSS algorithm can increase the number of estimable cells by simultaneously performing forward and backward smoothing. The covariance matrix is the average of forward smoothing and backward smoothing covariance matrices

$$\mathbf{R}_{fb} = \frac{1}{2}\left(\mathbf{R}_f + \mathbf{R}_b\right). \tag{13}$$

3.2. DOA Estimation of Subarrays. The MUSIC algorithm is

the most classic superresolution DOA estimation algorithm, which obtains the cell direction by searching the spectrum peak in the spatial domain. Compared with multidimensional algorithms such as maximum likelihood (ML) and weighted subspace fitting (WSF), the algorithm has less computation. The basic idea of the MUSIC algorithm is to eigendecompose the covariance matrix of the array output data to obtain the signal subspace corresponding to the signal component and the noise subspace orthogonal to the signal component and then use the orthogonality of the two subspaces to estimate the signal parameters.

The covariance matrices of the two submatrices are eigendecomposed, respectively, to obtain

$$\mathbf{R}_M = \mathbf{U}_{SM}\sum_{SM}\mathbf{U}_{SM}^H + \mathbf{U}_{NM}\sum_{NM}\mathbf{U}_{NM}^H,$$

$$\mathbf{R}_M = \mathbf{U}_{SN}\sum_{SN}\mathbf{U}_{SN}^H + \mathbf{U}_{NN}\sum_{NN}\mathbf{U}_{NN}^H. \tag{14}$$

Among them, matrices $\mathbf{U}_{SM} \in \mathbb{C}^{M*K}$ and $\mathbf{U}_{SN} \in \mathbb{C}^{N*K}$ are the signal subspaces formed by the eigenvectors corresponding to K large eigenvalues in $\hat{\mathbf{R}}_M$ and $\hat{\mathbf{R}}_N$, matrices $\sum_{SM} \in \mathbb{C}^{K*K}$ and $\sum_{SN} \in \mathbb{C}^{K*K}$ are the diagonal matrices formed by K large eigenvalues in $\hat{\mathbf{R}}_M$ and $\hat{\mathbf{R}}_N$, matrices $\mathbf{U}_{NM} \in \mathbb{C}^{M*(M-K)}$ and $\mathbf{U}_{NN} \in \mathbb{C}^{M*(N-K)}$ are the noise subspaces formed by the eigenvectors corresponding to $M - K$ and $N - K$ small eigenvalues in $\hat{\mathbf{R}}_M$ and $\hat{\mathbf{R}}_N$, and the matrices $\sum_{NM}$ and $\sum_{NN}$ are diagonal matrices composed of $M - K$ and $N - K$ small eigenvalues in $\hat{\mathbf{R}}_M$ and $\hat{\mathbf{R}}_N$, respectively (these small eigenvalues are equal, which is the noise power σ^2).

Under ideal conditions, the signal subspace $\mathbf{U}_S$ and the noise subspace $\mathbf{U}_N$ are orthogonal to each other, so the array flow pattern vector $\mathbf{a}^H(\theta)$ corresponding to the signal subspace is also orthogonal to the noise subspace $\mathbf{U}_N$, namely,

$$\mathbf{a}^H(\theta)\mathbf{U}_N = 0. \tag{15}$$

In practice, the steering vector and the noise subspace cannot be completely orthogonal due to the existence of other noises. Usually, the minimum optimization search process is used to find the minimum value to realize the direction-of-arrival estimation. This process can be expressed as

$$\theta_{\text{MUSIC}} = \arg_\theta \min \mathbf{a}^H(\theta)\hat{\mathbf{U}}_N\hat{\mathbf{U}}_N^H\mathbf{a}(\theta). \tag{16}$$

Based on the orthogonality between the signal subspace and the noise subspace, the spectral function of the MUSIC space power spectrum of the two subarrays can be expressed as

$$P_{\text{MUSIC\_}M} = \frac{1}{\mathbf{a}_M^H(\theta)\hat{\mathbf{U}}_M\hat{\mathbf{U}}_M^H\mathbf{a}_M(\theta)},$$

$$P_{\text{MUSIC\_}N} = \frac{1}{\mathbf{a}_N^H(\theta)\hat{\mathbf{U}}_N\hat{\mathbf{U}}_N^H\mathbf{a}_N(\theta)}. \tag{17}$$

Among them, the value θ range is generally $(-\pi/2, \pi/2)$.
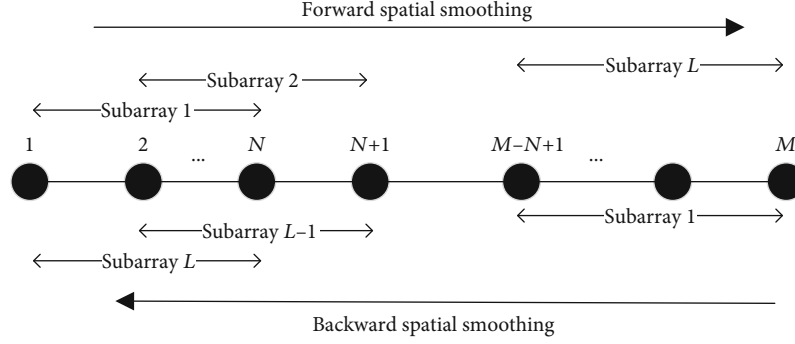
Figure 2: Schematic diagram of FBSS algorithm.

The accurate DOA estimate can be obtained by searching for the coincident peaks of the two subarray spectral functions of the coprime array.

3.3. Ambiguity Elimination. The calculation formula of phase defuzzification for incoherent sources is given in Reference [31]. For coherent sources, if there is phase ambiguity after using the spatial smoothing algorithm for the subarray with M elements, it can be seen from (13) that the steering vector between the real angle θ_k and the blurred angle $\overline{\theta}_k$ should be equal, that is,

$$A_M(\theta_k) = A_M\left(\overline{\theta}_k\right),$$
$$\exp\left(-jN\pi \sin\left(\theta_k\right)\right) = \exp\left(-jN\pi \sin\left(\overline{\theta}_k\right)\right). \tag{18}$$

After simplification, we get

$$\sin\left(\theta_k\right) - \sin\left(\overline{\theta}_k\right) = \frac{2P_M}{N}, \tag{19}$$

where P_M is a nonzero integer, $\theta_k, \overline{\theta}_k \in (-\pi/2, \pi/2)$. For any θ_k and $\overline{\theta}_k$, it must satisfy $|\sin\left(\theta_k\right) - \sin\left(\overline{\theta}_k\right)| < 2$, that is, $|2P/N| < 2$. The value range of P_M can be $-(N-1), -(N-2), \cdots, -1, 1, \cdots, N-1$; there are $2(N-1)$ values in total. Consider that θ_k and $\overline{\theta}_k$ can be exchanged. In addition to the real angle, there are $N-1$ fuzzy angles. That is to say, for a single subarray M whose element spacing is $N\lambda/2$ in the coprime array, there must be phase ambiguity. There are N peaks in the MUSIC spectrum using spatial smoothing, and the $N-1$ peaks correspond to the fuzzy angle.

In the same way, when considering the single subarray N whose element spacing is $M\lambda/2$ in the coprime array, the fuzzy angle needs to meet the requirement:

$$\sin\left(\theta_k\right) - \sin\left(\overline{\theta}_k\right) = \frac{2P_N}{M}. \tag{20}$$

The value range of P_N can be $-(M-1), -(M-2), \cdots, -1,$

$1, \cdots, M-1$. Combined with (19), the condition of phase ambiguity is obtained as follows:

$$\frac{2P_M}{N} = \frac{2P_N}{M}. \tag{21}$$

After simplification, we can get $NP_N = MP_M$. Since M and N are relatively prime, it cannot make the equation hold in the range of value; that is to say, $\overline{\theta}_k$ does not exist and there is no angle ambiguity. Therefore, the unique DOA estimation can be determined by using the spatial smoothing algorithm and MUSIC algorithm, respectively, for the subarrays of the coprime array, and then finding the overlapped peaks in the two groups of spectrum.

3.4. Complexity Analysis. The spatial smoothing algorithm, SVD algorithm, and Toeplitz algorithm can process coherent signals well, and the computational complexity of these three decoherence algorithms increases gradually. At present, the spatial smoothing algorithm has the least amount of computation; that is, the time of DOA processing is the shortest. In addition, spatial smoothing technology is also more mature, which is a more practical algorithm for processing coherent signals. The uniform linear array with M elements can distinguish $2M/3$ coherent targets by using the spatial smoothing algorithm. And the virtual element number of the coprime array with two subarray elements M and N is $O(MN)$. Therefore, $O(2MN/3)$ coherent targets can be distinguished by the coprime array with this algorithm. In the same case, only $O(2(M+N-1)/3)$ coherent targets can be distinguished by the uniform linear array with this algorithm.

4. Simulation Results

In this section, we have carried out the corresponding simulation analysis to prove the effectiveness of FBSS and MUSIC algorithms for coherent signals under the coprime array model. In the simulation process, the number of elements of two subarrays of the coprime array is $M = 7$ and $N = 5$, and the spacing between elements is $5\lambda/2$ and $7\lambda/2$, where λ is half wavelength. For a fair comparison, a 12 uniform linear array with half-wavelength spacing is also simulated with the FBSS and MUSIC algorithms. Consider two coherent signals in the space, which are incident from 0° and 30° to the
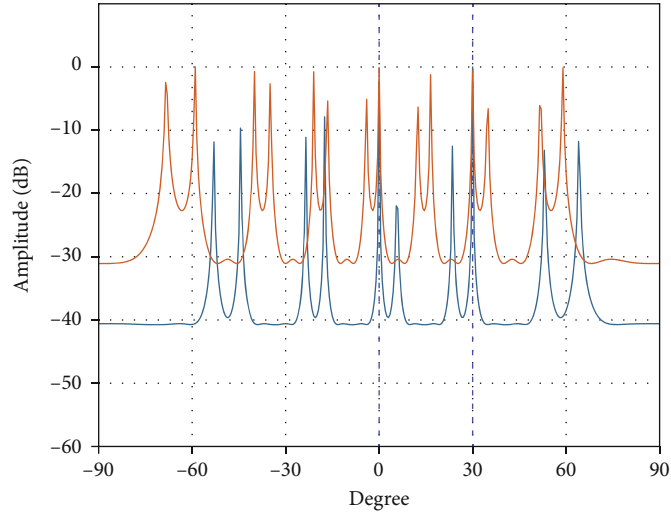
FIGURE 3: Spatial spectrum of DOA estimation simulation of two subarrays for coherent sources.
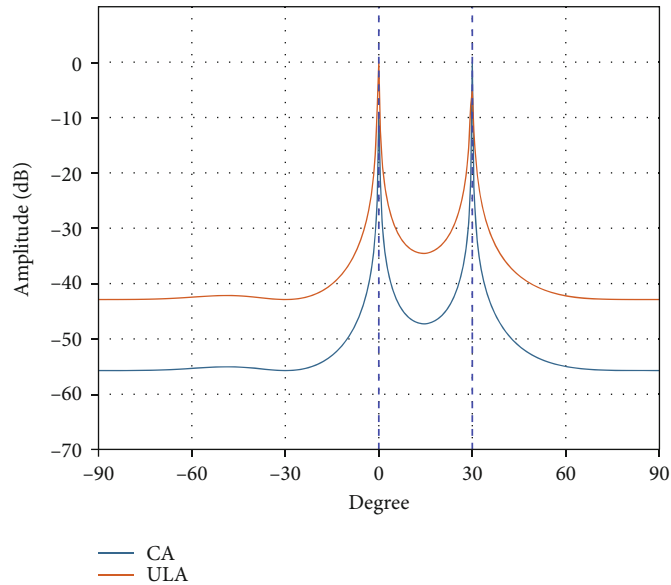
FIGURE 4: Spatial spectrum of DOA estimation simulation of uniform linear array and coprime array.

coprime array, respectively, and the noise is Gaussian white noise. The searching steps for all methods are set to be 0.02°.

4.1. Spatial Spectrum. We then show the spatial spectrum using FBSS and MUSIC algorithms in Figure 3, where we assume the signal to noise (SNR) as 10 dB and snapshot $n = 200$. The red spectral line is the subarray spectrum with $M = 7$, and the blue spectral line is the subarray spectrum with $N = 5$. It can be seen from the previous derivation that phase ambiguity will be generated when using spatial smoothing and MUSIC algorithm for a single subarray of coprime array. For a subarray with an element spacing of $N\lambda/2$, estimating a DOA will produce $N - 1$ ambiguity angles. Therefore, the subarray with $M = 7$ has 10 peaks, 8 of which are fuzzy angles. And the subarray with $N = 5$ has

14 peaks, 12 of which are fuzzy angles. However, the common spectral peak formed by the two subarrays is only at 0° and 30°, which proves the correctness of the algorithm for DOA estimation of coherent signals.

Under this condition, we further compare the DOA estimation spectrum of 11 elements uniform linear array and coprime array. The specific results are shown in Figure 4. Through the comparison of DOA estimation spectrum peaks, we can intuitively find that two coherent signals, whether uniform linear array or coprime array, can be well distinguished. But the coprime array is better than the uniform linear array in suppressing interference. Because the number of virtual elements of the coprime array is much larger than that of the uniform linear array, it has a higher degree of freedom and better estimation performance.
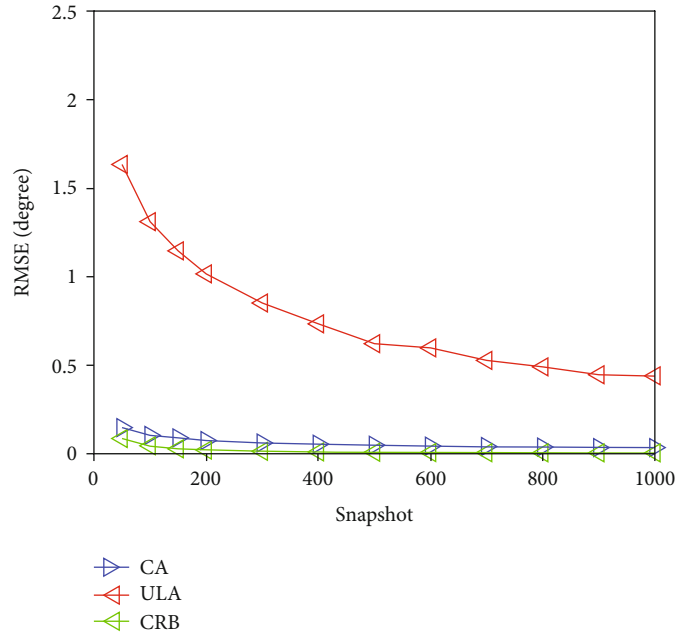
FIGURE 5: RMSE versus the number of snapshots.
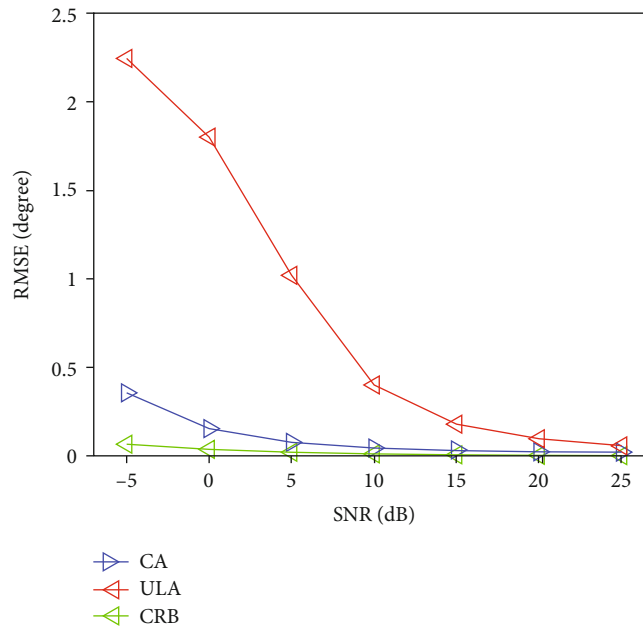
FIGURE 6: RMSE versus the SNR.

4.2. Root Mean Square Error (RMSE). In this simulation, we study the RMSE performance of the two arrays under different configurations. The root mean square error (RMSE) of the estimates is defined as the performance metric:

$$\text{RMSE} = \sqrt{\frac{1}{NK}\sum_{n=1}^{N}\sum_{k=1}^{K}\left[\left(\theta_k - \overline{\theta}_k^{(i)}\right)\right]}, \tag{22}$$

where N_0 denotes the times of Monte-Carlo simulations

and $\overline{\theta}_k^{(i)}$ and θ_k are the estimate and real values of the k th DOA for the nth trial, $n = 1, 2, 3, \cdots, N$. The targets are located at $\theta_1 = 10°$, $\theta_2 = 20°$, $\theta_3 = 30°$, and θ_1, θ_2 are coherent signals. For each simulation scenario, $S = 500$ rounds of Monte-Carlo runs are conducted. The Cramer-Rao bound (CRB) is plotted as a benchmark.

Figure 5 depicts the RMSEs of different configurations in terms of SNR, where the number of snapshots is 200. In Figure 6, we compare the RMSEs of the two arrays versus the number of snapshots, where the SNR is set as 5 dB. It is

obvious that the performance of all these configurations improves with the increase of the SNR and number of snapshots. But the performance of the coprime array is better than that of the uniform linear array in any case. Even in the case of low snapshot number and low signal-to-noise ratio, the coprime array can also show good DOA estimation performance.

Although the phase ambiguity of a single subarray of the coprime array is caused by the large element spacing, accurate DOA estimation can be achieved by comparing the peak values of the two subarrays. Through the simulation experiment, we can see that compared with the uniform linear array, the coprime array using the spatial smoothing algorithm greatly improves the resolution and reduces the computational complexity.

5. Conclusions

In this paper, we use FBSS and MUSIC algorithms for DOA estimation of coherent signals based on the structure of coprime arrays, where the spatial spectrum of each decomposed subarray can generate spectral peaks at the actual DOAs and multiple ambiguous DOAs simultaneously. And we solve the phase ambiguity by finding the common spectral peaks in the spectrum of the two subarrays. Theoretical analysis and simulation results show that the algorithm can effectively process DOA estimation of coherent signals, and the coprime array has better performance than the uniform linear array. However, some spatial degrees of freedom are sacrificed when using the spatial smoothing algorithm. How to increase the spatial degrees of freedom and improve the direction-finding accuracy under low snapshot numbers will be our further research direction.

Data Availability

Data will be made available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[2] J. Benesty, J. Chen, and Y. Huang, "Microphone array signal processing," *Journal of the Acoustical Society of America*, vol. 125, no. 6, pp. 4097-4098, 2008.

[3] J.-J. Jiang, F.-J. Duan, J. Chen, Y.-C. Li, and X.-N. Hua, "Mixed near-field and far-field sources localization using the uniform linear sensor array," *IEEE Sensors Journal*, vol. 13, no. 8, pp. 3136–3143, 2013.

[4] X. Wang, M. Huang, and L. Wan, "Joint 2D-DOD and 2D-DOA estimation for coprime EMVS–MIMO radar," *Circuits, Systems, and Signal Processing*, vol. 40, no. 6, pp. 2950–2966, 2021.

[5] Z. Zheng, C. Yang, W. Q. Wang, and H. C. So, "Robust DOA Estimation Against Mutual Coupling With Nested Array," *IEEE Signal Processing Letters*, vol. 27, no. 4, pp. 1360–1364, 2020.

[6] Z. Ye and X. Xu, "DOA estimation by exploiting the symmetric configuration of uniform linear array," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 12, pp. 3716–3720, 2007.

[7] J. He, L. Li, and T. Shu, "Sparse nested arrays with spatially spread square acoustic vector sensors for high accuracy under-determined direction finding," *IEEE Transactions on Aerospace and Electronic Systems*, 2021.

[8] J. Wang, H. Xu, G. J. Leus, and G. A. Vandenbosch, "Experimental Assessment of the Co-Array Concept for DoA Estimation in Wireless Communications," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 6, pp. 3064–3075, 2018.

[9] K. Almidfa, G. V. Tsoulos, and A. Nix, "Performance evaluation of direction-of-arrival (DOA) estimation algorithms for mobile communication systems," in *VTC2000-Spring. 2000 IEEE 51st Vehicular Technology Conference Proceedings (Cat. No.00CH37026)*, pp. 1055–1059, Tokyo, Japan, 2000.

[10] X. Zhang, L. Xu, L. Xu, and D. Xu, "Direction of departure (DOD) and direction of arrival (DOA) estimation in MIMO radar with reduced-dimension MUSIC," *IEEE Communications Letters*, vol. 14, no. 12, pp. 1161–1163, 2010.

[11] M. C. Dogan and J. M. Mendel, "Applications of cumulants to array processing .I. Aperture extension and array calibration," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1200–1216, 1995.

[12] J. He, L. Li, and T. Shu, "Sparse nested arrays with spatially spread orthogonal dipoles: high accuracy passive direction finding with less mutual coupling," *IEEE Transactions on Aerospace and Electronic Systems*, p. 1, 2021.

[13] G. He, M. Song, X. He, and Y. Hu, "GPS signal acquisition based on compressive sensing and modified greedy acquisition algorithm," *IEEE Access*, vol. 7, pp. 40445–40453, 2019.

[14] G. He, M. Song, S. Zhang, P. Song, and X. Shu, "Sparse GLONASS signal acquisition based on compressive sensing and multiple measurement vectors," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9654120, 11 pages, 2020.

[15] Zhongfu Ye, Jisheng Dai, Xu Xu, and Xiaopei Wu, "DOA estimation for uniform linear array with mutual coupling," *IEEE Transactions on Aerospace & Electronic Systems*, vol. 45, no. 1, pp. 280–288, 2009.

[16] N. Xi and L. Liping, "A computationally efficient subspace algorithm for 2-D DOA estimation with L-shaped array," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 971–974, 2014.

[17] J.-J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 702–709, 2001.

[18] J. He, Z. Zhang, T. Shu, and W. Yu, "Sparse nested array with aperture extension for high accuracy angle estimation," *Signal Processing*, vol. 176, p. 107700, 2020.

[19] D. A. Linebarger, I. H. Sudborough, and I. G. Tollis, "Difference bases and sparse sensor arrays," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 716–721, 1993.

[20] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Transactions on Antennas and Propagation*, vol. 16, no. 2, pp. 172–175, 1968.

[21] J. Yang, G. Liao, and J. Li, "An efficient off-grid DOA estimation approach for nested array signal processing by using sparse Bayesian learning strategies," *Signal Processing*, vol. 128, pp. 110–122, 2016.

[22] P. Pal and P. P. Vaidyanathan, "Nested arrays: a novel approach to array processing with enhanced degrees of freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, 2010.

[23] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.

[24] P. P. Vaidyanathan and P. Pal, "Theory of sparse coprime sensing in multiple dimensions," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3592–3608, 2011.

[25] Z. Weng and P. M. Djurić, "A search-free DOA estimation algorithm for coprime arrays," *Digital Signal Processing*, vol. 24, pp. 27–33, 2014.

[26] C. Zhou, Z. Shi, Y. Gu, and N. A. Goodman, "DOA estimation by covariance matrix sparse reconstruction of coprime array," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2369–2373, South Brisbane, QLD, Australia, 2015.

[27] Z. Shi, C. Zhou, Y. Gu, N. A. Goodman, and F. Qu, "Source estimation using coprime array: a sparse reconstruction perspective," *IEEE Sensors Journal*, vol. 17, no. 3, pp. 755–765, 2017.

[28] P. Pakrooh, L. L. Scharf, and A. Pezeshki, "Modal analysis using co-prime arrays," *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2429–2442, 2016.

[29] Q. Wu, F. Sun, P. Lan, G. Ding, and X. Zhang, "Two-dimensional direction-of-arrival estimation for co-prime planar arrays: a partial spectral search approach," *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5660–5670, 2016.

[30] J. Shi, G. Hu, X. Zhang, F. Sun, W. Zheng, and Y. Xiao, "Generalized co-prime MIMO radar for DOA estimation with enhanced degrees of freedom," *IEEE sensors journal*, vol. 18, no. 3, pp. 1203–1212, 2018.

[31] E. Gonen and J. M. Mendel, "Applications of cumulants to array processing. III. Blind beamforming for coherent signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 9, pp. 2252–2264, 1997.

[32] X. Wang, L. T. Yang, D. Meng, M. Dong, K. Ota, and H. Wang, "Multi-UAV cooperative localization for marine targets based on weighted subspace fitting in SAGIN environment," *EEE Internet of Things Journal*, 2021.

[33] F. Wen, J. Shi, and Z. Zhang, "Closed-form estimation algorithm for EMVS-MIMO radar with arbitrary sensor geometry," *Signal Processing*, vol. 186, p. 108117, 2021.

[34] C. L. Liu and P. P. Vaidyanathan, "Remarks on the spatial smoothing step in coarray MUSIC," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1438–1442, 2015.

[35] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas & Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[36] P. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[37] J. H. Cozzens and M. J. Sousa, "Source enumeration in a correlated signal environment," *IEEE Transactions on Signal Processing*, vol. 42, no. 2, pp. 304–317, 1994.

[38] E. BouDaher, F. Ahmad, and M. G. Amin, "Sparsity-based direction finding of coherent and uncorrelated targets using active nonuniform arrays," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1628–1632, 2015.

[39] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 37, no. 1, pp. 8–15, 1989.

[40] Chongying Qi, Yongliang Wang, Yongshun Zhang, and Ying Han, "Spatial difference smoothing for DOA estimation of coherent signals," *IEEE Signal Processing Letters*, vol. 12, no. 11, pp. 800–802, 2005.

[41] C. Zhou, Z. Shi, Y. Gu, and X. Shen, "DECOM: DOA estimation with combined MUSIC for coprime array," in *2013 International Conference on Wireless Communications and Signal Processing*, pp. 1–5, Hangzhou, China, 2013.

WILEY | Hindawi

Research Article

Partial Dictionary Based Off-Grid DOA Estimation Using Combined Coprime and Nested Array

Jianfeng Li [ID],[1,2] Xiong Xu,[1] Ping Li,[2] and Qiting Zhang[2]

[1]*State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), Luoyang 471003, China*
[2]*College of Electronic Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*

Correspondence should be addressed to Jianfeng Li; lijianfengtin@126.com

A partial dictionary based direction of arrival (DOA) estimation method which addresses the off-grid problem and exploits combined coprime and nested array (CCNA) is proposed. Compared to general coprime array, CCNA yields two sparse coprime subarrays in the coarray domain by adding a third subarray in the physical-array domain. To ensure the DOA estimation performance, the subarray with larger aperture is chosen, and the cyclic phase ambiguity caused by the sparse subarray allows partial dictionary covering arbitrary cycle to represent the whole atoms, and then, the off-grid sparse reconstruction method is developed to amend the grid mismatch. After the sparse recovery and off-grid compensation, ambiguous DOA estimations can be eliminated by substituting the estimations into the whole virtual array. Multiple simulations verify that the proposed algorithm outperforms the other state-of-the-art methods in terms of DOA estimation accuracy and angular resolution.

1. Introduction

Direction of arrival (DOA) estimation using antenna array is an important issue in many systems, e.g., radar, sonar, and wireless communication [1–5]. Compared to conventional subspace based methods, sparse representation based DOA estimation methods have been attractive since they can provide higher resolution and require fewer samples [6], and many effective sparse representation based methods have been proposed. The greedy methods [7, 8] require the prior information of source number and are sensitive to the noise, and the l_1-norm based algorithms, such as the l_1-norm singular value decomposition (l_1-SVD) method [9], sparse recovery using weighted subspace fitting (SRWSF) method [10], sparse representation of array covariance vector (SRACV) method [11], and sparse iterative covariance-based estimation (SPICE) method [12], can reduce the sensitivity to noise and estimate the angles via convex optimization. However, these methods discretize the whole spatial range into a grid, which will result in performance degradation when the sources are not exactly located on the grid, i.e., the grid mismatch problem [13]. In

[14], the off-grid sources were considered and estimated by introducing grid offsets in the sparse Bayesian inference (SBI). Based on the joint sparsity between original signal and the grid mismatch variables, joint sparse recovery method was proposed in [15]. Meanwhile, the grid-less methods are developed to directly recover the covariance matrix based on atomic norm or nuclear norm minimization [16, 17]. However, these methods only concentrate on the physical array model, which has limited degree of freedom (DOF).

Sparse array design has been developed to increase the virtual DOF in the difference coarray domain. Nested array was proposed in [18], which can generate $O(N^2)$ DOF in the difference coarray domain with only $O(N)$ physical antennas [19, 20]. Nested array can also be applied in radar system to increase the virtual DOF and enhance the spatial resolution [21]. However, the nested array has a dense subarray, which suffers from the mutual coupling problem. Coprime array [22], another well-known sparse array, was proposed to reduce the mutual coupling influence. With $O(M + N)$ sparsely spaced physical antennas, coprime array can achieve $O(MN)$ DOF [23], which is generally

nonuniform, and the coarray has more holes compared to that of nested array. Therefore, many works are developed to modify the coprime array to generate more continuous virtual elements in the coarray domain, such as the augmented coprime array (ACA) [24], generalized coprime array (GCA) [25], and thinned coprime array (TCA) [26]. To deal with the one-snapshot situation in the coarray domain, sparse representation based methods have also been introduced for sparse array [27–29]. For off-grid sources, a joint reconstruction method named joint LASSO (JLASSO) was proposed in [30], which can exploit the large DOF in the coarray domain of coprime array and amend the grid mismatch via joint sparse recovery. However, the computation complexity is very high due to the dictionary covering whole spatial range.

There is also another way to utilize coprime array, i.e., the separate processing of the two subarrays of coprime array, and the unique estimation is determined from the coincide results from the two subarrays, such as the combined multiple signal classification (MUSIC) method [31], partial search MUSIC method [32], root MUSIC method [33], and combined estimation of signal parameters via rotational invariance technique (ESPRIT) based method [34]. However, as the subarrays are processed separately in the physical-array domain, the exploited DOF is limited, and an alternative way is to transform the coprime relationship into the coarray domain [35]. A combined coprime and nested array (CCNA) geometry, which is obtained by adding a third subarray nested to both of the two subarrays, was proposed in [36], where the coprime subarrays are transformed into the coarray domain to achieve large aperture and DOF. However, the utilized MUSIC method results in aperture loss.

In this paper, we propose an off-grid DOA estimation method, which requires only partial dictionary based on CCNA. Due to the nested relationship within the subarrays, the two virtual subarrays are still sparse but uniform after the vectorization of the covariance matrices. Different with conventional schemes, we adopt the subarray with larger aperture for DOA estimation to avoid the negative effective brought by the smaller subarray. The sparsity of the virtual array enables partial dictionary covering partial spatial range to represent the whole-range atoms. Meanwhile, the off-grid sparse reconstruction method is developed to amend the grid mismatch. Finally, ambiguous DOA estimations can be eliminated based on coprime-ness by substituting the estimations into the whole virtual array. Numerical simulations show that the proposed algorithm outperforms the ACA method [24], partial search (PS) MUSIC [32], root MUSIC method [33], and CCNA with root MUSIC [36] in terms of estimation accuracy and angular resolution.

Notation: $(.)^T$, $(.)^*$, $(.)^H$, and $(.)^+$ denote transposition, conjugation, conjugate-transposition, and pseudo-inversion, respectively. $E[.]$ and vec$(.)$ denote the operations of expectation and vectorization, respectively. diag $(\mathbf{a})$ is a diagonal matrix with vector $\mathbf{a}$ being the diagonal elements, and $\mathbf{I}_p$ is a $p \times p$ identity matrix. $\|\cdot\|_2$ means l_2 norm, and angle(a) means the phase of a. $\otimes$, $\circ$, and ./ denote kronecker product, Khatri_rao product, and element-wise division, respectively.
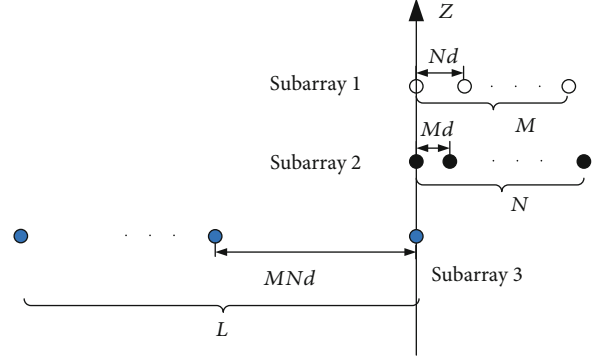
FIGURE 1: The structure of CCNA.

2. Data Model

Figure 1 shows the structure of CCNA, which is composed of three subarrays. Subarray 1 and subarray 2 form the original coprime array, where M and N are coprime integers. The third subarray is arranged along the negative side with L elements and interelement spacing being MNd, where d is the unit spacing, which is generally set as half-wavelength. It is also indicated that subarray 1 and subarray 3 form a nested array with the minimum interelement spacing being Nd, and subarray 1 and subarray 3 form another nested array with the minimum interelement spacing being Md. The total antenna number of CCNA is $M + N + L - 2$ as the subarrays share the same element in the origin.

Assume that there are K plane waves impinging upon the array with DOAs being $\theta_k, k = 1, \cdots, K$, which is angle between the wave line and Z axis. Then, the outputs of the subarrays are expressed as

$$
\begin{aligned}
\mathbf{x}_1(t) &= \mathbf{A}_1\mathbf{s}(t) + \mathbf{n}_1(t), \\
\mathbf{x}_2(t) &= \mathbf{A}_2\mathbf{s}(t) + \mathbf{n}_2(t), \\
\mathbf{x}_3(t) &= \mathbf{A}_3\mathbf{s}(t) + \mathbf{n}_3(t),
\end{aligned}
\tag{1}
$$

where $\mathbf{s}(t) = [s_1(t), \cdots, s_K(t)]^T \in C^{K \times 1}$ is the signal vector. $\mathbf{n}_1(t)$, $\mathbf{n}_2(t)$, and $\mathbf{n}_3(t)$ are the additive white Gaussian noise (AWGN) vectors with the same noise power σ^2. $\mathbf{A}_1 = [\mathbf{a}_1(\theta_1), \cdots, \mathbf{a}_1(\theta_K)]$, $\mathbf{A}_2 = [\mathbf{a}_2(\theta_1), \cdots, \mathbf{a}_2(\theta_K)]$, and $\mathbf{A}_3 = [\mathbf{a}_3(\theta_1), \cdots, \mathbf{a}_3(\theta_K)]$ denote the direction matrices of subarray 1, subarray 2, and subarray 3, respectively. The columns are the corresponding steering vectors, which are expressed as

$$
\mathbf{a}_1(\theta_k) = \left[1, e^{-jN\pi \sin \theta_k}, \cdots, e^{-j(M-1)N\pi \sin \theta_k} \right]^T, k = 1, \cdots, K,
$$

$$
\mathbf{a}_2(\theta_k) = \left[1, e^{-jM\pi \sin \theta_k}, \cdots, e^{-j(N-1)M\pi \sin \theta_k} \right]^T, k = 1, \cdots, K,
$$

$$
\mathbf{a}_3(\theta_k) = \left[1, e^{jMN\pi \sin \theta_k}, \cdots, e^{j(L-1)MN\pi \sin \theta_k} \right]^T, k = 1, \cdots, K.
$$

$$\tag{2}$$

3. Partial Dictionary Based Off-Grid DOA Estimation Method

3.1. Sparse Representation Using Partial Dictionary. Combine the outputs of subarray 1 and subarray 3 to form the first nested array $\mathbf{y}_1(t) = [\mathbf{x}_1{}^T(t), \mathbf{x}_3{}^T(t)]^T$, whose covariance matrix is

$$\mathbf{R}_1 = E[\mathbf{y}_1(t)\mathbf{y}_1^H(t)] = \mathbf{A}_{n1}\mathbf{R}_s\mathbf{A}_{n1}^H + \sigma^2\mathbf{I}_{M+L-1}, \qquad (3)$$

where $\mathbf{A}_{n1}(t) = [\mathbf{A}_1{}^T(t), \mathbf{A}_3{}^T(t)]^T$ is the combined direction matrix and $\mathbf{R}_s = E[\mathbf{s}(t)\mathbf{s}^H(t)] = \mathrm{diag}\,(\sigma_1{}^2, \cdots, \sigma_K{}^2)$ is a diagonal matrix containing signal powers. To obtain the virtual array in the coarray domain, the vectorization of the covariance matrix is

$$\mathbf{r}_1 = \mathrm{vec}(\mathbf{R}_1) = (\mathbf{A}_{n1}^* \circ \mathbf{A}_{n1})\mathbf{p} + \sigma^2\mathrm{vec}(\mathbf{I}_{M+L-1}), \qquad (4)$$

where $\mathbf{p} = [\sigma_1{}^2, \cdots, \sigma_K{}^2]^T$. Due to the nested relationship, there are $2LM - 1$ continuous elements located from $-(LM - 1)Nd$ to $(LM - 1)Nd$ in the virtual array with interelement spacing being Nd [36]. After selecting continuous elements from $\mathbf{r}_1$, then, we obtain

$$\mathbf{r}_{s1} = \mathbf{W}_{s1}\mathbf{r}_1 = \mathbf{A}_{s1}\mathbf{p} + \sigma^2\mathbf{e}_1, \qquad (5)$$

where $\mathbf{W}_{s1}$ is the selecting matrix and $\mathbf{e}_1$ denotes the column vector after the same selecting operation from $\mathrm{vec}(\mathbf{I}_{M+L-1})$. $\mathbf{A}_{s1} = [\mathbf{a}_{s1}(\theta_1), \mathbf{a}_{s1}(\theta_2), \cdots, \mathbf{a}_{s1}(\theta_K)]$ is the direction matrix of the continuous part, where

$$\mathbf{a}_{s1}(\theta_k) = \left[e^{-j(LM-1)N\pi\,\sin\,\theta_k}, \cdots, e^{-jN\pi\,\sin\,\theta_k}, 1, e^{jN\pi\,\sin\,\theta_k}, \cdots, e^{j(LM-1)N\pi\,\sin\,\theta_k}\right]^T, k = 1, \cdots, K. \qquad (6)$$

Similar with the steps from Eq. (3) to Eq. (5), we can obtain another virtual array from the overall output of subarray 2 and subarray 3, which can be expressed as

$$\mathbf{r}_{s2} = \mathbf{A}_{s2}\mathbf{p} + \sigma^2\mathbf{e}_2, \qquad (7)$$

where $\mathbf{e}_2$ denotes a column vector after the selecting operation and $\mathbf{A}_{s2} = [\mathbf{a}_{s2}(\theta_1), \mathbf{a}_{s2}(\theta_2), \cdots, \mathbf{a}_{s2}(\theta_K)]$ is the direction matrix corresponding to an $(2LN - 1)$-element array located from $-(LN - 1)Md$ to $(LN - 1)Md$ with interelement spacing being Md. The steering vector is

$$\mathbf{a}_{s2}(\theta_k) = \left[e^{-j(LN-1)M\pi\,\sin\,\theta_k}, \cdots, e^{-jM\pi\,\sin\,\theta_k}, 1, e^{jM\pi\,\sin\,\theta_k}, \cdots, e^{j(LN-1)M\pi\,\sin\,\theta_k}\right]^T, k = 1, \cdots, K. \qquad (8)$$

Now, the two virtual coprime subarrays in Eq. (5) and Eq. (7) are obtained, and the large interelement spacing will result in parameter estimation ambiguity problem. However, our method will in turn exploit the phase ambiguity to reduce the complexity and then eliminate the ambiguity based on the coprime-ness between M and N. Suppose $M > N$, then the first subarray in Eq. (5) achieves larger aperture than that in Eq. (7), so we choose $\mathbf{r}_{s1}$ to estimate the DOA for better estimation performance.

As the virtual output has only one snapshot, sparse representation framework will be established to avoid the aperture loss caused by the spatial smoothing [24]. Discretize the whole spatial range as a grid $\widetilde{\theta}_1, \widetilde{\theta}_2, \cdots, \widetilde{\theta}_P(P \gg K)$, and suppose that all the true DOAs fall in the grid, i.e., the dictionary $\mathbf{\Omega} = [\mathbf{a}_{s1}(\widetilde{\theta}_1), \cdots, \mathbf{a}_{s1}(\widetilde{\theta}_P), \mathbf{e}_1]$ contains the columns of $\mathbf{A}_{s1}$, then Eq. (5) can be rewritten in a sparse form as

$$\mathbf{r}_{s1} = \mathbf{\Omega}\boldsymbol{\rho}, \qquad (9)$$

where $\boldsymbol{\rho} \in \mathbb{C}^{(P+1)\times 1}$ is a sparse vector, whose elements corre-

sponding to the true DOAs keep the same with those in $\mathbf{p}$ and last element is noise power σ^2. After sparse recovery, the positions of nonzero elements (except the last element) in $\boldsymbol{\rho}$ will give the estimations of the DOAs. However, as the interelement spacing of the virtual array is Nd, which is larger than half-wavelength, then there are phase ambiguities in $\mathbf{\Omega}$. To clearly elaborate this problem, let $z_p = e^{-jN\pi\,\sin\,\widetilde{\theta}_p}$; then, z_p determines the uniqueness of $\mathbf{a}_{s1}(\widetilde{\theta}_p)$ due to the Vandermonde structure, i.e., if $z_p = z_q$, then $\mathbf{a}_{s1}(\widetilde{\theta}_p) = \mathbf{a}_{s1}(\widetilde{\theta}_q)$.

As $N > 1$, there is a cyclic phase ambiguity in z_p. Except for $\widetilde{\theta}_p$, there are other $(N - 1)$ angles $\widetilde{\theta}_{p,n}, n = 2, \cdots, N$ satisfying

$$e^{-jN\pi\,\sin\,\widetilde{\theta}_{p,n}} = z_p, n = 2, \cdots, N. \qquad (10)$$

It can be derived from Eq. (10) that the relationship between $\widetilde{\theta}_p$ and $\widetilde{\theta}_{p,n}, n = 2, \cdots, N$ is

$$\sin \widetilde{\theta}_{p,n} = \sin \widetilde{\theta}_p - \frac{2m}{N}, n = 2, \cdots, N, \qquad (11)$$

where m is an integer making $\sin \widetilde{\theta}_{p,n}$ locate at the range [-1, 1]. If angles $\widetilde{\theta}_{p,n}, n = 2, \cdots, N$ are also located in the grid $\widetilde{\theta}_1$, $\widetilde{\theta}_2, \cdots, \widetilde{\theta}_P$, then $\widetilde{\theta}_p$ and $\widetilde{\theta}_{p,n}, n = 2, \cdots N$ will provide N identical columns in the dictionary matrix $\mathbf{\Omega}$ due to Eq. (10). This will not only cause the estimation ambiguity but may also make the sparse recovery fail. However, we can in turn exploit the ambiguity to reduce the size of the dictionary and then reduce the complexity of sparse recovery accordingly.

From Eq. (11), it is shown that the N solutions $\sin \widetilde{\theta}_p$ and $\sin \widetilde{\theta}_{p,n}, n = 2, \cdots, N$ are uniformly distributed among the range [-1, 1] following a circle $2/N$. An example is shown in Figure 2, where $\sin \widetilde{\theta}_p = 0.5$ and $N = 3$, then other two solutions are $\sin \widetilde{\theta}_{p,2} = -1/6$ and $\sin \widetilde{\theta}_{p,3} = -5/6$, respectively. It is also shown by the dashed lines in Figure 2 that if we divide the whole range into N cycles with width being $2/N$, then there is only one solution in one cycle based on Eq. (11). As these solutions provide identical atoms in the dictionary, we can choose one cycle as a representative to construct the dictionary. Without loss of generality, we choose range $[-1/N, 1/N]$, and the angle range is $[\arcsin(-1/N), \arcsin(1/N)]$, whose corresponding dictionary is denoted by $\mathbf{\Omega}_{\mathrm{sub}}$; then, the sparse form in Eq. (9) becomes

$$\mathbf{r}_{s1} = \mathbf{\Omega}_{\mathrm{sub}} \mathbf{\rho}, \qquad (12)$$

where $\mathbf{\rho} \in \mathbb{C}^{(P+1) \times 1}$ is a K-sparse vector, whose elements corresponding to the true steering vectors (maybe not the true DOAs) keep the same with those in $\mathbf{p}$, and the others are zero (except the last element). For example, as shown in Figure 2, the true solution is 0.5, and the representative solution in the partial dictionary $\mathbf{\Omega}_{\mathrm{sub}}$ is -1/6.

Due to Eq. (12), the phase ambiguity in the dictionary can be avoided now. Meanwhile, as now the size of the dictionary $\mathbf{\Omega}_{\mathrm{sub}}$ is only $1/N$ of its original size, the computation complexity of sparse recovery can be reduced.

3.2. Off-Grid Sparse Representation Framework. Now a partial dictionary based sparse representation framework is established, but it is built based on the assumption that the true DOAs or their representative angles are located in the grid. However, the angles are very likely to lie off the discretized grid, no matter how fine the grid is defined. Off-grid sources will bring in grid mismatch problem and degrade the sparse recovery performance significantly. In this section, we take the off-grid problem into account and reformulate the sparse representation to enhance the robustness to grid mismatch.

Within the range $[\arcsin(-1/N), \arcsin(1/N)]$, we denote the uniformly sampled grid as $\widetilde{\theta}_1, \widetilde{\theta}_2, \cdots, \widetilde{\theta}_Q$ with adjacent interval being g. Then, the true DOA or its representative angle $\overline{\theta}_k$ can be represented by a nearest grid $\widetilde{\theta}_{q,k}$ plus an offset α_k, which is among the range $[-g/2, g/2]$. Based on first order Taylor expansion around the grid [14], the true steering vector can be approximately expressed as
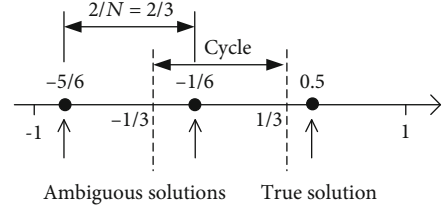
FIGURE 2: The relationship between the real and ambiguous solutions.

$$\mathbf{a}_{s1}\left(\overline{\theta}_k\right) \approx \mathbf{a}_{s1}\left(\widetilde{\theta}_{q,k}\right) + \frac{\partial \mathbf{a}_{s1}\left(\widetilde{\theta}_{q,k}\right)}{\partial \widetilde{\theta}_{q,k}} \alpha_k, \qquad (13)$$

where $\alpha_k = \overline{\theta}_k - \widetilde{\theta}_{q,k}$. Then, Eq. (12) is revised as

$$\mathbf{r}_{s1} = \left(\mathbf{\Omega}_{\mathrm{sub}} + \mathbf{\Omega}'_{\mathrm{sub}} \mathbf{\Lambda}\right) \mathbf{\rho}, \qquad (14)$$

where $\mathbf{\Omega}'_{\mathrm{sub}} = [\partial \mathbf{a}_{s1}(\widetilde{\theta}_1)/\partial \widetilde{\theta}_1, \cdots, \partial \mathbf{a}_{s1}(\widetilde{\theta}_Q)/\partial \widetilde{\theta}_Q, \mathbf{e}_1]$, $\mathbf{\Lambda} = \mathrm{diag}$ $(\mathbf{\beta})$, and

$$\mathbf{\beta}(q) = \begin{cases} \alpha_k, \mathrm{if} \widetilde{\theta}_{q,k} = \widetilde{\theta}_q, k = 1, \cdots, K \\ 0, \mathrm{others} \end{cases}, q = 1, \cdots, Q. \qquad (15)$$

Let $\mathbf{\omega} = \mathbf{\Lambda}\mathbf{\rho}$, then it is easy to verify that $\mathbf{\omega}$ and $\mathbf{\rho}$ are joint sparse [30]. So the off-grid sparse formulation can be expressed as

$$\min \|\mathbf{h}\|_{2,1}$$
$$\mathrm{s.t.} \ \mathbf{r}_{s1} = \mathbf{\Omega}_{\mathrm{sub}} \mathbf{\rho} + \mathbf{\Omega}'_{\mathrm{sub}} \mathbf{\omega} \qquad (16)$$
$$-\frac{g}{2} \mathbf{\rho} \le \mathbf{\omega} \le \frac{g}{2} \mathbf{\rho},$$

where $\mathbf{h} = [\mathbf{\rho}^T, \mathbf{\omega}^T]^T$, and $\|\mathbf{h}\|_{2,1} = \sum_{i=1}^Q \sqrt{\mathbf{\rho}_i^2 + \mathbf{\omega}_i^2}$, where $\mathbf{\rho}_i$ means the i-th element of $\mathbf{\rho}$. It should be noted that the covariance matrix in Eq. (3) can only be estimated via finite snapshots

$$\widehat{\mathbf{R}}_1 = \left(\frac{1}{T}\right) \sum_{t=1}^T \mathbf{y}_1(t) \mathbf{y}_1^H(t), \qquad (17)$$

where T denotes the snapshot number. So the sparse form in Eq. (16) is not robust due to the residual error. Use $\Delta \mathbf{r}_1$ to denote the deviation of $\mathbf{r}_1$ in Eq. (6), then according to [37], $\Delta \mathbf{r}_1$ follows asymptotic normal distribution with zero mean and covariance matrix being $1/T(\mathbf{R}_1^T \otimes \mathbf{R}_1)$

$$\Delta \mathbf{r}_1 \sim \mathrm{As}N\left(0, \frac{1}{T}\left(\mathbf{R}_1^T \otimes \mathbf{R}_1\right)\right). \qquad (18)$$
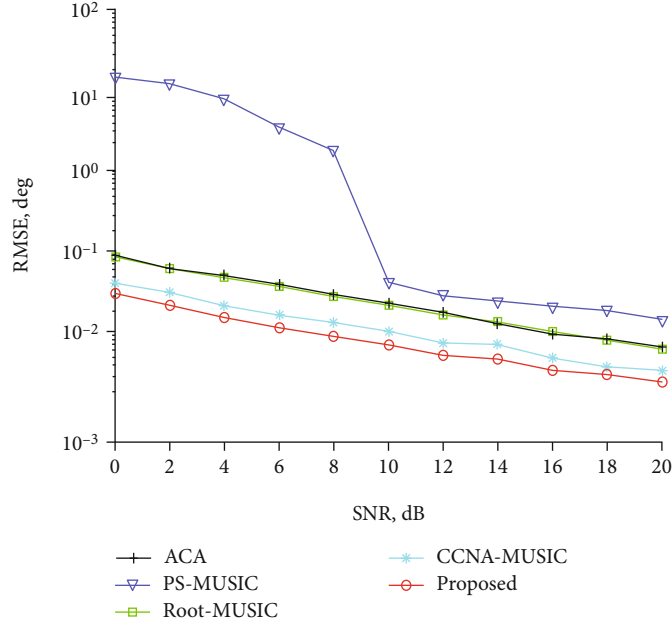
FIGURE 3: Angle estimation accuracy comparison versus SNR.

After selecting operation

$$\Delta \mathbf{r}_{s1} \sim \mathrm{A}s N\left(0, \frac{1}{T}\mathbf{W}_{s1}\left(\mathbf{R}_1^T \otimes \mathbf{R}_1\right)\mathbf{W}_{s1}^{\ H}\right). \qquad (19)$$

Define weight matrix $\mathbf{W} = (1/T)\mathbf{W}_{s1}(\mathbf{R}_1^T \otimes \mathbf{R}_1)\mathbf{W}_{s1}^{\ H}$, then

$$\mathbf{W}^{-1/2}\Delta \mathbf{r}_{s1} \sim \mathrm{A}s N(0, \mathbf{I}_{2LM-1}). \qquad (20)$$

From Eq. (20), $\|\mathbf{W}^{-1/2}\Delta \mathbf{r}_{s1}\|_2^2$ follows an asymptotic chi-square distribution with $2LM - 1$ DOF. Consequently, the enhanced sparse recovery problem can be formulated as

$$\min. \|\mathbf{h}\|_{2,1}$$
$$\mathrm{s.t.} \left\|\mathbf{W}\wedge^{-1/2}\left(\mathbf{r}_{s1} - \Omega_{\mathrm{sub}}\rho - \Omega'_{\mathrm{sub}}\omega\right)\right\|_2 \leq \xi \qquad (21)$$
$$-\frac{g}{2}\rho \leq \omega \leq \frac{g}{2}\rho,$$

where $\widehat{\mathbf{W}} = (1/T)\mathbf{W}_{s1}(\widehat{\mathbf{R}}_1^T \otimes \widehat{\mathbf{R}}_1)\mathbf{W}_{s1}^{\ H}$ is the approximate weight matrix; ξ is the up bound of the fitting error, which can be set as $\xi = \sqrt{\mathrm{chi2inv}(1-p, 2LM-1)}$ [11], where chi2 inv$(1-p, 2LM-1)$ denotes the inverse cumulative distribution function that makes the inequality holds with a probability $(1-p)$. Generally, it is enough to set $p = 0.001$ to make it nearly a sure event.

After solving Eq. (21) via CVX [38, 39], we can obtain the estimations of ρ and ω, which are denoted as $\widehat{\rho}$ and $\widehat{\omega}$, respectively.

3.3. Ambiguity Elimination. The positions of nonzero elements in the first Q elements of $\widehat{\rho}$ and $\widehat{\omega}$ give the initial DOA estimations $\widetilde{\theta}_{q,k}, k = 1, \cdots, K$, which are the grids near-

est to the true DOAs or their representative angles. Besides, the offset vector can be obtained via

$$\beta = \mathrm{diag}\,(\widehat{\rho}./\widehat{\omega}), \qquad (22)$$

where ./ means element-wise division. Then, the angles are obtained via

$$\overline{\theta}_k = \widetilde{\theta}_{q,k} + \alpha_k, k = 1, \cdots, K, \qquad (23)$$

where the offsets $\alpha_k, k = 1, \cdots, K$ are obtained from the first Q elements of β in Eq. (22).

Now, the angles are estimated with offsets being compensated, but the angles in Eq. (23) may be true DOAs and also may be representative angles. As been discussed in Eq. (10) and Eq. (11), there are totally N angles including $\overline{\theta}_k$ sharing the same atom, and their relationship is

$$\sin \overline{\theta}_{k,n} = \sin \overline{\theta}_k - \frac{2m}{N}, n = 2, \cdots, N, \qquad (24)$$

where m is an integer making $\sin \overline{\theta}_{k,n}$ locate at the range $[-1, 1]$.

To determine the unique DOA without ambiguity, we substitute the N angles in Eq. (24) into the whole virtual array

$$\max \mathbf{a}_n^{\ H}\mathbf{r}, \qquad (25)$$

where $\mathbf{r} = [\mathbf{r}_{s1}^{\ T}, \mathbf{r}_{s2}^{\ T}]^T$ and $\mathbf{a}_n = [\mathbf{a}_{s1}^{\ T}(\overline{\theta}_{k,n}), \mathbf{a}_{s2}^{\ T}(\overline{\theta}_{k,n})]^T$. Due to the coprime relationship between the two subarrays, the unique DOA can be determined from the coincide results from the two subarrays. Consequently, if the whole array containing both two subarrays is exploited, unique angle is determined by finding the maximum value in Eq. (25).
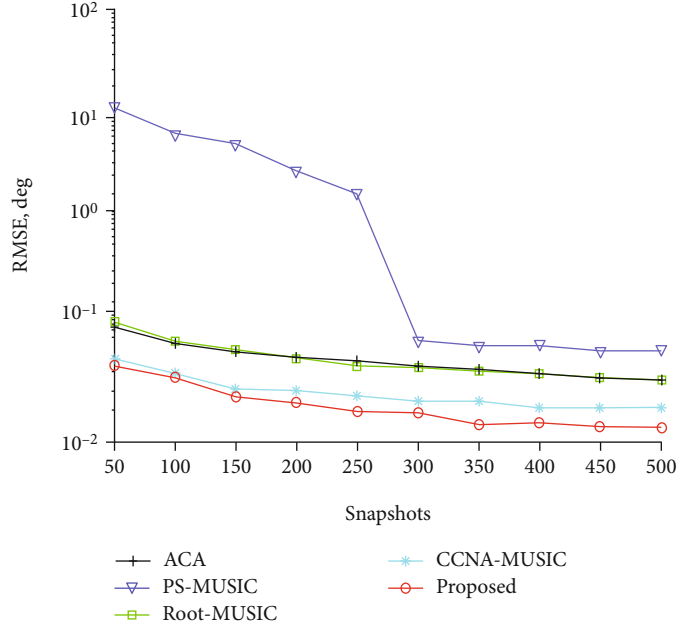
FIGURE 4: Angle estimation accuracy comparison versus snapshot number (SNR = 10 dB).
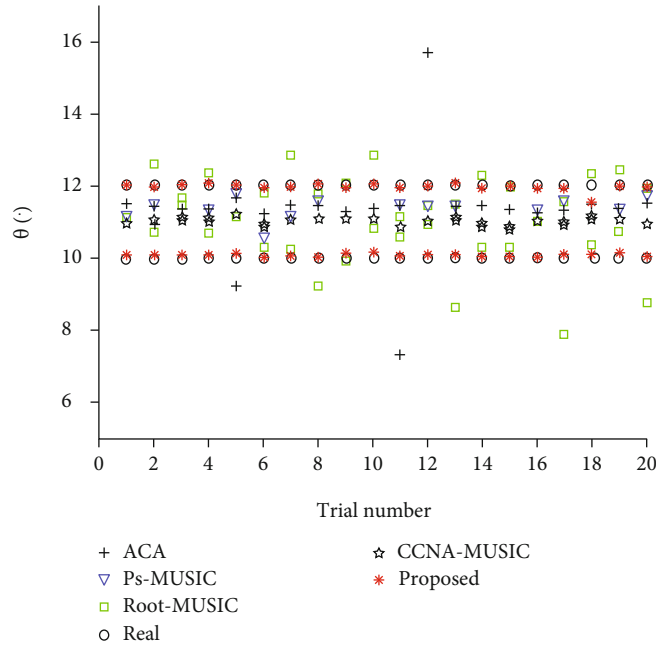
FIGURE 5: Angle estimation results of closely spaced sources (SNR = 0 dB).

For the complexity, the proposed method only requires partial dictionary, and the main complexity lies in the construction of two covariance matrices and sparse recovery. The total number of complex multiplications is about $((M + L - 1)^2 + (N + L - 1)^2)T + (2LM - 1)n^3 + N(2LM + 2LN - 2)$, where n denotes the dictionary size. The proposed method has lower complexity than peak search method [24] and other sparse representation methods that require whole dictionary [27–30]. Compared to DOA estimation methods with closed-form solution, e.g., ESPRIT and root-MUSIC, the proposed method costs more but achieves better estimation performance, which will be verified in the simulation section below.

4. Simulation Results

In the simulations, the CCNA is configured with $M = 4$, $N = 3$, and $L = 3$. $T = 500$ snapshots are collected to estimate the covariance matrix, and the root mean square error (RMSE) is defined below to measure the DOA estimation performance

$$\text{RMSE} = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{D}\sum_{i=1}^{D}(\theta\wedge_{k,i} - \theta_k)^2]}, \qquad (26)$$

where $\widehat{\theta}_{k,i}$ denotes the estimations of θ_k of the i-th Monte Carlo trial and $D = 200$ trials are carried out.

With the measurement of RMSE, Figures 3 and 4 present the angle estimation accuracy comparisons between the proposed algorithm and other methods versus SNR and snapshot number, respectively. The ACA method [24], PS-MUSIC using prototype coprime array [32], root-MUSIC method using prototype coprime array [33], and root-MUSIC using CCNA [36] all adopt the same number of physical antennas with the proposed method for fair comparison. It is indicated from Figures 3 and 4 that the proposed algorithm outperforms the other methods, and the main reasons include (1) the virtual subarray with larger aperture is chosen to avoid the negative effect from the smaller subarray, and (2) the off-grid sparse representation is established to amend the grid mismatch problem. The PS-MUSIC has the worst performance, especially with low SNR, as it has the additional pairing problem, and it utilizes the data from the physical array, which has limited DOF.

To test the resolution performance of the algorithms, we choose two closely spaced sources with angles being $\theta_1 = 10°$ and $\theta_2 = 12°$, respectively. Figure 5 shows the estimation results of the algorithms over 20 trials with SNR = 0 dB. It is indicated that the proposed method can always clearly identify the two sources, while the other methods have big deviations. Consequently, the proposed method achieves the best angular resolution.

5. Conclusions

An off-grid DOA estimation method exploiting CCNA is proposed. Based on the nested relationships within the three subarrays, two virtual coprime subarrays are obtained firstly in the coarray domain. Thereafter, subarray with larger aperture is chosen for enhanced estimation performance, and cyclic phase ambiguity is exploited to reduce the size of the dictionary. Meanwhile, off-grid sparse reconstruction method is established to amend the grid mismatch. Finally, DOA is uniquely determined by substituting the ambiguous into the whole array. Compared to other methods with simulations, the proposed approach is verified that it has better DOA estimation performance and angular resolution.

Data Availability

Data are available in the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

References

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[2] L. Wan, G. Han, L. Shu, S. Chan, and T. Zhu, "The application of DOA estimation approach in patient tracking systems with high patient density," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2353–2364, 2016.

[3] F. Wen, Z. Zhang, K. Wang, G. Sheng, and G. Zhang, "Angle estimation and mutual coupling self-calibration for ULA-based bistatic MIMO radar," *Signal Processing*, vol. 144, pp. 61–67, 2018.

[4] X. Wang, L. Wan, M. Huang, C. Shen, Z. Han, and T. Zhu, "Low-complexity channel estimation for circular and noncircular signals in virtual MIMO vehicle communication systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 3916–3928, 2020.

[5] X. Wang, L. Yang, D. Meng, M. Dong, K. Ota, and H. Wang, "Multi-UAV cooperative localization for marine targets based on weighted subspace fitting in SAGIN environment," *IEEE Internet of Things Journal*, 2021.

[6] D. Meng, X. Wang, and M. Huang, "Robust weighted subspace fitting for DOA estimation via block sparse recovery," *IEEE Communications Letters*, vol. 24, no. 3, pp. 563–567, 2020.

[7] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[8] J. Li, Z. Li, and X. Zhang, "Partial angular sparse representation based DOA estimation using sparse separate nested acoustic vector sensor array," *Sensors*, vol. 18, no. 12, pp. 4465–4479, 2018.

[9] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[10] N. Hu, Z. Ye, D. Xu, and S. Cao, "A sparse recovery algorithm for DOA estimation using weighted subspace fitting," *Signal Processing*, vol. 92, no. 10, pp. 2566–2570, 2012.

[11] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4489–4493, 2011.

[12] P. Stoica, P. Babu, and J. Li, "SPICE: a sparse covariance-based estimation method for array processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, 2011.

[13] L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[14] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, 2011.

[15] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4997–5008, 2014.

[16] X. Wang, L. Wang, X. Li, and G. Bi, "Nuclear norm minimization framework for DOA estimation in MIMO radar," *Signal Processing*, vol. 135, pp. 147–152, 2017.

[17] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[18] P. Pal and P. P. Vaidyanathan, "Nested arrays: a novel approach to array processing with enhanced degrees of freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, 2010.

[19] J. Li, P. Ma, and X. Zhang, "Improved DFT algorithm for 2D DOA estimation based on 1D nested array motion," *IEEE Communications Letters*, vol. 24, no. 9, pp. 1953–1956, 2020.

[20] J. He, L. Li, and T. Shu, "Sparse nested arrays with spatially spread orthogonal dipoles: high accuracy passive direction finding with less mutual coupling," *IEEE Transactions on Aerospace and Electronic Systems*, p. 1, 2021.

[21] J. Shi, F. Wen, and T. Liu, "Nested MIMO radar: coarrays, tensor modeling, and angle estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, pp. 573–585, 2021.

[22] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.

[23] X. Wang, M. Huang, and L. Wan, "Joint 2D-DOD and 2D-DOA estimation for coprime EMVS–MIMO radar," *Circuits, Systems, and Signal Processing*, vol. 40, pp. 2950–2966, 2021.

[24] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the music algorithm," in *2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, pp. 289–294, Sedona, AZ, USA, January 2011.

[25] S. Qin, Y. D. Zhang, and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1377–1390, 2015.

[26] W. Liu, A. R. Raza, and Q. Shen, "Thinned coprime array for second-order difference co-array generation with reduced mutual coupling," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2052–2065, 2019.

[27] P. Pal and P. P. Vaidyanathan, "Correlation-aware sparse support recovery: Gaussian sources," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5880–5884, Vancouver, BC, Canada, May 2013.

[28] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based DOA estimation using co-prime arrays," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3967–3971, Vancouver, BC, Canada, May 2013.

[29] C. Zhou, Z. Shi, Y. Gu et al., "DOA estimation by covariance matrix sparse reconstruction of coprime array," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2369–2373, South Brisbane, QLD, Australia, April 2015.

[30] Z. Tan and A. Nehorai, "Sparse direction of arrival estimation using co-prime arrays with off-grid targets," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 26–29, 2013.

[31] C. Zhou, Z. Shi, Y. Gu, and X. Shen, "DECOM: DOA estimation with combined MUSIC for coprime array," in *2013 Inter-*

national Conference on Wireless Communications and Signal Processing, pp. 1–5, Hangzhou, China, October 2013.

[32] F. Sun, P. Lan, and B. Gao, "Partial spectral search-based DOA estimation method for co-prime linear arrays," *Electronics Letters*, vol. 51, no. 24, pp. 2053–2055, 2015.

[33] D. Zhang, Y. Zhang, G. Zheng, C. Feng, and J. Tang, "Improved DOA estimation algorithm for co-prime linear arrays using root-MUSIC algorithm," *Electronics Letter*, vol. 53, no. 18, pp. 1277–1279, 2017.

[34] J. Li, D. Jiang, and X. Zhang, "DOA estimation based on combined unitary ESPRIT for coprime MIMO radar," *IEEE Communications Letters*, vol. 21, no. 1, pp. 96–99, 2017.

[35] J. Li, Y. He, P. Ma, X. Zhang, and Q. Wu, "Direction of arrival estimation using sparse nested arrays with coprime displacement," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5282–5291, 2021.

[36] J. Li, Y. Li, and X. Zhang, "Direction of arrival estimation using combined coprime and nested array," *Electronics Letters*, vol. 55, no. 8, pp. 487–489, 2019.

[37] J. Li, D. Jiang, and X. Zhang, "Sparse representation based two-dimensional direction of arrival estimation using co-prime array," *Multidimensional Systems and Signal Processing*, vol. 29, no. 1, pp. 35–47, 2018.

[38] G. Michael and B. Stephen, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," 2013, http://cvxr.com/cvx.

[39] G. Michael and B. Stephen, "Graph implementations for non-smooth convex programs," in *Recent Advances in Learning and Control. Lecture Notes in Control and Information Sciences, vol 371*, V. D. Blondel, S. P. Boyd, and H. Kimura, Eds., pp. 95–110, Springer, London, 2008.

WILEY | Hindawi

Research Article

SDN Controller Deployment for QoS Guarantees in Tactical Ad Hoc Networks

Xin Yan [ID],[1] Xiaodong Hu,[2] and Wen Liu[1]

[1]*School of Computer Science & Technology, Wuhan University of Technology, Wuhan 430070, China*
[2]*School of Mathematics & Computer Science, Wuhan Textile University, Wuhan 430200, China*

Correspondence should be addressed to Xin Yan; yanxin@whut.edu.cn

Tactical ad hoc networks (TANET) accomplish the corresponding tasks via a hopeful device-to-device connection mechanism for data transmission and resource management without a centralized foundation. Software-defined networking (SDN) provides an evolution from the previous networks by decoupling the network control from data forwarding and providing a novel paradigm for network handling. Nevertheless, a SDN-based strategy in TANET leads to various novel problems since the primary construction cannot be employed in mobile ad hoc networks anymore. In this paper, a new SDN-based structure is constructed for TANET. Then, both delay and energy consumption (EC) are utilized to model the controller deployment and data-plane assignment problems, by which quality-of-service (QoS) guarantees are realized. Afterwards, mixed-integer programming (MIP) is adopted to solve the mentioned model. Finally, according to the experimental outcomes, it can be observed that the presented approach ensures the data-plane delay and optimizes the EC.

1. Introduction

With the increasing growth of the technology of wireless communications, data traffic transmitted through wireless networks is continuously increasing. Wireless mobile communication networks commonly contain wireless local area networks, cellular networks, mobile ad hoc networks, wireless mesh networks, etc. Local traffic may be caused when the source and the destination are close. In popular network situations, local traffic transmission and redistribution can be realized from the cellular network to the main structure. As a more suitable strategy, a distributed organization for devices can be realized using device-to-device technologies. Lack of centralized foundation leads to independent operation, making routing decisions, and dynamic adaptation of each node to topology variations in a mobile network. The mentioned features lead to significant achievements for mobile ad hoc networks in a tactical network scenario, in other words, tactical ad hoc networks (TANET) [1].

As a hopeful novel pattern, software-defined networking (SDN) makes a remarkable change in the construction and operation of communication networks. Due to its universal view on the whole network, the centralized SDN controller leads to a remarkable improvement in its service quality and management of resources. SDN adopts task-specific benchmark and necessary to know restrictions within the tactical networks to make globally optimal routing decisions and provide data transmission. The SDN has been employed in ad hoc networks to improve efficiency and flexibility [2]. In comparison to the traditional distributed routing procedures (like optimized link state routing, OLSR), the SDN-enabled structure leads to lower response delay, superior throughput, and energy consumption (EC).

SDN provides an improvement in management intelligence and flexible control in TANET. Nevertheless, due to the distribution of resources and management in mobile ad hoc networks, employing a centralized SDN pattern in TANET (i.e., software-defined tactical ad hoc networks, SD-TANET) might be complicated. For the SDN unit's proper operation, specific communications should be established among controllers and data-plane services using network state data (e.g., topology discovery) and flow tables.

The mentioned communications can be influenced in TANET due to the low transfer rates and unreliable link connection. Due to the intrinsic limitations for devices in SD-TANET, the controllers should be distributed within the network to achieve superior performance and reliability.

A redundancy is incorporated into the control plane to overcome the mentioned deficiencies and guarantee the connection reliability among data-plane nodes and controllers. In a large-scale SD-TANET, multiple controller deployment can be considered an appropriate approach. Due to the critical role of the SDN controller within the network and the movement of the energy-intensive functions from nodes to the controller, the SDN controller's optimum deployment has a considerable impact on decreasing the response delay and EC. An essential issue here is to decide where to locate multiple controllers in the network. This decision is able to significantly affect the efficiency of the SDN-based TANET.

The other sections of the present article are classified as given below: Section 2 is allocated to the relevant studies, which employ SDN in mobile ad hoc networks and the issues of wireless controller deployment. In Section 3, the presented design and SDN-based TANET are introduced, which are adopted to describe the deployment issue of controllers and the data-plane assignment to the controllers in Section 4. The solution of the mentioned model is given in Section 5, and the presented approach is evaluated through a practical dataset. The obtained conclusions are drawn in Section 6.

2. Relevant Studies

Current controller software, including ONOS [3] and the software switch Open vSwitch [4], could be implemented on various instruments, even certain weightless network ones. This makes it feasible to employ SDN in mobile ad hoc networks. Besides, various controller deployment approaches can be realized using common software execution.

In [5], the authors introduced a feasible implementation of an SDN-based MANET (namely, mobile ad hoc networks) and characterized the developed software components. The authors in [6] constructed a framework and a related prototype for SDN-based quality management, which provides high flexibility via novel flow management rules at preparation time and can appropriately handle node join/leave events. We see in [1] a structure constructed for SDN-based mobile ad hoc networks in the tactical area and illustrated the difficulties caused via the ad hoc and coalition network ambience. In [7], the technology of SDN has been employed for the ad hoc sensor networks of flying, and a clustering structure with a clustered controller of SDN is constructed to realize hierarchical management and integrated dispatch. In [8], a flexible procedure has been presented, which could dynamically select to route under the controller's commands or make routing choices through a distributed approach. The mentioned studies demonstrate the possi-

bility of employing SDN in ad hoc networks and its functionality aspects in tactical field networks.

Controller deployment problem (CPP) can spread out a fair quantity of controllers at optimum positions. Several studies have been performed about wired networks based on various targets and limitations. Nodal mobility and channel uncertainty are two essential features of the wireless network, imposing unique CPP in software-defined wireless networks (SDWN) [9]. The CPP issue has been considered in SDWN in various studies.

The authors in [10] presented a TDMA-based controller, in which wireless control plane and southbound interfaces have been considered to diminish the quantity of controllers in a specified controller reaction time. We see in [11] a novel assessment named transparency and optimized multiple goals taking channel conflict and packet loss in wireless networks. In [12], the authors verified the influence of delayed state data and constructed a deployment structure of dynamic controller, where delayed queue length data has been utilized to relocate the controller.

The data mentioned above give us the possibilities of not only employing SDN in TANET but also implementing CPP with wireless links. Although the CPP with delay or packet loss guarantee in a wireless scenario has been studied, the issue with multiple QoS constraints has not been involved yet. In particular, EC constraint is a crucial QoS guarantee for both controllers and data nodes in wireless mobile networks. Thus, in addition to channel conflict, both delay and EC are considered in this paper, which is our major contribution as well.

3. The SD-TANET Structure

The constructed structure of SDN-based TANET and their operation is illustrated in the current section. An appropriate MAC layer should be selected in the data plane for the TANET organization. Unlicensed bands without centralized coordination determine the mentioned selection rule for the MAC protocol. Based on [5], IEEE 802.11 P2P mode independent basic service set (IBSS) is the ad hoc mode standard of IEEE 802.11, providing complete control on the upper layers, which is necessary for the utilized method.

Although the SDN is founded on the decoupling of the data and control planes, it is more reasonable to employ a hybrid structure in the SD-TANET for the control plane. In a hybrid structure, data transmission via data-plane nodes could be realized using the flow tables transmitted via controllers. If the target node is undetectable or the flow tables are not updated by the controller, the old flow tables could be neglected via the data-plane nodes, and messages can be transmitted via distributed routing rules. This design kind causes data-plane nodes to create distributed routing choices to improve the control plane's redundancy.

Further necessity in this architecture is to employ two distinct frequency bands to split network control and data sending, enhancing the scalability and performance. The
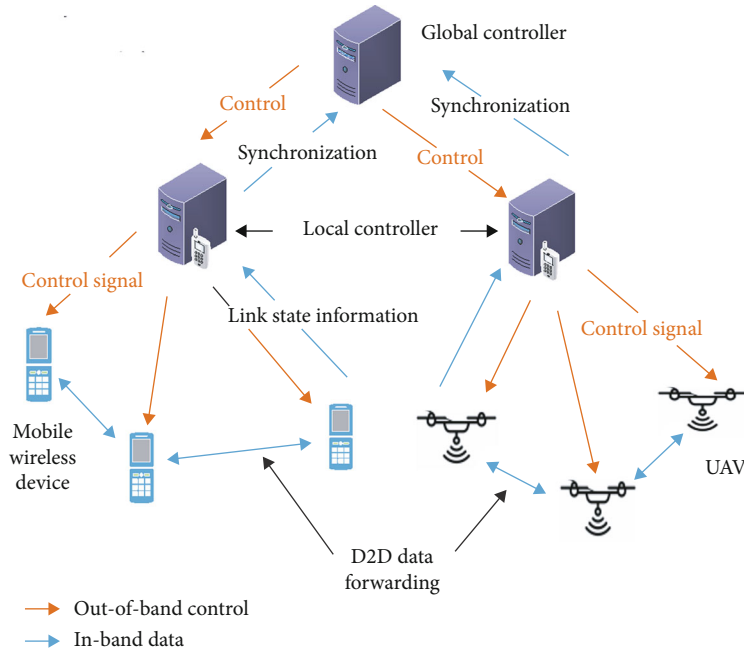
FIGURE 1: The SD-TANET structure, where the mobile wireless instruments and unmanned aerial vehicles (UAV) with SDN capability create the data plane.

Input: The **n**etwork topology, the objective function and the constraints
Output: The optimal controller location and assignment $<x, Y>$
1. $C \leftarrow \Phi$
2.   Initialize x, Y
3.   For $n=1$ to N_c do
4. $x \leftarrow$ place(x_n) //calculated by CPLEX
5. $Y \leftarrow$ assign(x_n, y_{sn}) //optimal assignment
6.   Calculate $\alpha D(y) + E(x, y)$
7.   End
8. $C \leftarrow \arg\ \min_{x,y} \alpha D(y) + E(x, y)$
9.   Return $<x, Y>$
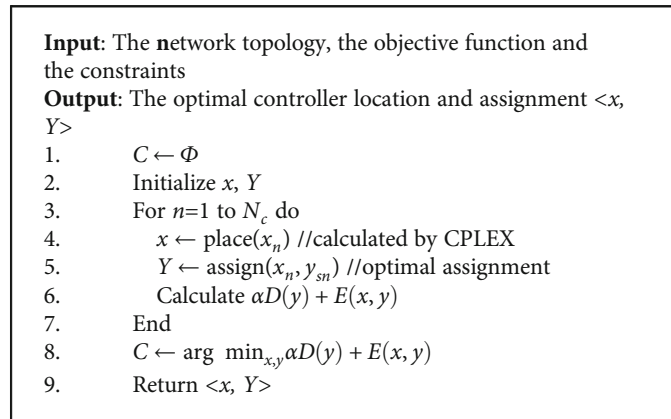
ALGORITHM 1: The controller deployment and data-plane assignment.

"in-band" phrase is employed for point-to-point data transfer among distributed devices, while the "out-of-band" phrase is utilized for point-to-multipoint control signaling among the controller and distributed devices [13, 14]. Figure 1 describes the details of the hybrid structure, which consists of a global controller and some local controllers, forming the control plane.

An instrument with adequate processing of data abilities is able to operate as the local controller, deployed on a portable wireless foundation or SDN-based mobile instruments. It combines link-state data with local topology from all corresponding distributed instruments. The global controller provides a general perspective of the network and preprocesses link-state data with a centralized strategy. It can be spread out in an edge computing center like the center of commanding in the networks of tactical

field. The local controller operates as a link connecting the global controller to distributed devices. Optimal routing tables obtained via the global controller are sent via the local controller. The local controller can perform the overall control under a global controller's failure or real-time response requirements. The routing of devices can be realized using their processed data.

4. System Model and Strategy

The control plane consists of the controllers' deployment and the data-plane nodes' assignment. The network, especially the response delay and EC for transmission and synchronization, can be effectively influenced by the control-plane structure. The following challenges can be solved using the presented methodology: (1) the number of controllers; (2)
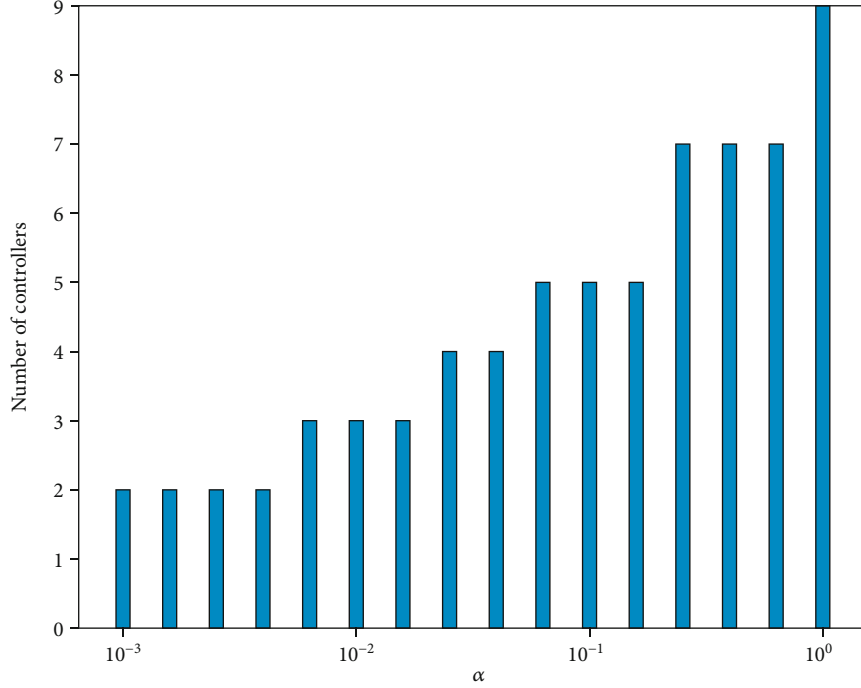
FIGURE 2: Quantity of controllers for various weight values α.
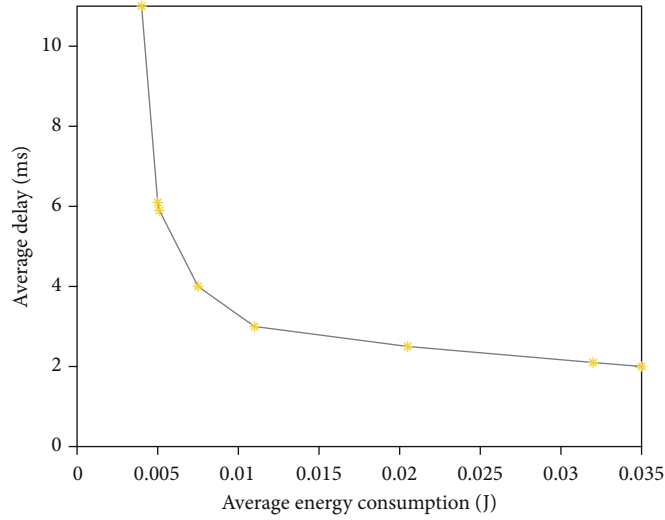
FIGURE 3: Trade-off among EC and delay.

the controller position; and (3) the assignment of data-plane nodes to the controller.

A network is described with $G(N, E)$. Suppose that the data-plane nodes and controllers are indicted with the set N. The wireless connection among data-plane nodes and controller is described by set E.

As shown in Figure 1, the smartphones and unmanned aerial vehicles (UAV) could be described via data-plane nodes if they are SDN-consistent and support the MAC layer explained in the prior section. Virtual SDN switches can be simulated with the network nodes. The mentioned mobile ad hoc nodes and the controller set are denoted by

$S = \{s_1, s_2, \cdots, s_k\}$, and $C = \{c_1, c_2, \cdots, c_t\}$, respectively. Consider that all TANET nodes can operate as a controller. A series of nodes with sufficient resources that are able to operate as a controller is denoted by $N_c \subseteq N$.

Now, $x_c \in \{0, 1\}$ is defined as binary decision variables, as the following (1), indicating whether a local controller is positioned at node $c \in N_c$. The mentioned variables can be employed to develop a controller deployment approach.

$$x = (x_n \in \{0, 1\}: n \in N). \tag{1}$$
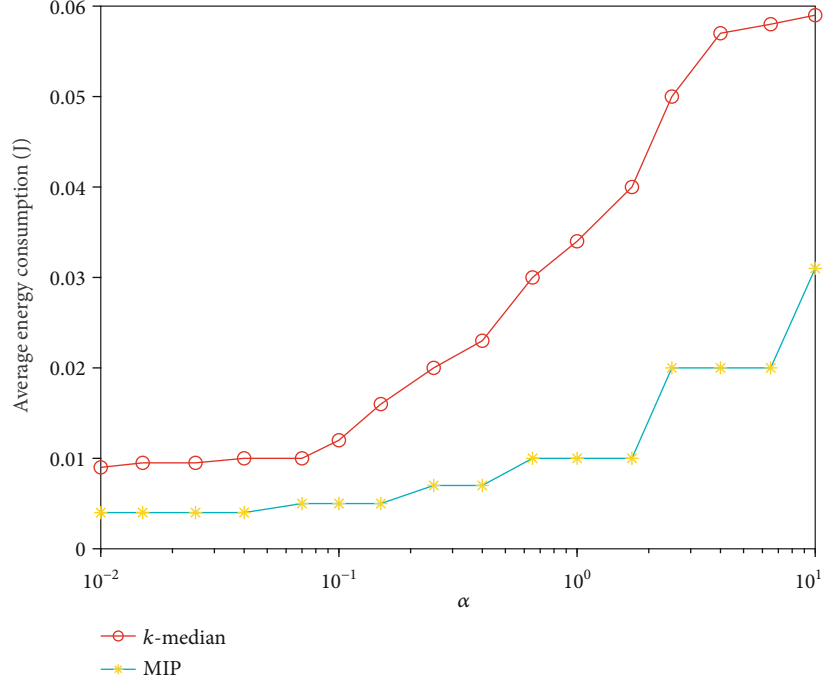
FIGURE 4: Comparison with the k-median approach.

Since a controller should be positioned at a node with sufficient resources, the following condition should be fulfilled:

$$x_n = 0, \quad \forall n \notin N_c. \tag{2}$$

The presented approach should determine the assignment of data-plane nodes to controllers. This implies that an appropriate controller must be selected to manage it. In a similar manner, decision variables $y_{sc} \in \{0, 1\}$ indicate whether a node $s \in S$ corresponds to the controller at $c \in N_c$ and $y_{sc} = 1$ or not $y_{sc} = 0$. The assignment rule is described as

$$y = (y_{sc} \in \{0, 1\}: s \in S, c \in N_c). \tag{3}$$

Since any data-plane node must be devoted to an individual controller at a determined time, the following constraint should be satisfied:

$$\sum_{c \in N_c} y_{sc} = 1, \quad \forall s \in S. \tag{4}$$

As another constraint, if node s corresponds to controller c, controller c should be necessarily positioned there. Thus, the following constraint should be fulfilled:

$$y_{sc} \leq x_c, \quad \forall s, c \in N. \tag{5}$$

The presented controller deployment and assignment strategy's primary purpose is to decrease the reaction delay and the communication overhead for state data collection and synchronization. Then, the model of delay and overhead can be studied.

It can be seen from Figure 1 that control and data planes send control signals and status data through wireless links. Any wireless link is related to probability p_{sc}, which defines the probability that a message can be successfully transmitted on the mentioned link. p_{sc} is dependent on the model of wireless channel that considers a combined path loss and shadowing model, presented through [15].

$$p_{sc} = Q\left(\frac{P_{\min} - (P_t + 10 \lg k - 10\gamma \lg (d/d_0))}{\sigma_{\psi_{dB}}}\right). \tag{6}$$

Q describes the probability that a Gaussian random variable x with the zero mean and the unit variance can be higher than z:

$$Q(z) \triangleq p(x > z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \tag{7}$$

$P_{\min}$ defines the least delivered power, P_t denotes the transmitted power, k indicates a nondimensional parameter that is related to the features of antenna and the mean attenuation of channel, γ denotes the path loss exponent, d describes the distance among the controller and the corresponding data-plane nodes, and d_0 denotes the reference distance for the antenna far field [15]. Their corresponding values could be calculated to approximate either an experimental or analytical model. ψ_{dB} describes a Gaussian-distributed stochastic variable with zero mean and variance $\sigma_{\psi_{dB}}$.

Within the situation where the messages transmitted via the source, including data-plane nodes or controller, was not effectively sent to the receiver, the quantity of retransmission is indicated by n_{sc}, where its distribution is described as

$$P_r\{n_{sc} = i\} = p_{sc}(1 - p_{sc})^{i-1}, \quad i \in \{1, 2, 3, \cdots\}. \tag{8}$$

The nodes controlled by c_i are indicated by s_{i,c_i}. Now, the rate of data transfer of the wireless link $r(c_i, s_{i,c_i})$ could be obtained as

$$r_i(c_i, s_{i,c_i}) = B \log_2 \left(1 + \frac{p_t(c_i)h(c_i, s_{i,c_i})}{\sigma^2 + I(c_i, s_{i,c_i})}\right). \tag{9}$$

η_i describes the communication traffic used by upgrading the flow tables. $h(c_i, s_{i,c_i})$ indicates the channel gain among controller c_i and node s_{i,c_i} given via c_i, where it can be computed as $h = d^{-\gamma}$. B describes the bandwidth of the channel, and σ^2 indicates the white Gaussian noise variance.

The transfer delay among data-plane nodes and controller is given by

$$t_{sc} = \frac{\eta_i}{r_i(c_i, s_{i,c_i})}. \tag{10}$$

Now, the expected transfer delay can be obtained as $2t_{sc}/p_{sc}$.

According to the above analysis, the overall response delay can be obtained as

$$D(y) = \sum_{s \in S} \sum_{c \in N_c} y_{sc} \left(\frac{2t_{sc}}{p_{sc}}\right). \tag{11}$$

As for EC, we mainly assume the energy used by the data exchange among data-plane nodes and controllers and the synchronization among controllers. The mentioned two kinds of EC are influenced by the controller deployment, illustrated individually.

The communication overhead of the assignment of node s to controller c is indicated through w_{sc}^{α}. Assignment overhead is described as

$$W_{\alpha}(y) = \sum_{s \in S} \sum_{c \in N_c} y_{sc} w_{sc}^{\alpha}. \tag{12}$$

Increasing the data-plane nodes managed by the controller increases the number of messages exchanged with other controllers. The overhead produced through the fixed rate messages exchange among controller m and n is indicated by w_{mn}^{con}, while the extra cost related to the controller m's load is denoted by w_{mn}^{add}.

In a multicontroller network, controllers communicate data with each other using a consensus protocol for synchronization. Based on [16], various consensus approaches may be employed by several controllers to attain synchronization

goals. According to the presented structure, the leader-based case is discussed here.

In the leader-based approach, controllers just exchange data with the leader. The leader-based overhead can be described as

$$W_{lb}(x, y) = \sum_{m \in N_c} \sum_{n \in N_c} x_m \left(w_{mn}^{lb\_con} + w_{mn}^{lb\_add} \sum_{s \in S} y_{sm}\right). \tag{13}$$

According to the mentioned analysis, the overall overhead can be obtained as

$$W(x, y) = W_{\alpha}(y) + W_{lb}(x, y). \tag{14}$$

Considering the energy utilized through the abovementioned transition overhead, according to [17], the following EC model can be obtained, which describes the EC of the transmission of k bit/s information:

$$E_t(k, d) = \begin{cases} k(E_{elec} + \varepsilon_{fs}d^2), & d < d_{max}, \\ k(E_{elec} + \varepsilon_{amp}d^4), & d \ge d_{max}, \end{cases} \tag{15}$$

where $E_{elec} = 50$ nJ/bit, $\varepsilon_{fs} = 10$ pJ/bit/m^2, and $\varepsilon_{amp} = 0.0013$ pJ/bit/m^4 describe the energy parameters of the amplifier of power with various distances. The maximum communication radius is obtained as $d_{max} = \sqrt{\varepsilon_{fs}/\varepsilon_{amp}}$. The whole EC for transition and synchronization is given by

$$E(x, y) = W(x, y) + (E_{elec} + \varepsilon_{fs}d^2). \tag{16}$$

It is evident that the outspread deployment of more controllers within the network decreases the reaction delay because nodes can be managed at a lower distance. Nevertheless, the dense deployment of a smaller number of controllers decreases the EC for multicontroller synchronization. The mentioned two assessments are inconsistent and cannot be diminished at a similar time. Thus, an appropriate deployment approach should be obtained to make a balance between delay and EC and obtain the minimum attainable response delay and consumption. The weight parameter $\alpha > 0$ is chosen to balance between the mentioned two assessments and optimizing the controller deployment and assignment through solving the given minimization problem.

$$\min_{x,y} \alpha D(y) + E(x, y)$$
$$\text{subject to} \quad (1\text{-}5). \tag{17}$$

The standard linearization can be utilized to convert the problem (17) to a Mix-Integer Programming (MIP) problem. Now, a mathematical optimizer like CPLEX can be adopted for its solution [18]. The results can be obtained using Algorithm 1. It is performed on the global controller for a constant time range.

5. Numerical Simulations

In the current section, the presented approach's efficiency is evaluated through the mobile ad hoc networking interoperability and cooperation (MANIAC) datasets [19]. The ONOS is assumed to serve as the studied controller because it provides the deployment of a multicontroller. Mininet-WiFi is employed to perform simulations on similar topology in MANIAC to evaluate whether the presented approach can satisfy the desired goals. Mininet-WiFi incorporates novel categories to provide the possible addition of the mentioned wireless mobile devices in a Mininet network framework [20]. Moreover, measurements can be obtained through an actual traffic overhead and delay in a simulated approach [21].

The model parameters are chosen as an out-of-door ambient combining with a path loss and shadow channel for (6), $k = -31.54\,\text{dBm}$, $P_t = 24\,\text{dBm}$, $P_{\min} = -115\text{dBm}$, $\gamma = 3.7$, $\sigma_{\psi_{dB}} = 3.65\text{dBm}$, and $d_0 = 50\text{m}$ [15]. And $B = 5\,\text{MHz}$ and $\sigma^2 = -100\text{dBm}$ for (9). The packet scale η of the requested flow tables is chosen as 1500 bytes. The traffic of controller nodes and synchronization of controller-controller are utilized to generate the overhead. A certain amount is related to the controllers' type and their corresponding distance. The mathematical relations corresponding to the obtained overhead are utilized to obtain the EC. The CPLEX optimizer is adopted to solve the optimization problem.

Various controller deployment approaches are constructed based on the k-median clustering technique for optimizing the delay among data-plane nodes and controllers. The presented approach is compared with the k-median approach. As k-median considers the quantity of controllers as an initial data, a similar quantity of controllers is chosen for the comparison.

As shown in Figure 2, the presented approach can set the quantity of controllers corresponding to the parameter of weight α. The weight parameter α denotes the same as that in (17). Increasing the weights of delay leads to more controllers in the vicinity of nodes. According to Figure 3, there is a trade-off between delay and EC. As shown in Figure 4, lower EC can be obtained through the presented approach than the k-median clustering approach. The MIP in this figure represents our presented approach since it has been converted to a MIP problem.

When the EC weights increase, the presented approach prefers to place the smaller number of controllers to reduce the intercontroller and controller-node communications and EC. When the response delay is more significant, the number of controllers positioned near the nodes to attain a controller with a lesser reaction delay is more. According to simulation outcomes, the presented approach can make a balance between delay and EC and regulate the proper quantity of controllers.

6. Conclusions

In the current work, the controller deployment issue is verified in the SDN-based TANET. A structure of SDN-based mobile ad hoc networks is constructed. Now, an optimal controllers' deployment is developed for delay guarantee and optimal EC. Simulation results indicate significant efficiency in decreasing delay and EC. According to the presented approach, there is a trade-off between two assessments in various preferences. The hybrid control in the mentioned framework and controller deployment can be considered as future works.

Data Availability

Previously reported CRAWDAD dataset was used to support this study and are available at https://crawdad.org/vt/maniac/20110721/2009. These prior studies (and datasets) are cited at relevant places within the text as references [19].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] K. Poularakis, G. Iosifidis, and L. Tassiulas, "SDN-enabled tactical ad hoc networks: extending programmable control to the edge," *IEEE Communications Magazine*, vol. 56, no. 7, pp. 132–138, 2018.

[2] R. Amin, M. Reisslein, and N. Shah, "Hybrid SDN networks: a survey of existing approaches," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3259–3306, 2018.

[3] A. Giorgetti, A. Sgambelluri, R. Casellas, R. Morro, A. Campanella, and P. Castoldi, "Control of open and disaggregated transport networks using the open network operating system (ONOS) [invited]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 12, no. 2, pp. A171–A181, 2020.

[4] H. Pan and S. Wang, "Optimizing the SDN control-plane performance of the Openvswitch software switch," in *2015 IEEE Symposium on Computers and Communication (ISCC)*, pp. 403–408, Larnaca, Cyprus, 2015.

[5] H. C. Yu, G. Quer, and R. R. Rao, "Wireless SDN mobile ad hoc network: from theory to practice," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–7, Paris, 2017.

[6] P. Bellavista, A. Dolci, and C. Giannelli, "MANET-oriented SDN: motivations, challenges, and a solution prototype," in *2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, pp. 14–22, Chania, Greece, 2018.

[7] W. Qi, Q. Song, X. Kong, and L. Guo, "A traffic-differentiated routing algorithm in flying ad hoc sensor networks with SDN cluster controllers," *Journal of the Franklin Institute*, vol. 356, no. 2, pp. 766–790, 2019.

[8] K. Poularakis, Q. Qin, E. M. Nahum, M. Rio, and L. Tassiulas, "Flexible SDN control in tactical ad hoc networks," *Ad Hoc Networks*, vol. 85, pp. 71–80, 2019.

[9] M. J. Abdel-Rahman, E. A. Mazied, K. Teague, A. B. MacKenzie, and S. F. Midkiff, "Robust controller deployment and assignment in software-defined cellular networks," in *The 26th International Conference on Computer Communications and Networks (ICCCN 2017)*, pp. 1–9, Vancouver, BC, 2017.

[10] M. J. Abdel-Rahman, E. A. Mazied, A. MacKenzie, S. Midkiff, M. R. Rizk, and M. El-Nainay, "On stochastic controller placement in software-defined wireless networks," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, San Francisco, CA, 2017.

[11] A. Dvir, Y. Haddad, and A. Zilberman, "Wireless controller deployment problem," in *15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pp. 1–4, Las Vegas, NV, 2018.

[12] M. Johnston and E. Modiano, "Controller placement in wireless networks with delayed CSI," *IEEE/ACM Transactions on Networking*, vol. 25, no. 3, pp. 1775–1788, 2017.

[13] M. Abolhasan, J. Lipman, W. Ni, and B. Hagelstein, "Software-defined wireless networking: centralized, distributed, or hybrid?," *IEEE Network*, vol. 29, no. 4, pp. 32–38, 2015.

[14] X. Wang, M. Huang, and L. Wan, "Joint 2D-DOD and 2D-DOA estimation for coprime EMVS–MIMO radar," *Circuits, Systems, and Signal Process*, vol. 40, no. article 1605, 2021.

[15] A. Goldsmith, *Wireless Communications*, Cambridge University Press, New York, 2012.

[16] Q. Qin, K. Poularakis, G. Iosifidis, S. Kompella, and L. Tassiulas, "SDN controller placement with delay-overhead balancing in wireless edge networks," *IEEE Transactions on Network and Service Management*, vol. 15, no. 4, pp. 1446–1459, 2018.

[17] C. Li, J. Bai, J. Gu, X. Yan, and Y. Luo, "Clustering routing based on mixed integer programming for heterogeneous wireless sensor networks," *Ad Hoc Networks*, vol. 72, pp. 81–90, 2018.

[18] M. Barros and M. Casquilho, "Linear programming with CPLEX: an illustrative application over the Internet CPLEX in Fortran 90," in *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6, Coimbra, Portugal, 2019.

[19] L. A. DaSilva, A. B. MacKenzie, M. S. Thompson, and E. Q. Baumann, "The MANIAC challenge: educational experiences in ad hoc networking," *IEEE Pervasive Computing*, vol. 8, no. 1, pp. 7–11, 2009.

[20] R. R. Fontes, S. Afzal, S. H. B. Brito, M. A. S. Santos, and C. E. Rothenberg, "Mininet-WiFi: emulating software-defined wireless networks," in *2015 11th International Conference on Network and Service Management (CNSM)*, pp. 384–389, Barcelona, Spain, 2015.

[21] X. Wang, L. T. Yang, D. Meng, M. Dong, K. Ota, and H. Wang, "Multi-UAV cooperative localization for marine targets based on weighted subspace fitting in SAGIN environment," *IEEE Internet of Things Journal*, vol. 8, 2021.

WILEY | Hindawi

Research Article

GPS Sparse Multipath Signal Estimation Based on Compressive Sensing

Guodong He [iD],[1,2] **Maozhong Song** [iD],[1] **Shanshan Zhang**,[2] **Huiping Qin**,[3] **and Xiaojuan Xie**[2]

[1]*College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[2]*School of Physics and Electronic Information, Anhui Normal University, Wuhu 241003, China*
[3]*School of Electronics and Information, South China University of Technology, Guangzhou 510641, China*

Correspondence should be addressed to Maozhong Song; smz108@nuaa.edu.cn

A GPS sparse multipath signal estimation method based on compressive sensing is proposed. A new 0 norm approximation function is designed, and the parameter of the approximate function is gradually reduced to realize the approximation of 0 norm. The sparse signal is reconstructed by a modified Newton method. The reconstruction performance of the proposed algorithm is better than several commonly reconstruction algorithms at different sparse numbers and noise intensities. The GPS sparse multipath signal model is established, and the sparse multipath signal is estimated by the proposed reconstruction algorithm in this paper. Compared with several commonly used estimation methods, the estimation error of the proposed method is lower.

1. Introduction

In wireless communication system, due to the influence of scattering or reflection, radio wave may reach the receiver through multiple different paths, so that the received signal at the receiver is the superposition of multiple signals. The multipath interference is inevitable, which is one of the main reasons that affect the performance of GNSS receiver. It will lead to the distortion of the autocorrelation waveform of the pseudocode and affect the detection of the coherent peak, thus reducing the positioning accuracy of the receiver. It is necessary to detect the multipath signals [1, 2].

Candès et al. [3] and Donoho [4] proposed the theory of compressive sensing (CS) in 2006, which can sample the sparse signal at low sampling rate. Currently, the research and application fields of CS include radar signal arrival angle estimation [5, 6], satellite navigation signal processing [7–9], and sparse radar design [10–13]. The CS signal reconstruction needs to recover the original high-dimensional signal from a small amount of measurement data. There are several kinds of reconstruction algorithms. The relaxation algorithm transforms 0 norm into 1 norm. It converts the problem into a convex optimization problem which is easy to be solved [14, 15]. The greedy algorithm has the advantages of simple calculation and easy implementation. However, the reconstruction accuracy is not high and the sparse number is needed [16, 17]. In recent years, some 0 norm approximate solution algorithms have been developed, which transform the 0 norm into mathematical analyzable approximation functions. The sparse signal is reconstructed by a gradient method, Newton method, or other optimization methods [18, 19]. This method can reduce the difficulty of signal reconstruction and obtain better results.

A new 0 norm approximation function is proposed. The sparse reconstruction problem of CS is translated into a Lagrange multiplier problem. It gradually reduces the parameters and reconstructs the original signal by the modified Newton method. A GPS sparse multipath signal model is established, and the estimation of GPS sparse multipath signal is realized by using the approximate function which is proposed in this paper. The rest of this paper is organized as follows. In Section 2, the GPS sparse multipath model is

introduced. Section 3 analyzes the proposed 0 norm approximate function and reconstruction method. Sections 4 and 5 give the reconstruction and estimation results of the new proposed method. Finally, the paper is concluded in Section 6.

2. GPS Sparse Multipath Model

Assuming that the GPS receiver has accurately tracked the carrier frequency of the receiver signal and stripped the carrier, the received multipath signal in one navigation message can be expressed as

$$r(t) = \sum_{k=1}^{L} A_k s(t - \tau_k) e^{j\phi_k} + n(t). \tag{1}$$

In the formula, A_k, τ_k, and ϕ_k are the amplitude, delay time, and phase of the multipath signal, respectively. $s(t)$ is the spread spectrum code, and $n(t)$ is the noise [20]. The corresponding signals of all sampling points can be obtained by cyclic shift, so the upper formula can be represented as a matrix form:

$$\mathbf{r} = \mathbf{Sa} + \mathbf{n}, \tag{2}$$

where $\mathbf{S} = [\mathbf{s}_{\tau_1}, \mathbf{s}_{\tau_2}, \cdots, \mathbf{s}_{\tau_N}]$ and $a_k = A_k e^{j\phi_k}$. When there are multipath signals, the vector values equal to the amplitude value of the multipath signal, and the other values are 0. The vector a can be represented as

$$a[i] = \begin{cases} a_k, & \text{if } \tau_i = \tau_k, \\ 0, & \text{elsewhere.} \end{cases} \tag{3}$$

The length of a GPS spread spectrum chip is 1/1023 ms. Suppose the sampling frequency of the GPS signal is 51.15 MHz. So there are 50 sampling points in a chip. Assuming that the number of multipath in a chip is 3; that is to say, three elements are not 0 and the others are 0. $\mathbf{r}$ is a sparse vector. The GPS multipath signal represented by formula (3) is a sparse model.

There are many GPS multipath estimation methods. In reference [21], a TK operator method is easy to implement but the estimation effect is not ideal in strong noise environment. The method based on maximum likelihood (ML) estimation is proposed in reference [22]. The ML method is improved, and the Generalized Likelihood Ratio (GLRT) is used to improve the performance of the estimation [23]. The MEDLL is used to analyze the signal, and the effect of MEDLL is good, but the method is complex [24]. There are other wavelet-based analysis methods and so on [25, 26]. These methods estimate GPS multipath signals without considering the sparse characteristics of GPS multipath. If sparse constraints are added to the estimation, the performance of multipath estimation can be improved theoretically.

3. New 0 Norm Approximation Function

3.1. Design of Approximate 0 Norm. Donoho et al. proposed a CS-based sparse reconstruction method for magnetic reso-

nance imaging (MRI) images [27]. Instead of 0 norm, 1 norm is used to reconstruct the MRI images. The 1 norm is changed to the following formula:

$$\|\mathbf{x}\|_1 \approx \sum_{i=1}^{N} \left(x_i^2 + p\right)^{1/2}. \tag{4}$$

In the formula, p is a small positive constant. The approximate formula is used to transform the nondifferential 1 norm into a differentiable function. Then, the MRI image is reconstructed by using the optimization algorithm. A CS reconstruction algorithm based on 0 norm approximation function is proposed [18]. The approximation function is a Gaussian function:

$$\|\mathbf{x}\|_0 = \lim_{\sigma \to 0} \sum_{i=0}^{N} \exp\left(-\frac{x_i^2}{2\sigma^2}\right). \tag{5}$$

The function realizes the approximate expression of 0 norm by decreasing the parameter σ gradually. Then, the sparse signal is reconstructed by the gradient method. Both methods convert a nonconvex 0 norm problem to a convex function which has some deviation in principle. Furthermore, the parameter p of formula (4) remains unchanged, and the reconstruction may have some deviation.

Based on the approximation of 0 norm given by Donoho and Gaussian function, a new approximate function is proposed. The coefficient 1/2 is changed to $p/2$ by modifying the approximation of 1 norm given by Donoho. An approximate 0 norm can be obtained when the coefficient p decreases gradually to 0. The designed 0 norm approximation function is

$$\|\mathbf{x}\|_0 = \lim_{p \to 0} \sum_{i=0}^{N} \left(x_i^2 + p^2\right)^{p/2}. \tag{6}$$

The model has only one parameter, and the approximate function is a smooth differentiable function. So the gradient and Hessian matrix of the function can be obtained, which facilitates the solution of the model.

3.2. Sparse Signal Reconstruction Based on Approximate 0 Norm. Based on the proposed approximation function, the CS reconstruction problem can be described as the following mathematical constraint expression:

$$\min_x \lim_{p \to 0} \sum_{i=0}^{N} \left(x_i^2 + p^2\right)^{p/2} \tag{7}$$
$$\text{s.t.} \quad \mathbf{y} = \mathbf{\Phi x}.$$

$\mathbf{\Phi} \in R^{m \times n}$ is a measurement matrix. In reference [18], the gradient method is used to reconstruct the sparse signal. The descending direction is "zigzag," which affects the efficiency of reconstruction. In this paper, the Newton method is used to reconstruct the sparse signal for improving the efficiency [28, 29]. Newton's method requires positive definite Hessian

---

Input: random measurement matrix $\mathbf{\Phi}$, measurement value y;
Output: reconstruction signal xR;
(1) Initialization: Set the initial parameter: p_0, λ, c, α_k, Total iteration times J, inner iteration L_k, reconstruction error threshold ε, the signal initial value $x_0 = 0$, external iteration count k, internal iteration count t;
(2) The external iteration: if $k > J$, turn to step 5. Otherwise, x_k is calculated by Newton method;
(3) The internal iteration: $t = t + 1$, if $t > L_k$, the internal iteration is over and turn to the next step. Otherwise, calculate x_k^t by x_k and p_k^t which decrease gradually. If $E_t^k = \text{norm}(\mathbf{H}_t^k) < \varepsilon$, turn to the next step;
(4) $p_{k+1} = p_k$, $k = k + 1$, and turn to step 2, searching the optimal value based on x_k^t;
(5) All the iteration is over and get the sparse solution $xR = x_k^t$;
Output2: There is no satellite in the line of sight.

ALGORITHM 1: GPS sparse multipath signal estimation based on compressive sensing.

matrix, so the Hessian matrix needs to be modified in iteration. The CS reconstruction problem is transformed into a Lagrange solution model:

$$\min_x f = \frac{1}{2} \|\mathbf{\Phi}\mathbf{x} - \mathbf{y}\|^2 + \lambda \sum_{i=0}^{N} \left(x_i^2 + p^2\right)^{p/2}. \quad (8)$$

y is Lagrange multiplier. The gradient of formula (8) is

$$\mathbf{g} = \mathbf{\Phi}^{\mathrm{T}}(\mathbf{\Phi}\mathbf{x} - \mathbf{y}) + \lambda \begin{bmatrix} p\left(x_1^2 + p^2\right)^{p/2-1} x_1 \\ . \\ . \\ . \\ p\left(x_N^2 + p^2\right)^{p/2-1} x_N \end{bmatrix}. \quad (9)$$

The Hessian matrix of formula (8) is obtained by derivation of expression (9):

$$\mathbf{H} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi} + \lambda * \begin{bmatrix} u_1 & 0 & \cdots & 0 \\ 0 & u_2 & \cdots & 0 \\ . & . & . & . \\ 0 & 0 & \cdots & u_N \end{bmatrix}. \quad (10)$$

$u_i = p(x_i^2 + p^2)^{p/2-2}[(p-1)x_i^2 + p^2]$. When $p = 1$, u_i is a positive value, so the Hessian matrix is a positive definite matrix. That is to say, the function is a convex function. The value $p = 1$ can be used as the starting point of iteration. According to Newton's method, the iterative formula can be expressed as

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, \quad (11)$$

$$\mathbf{d}_k = -\mathbf{H}^{-1}\mathbf{g}_{k-1}. \quad (12)$$

In the formula ((11) and (12)), α_k is the scale factor and d_k is the direction of gradient descent. g_{k-1} is the gradient of the $k - 1$ iteration which can be obtained by formula (9). In the iteration, u_i may be a negative value in formula (10). Formula (12) requires a positive definite Hessian matrix. It needs to modify the value of u_i to guarantee a positive definite

matrix in the iteration. The modification formula is

$$u_i = \begin{cases} u_i, & \text{if} \quad u_i > \delta > 0, \\ \delta, & \text{else.} \end{cases} \quad (13)$$

δ is a very small positive value which ensures the Hessian matrix is a positive definite matrix. The empirical value is 10^{-5}. The initial value of the parameter p in the approximate function is 1. The value is decreased in iteration according to the following formula:

$$p_k = cp_{k-1}, \quad k = 1, \cdots, J. \quad (14)$$

It can obtain better reconstruction effect when c belong to $(0.5, 1)$.

The general idea of sparse reconstruction algorithm in this paper is given as below. The reconstruction has two search processes. For each given value p_k in the external loop, an optimal value is obtained by the modified Newton method. An upgrading value is gotten by decreasing p gradually. Then, p_k in the outer loop is reduced to the value p_{k+1}, and the new optimal value is searched again according to the previous optimal solution. The sparse solution is found iteratively. The implementation steps of this reconstruction Algorithm 1 are as follows:

3.3. Convergence Analysis of Approximate Function.
In this section, the convergence of the algorithm is analyzed. In order to facilitate the analysis, a lemma is first introduced.

Lemma 1. *Suppose that $\mathbf{A} \in R^{m \times n}$ is a Unique Representation Property (URP) matrix. All $m \times m$ submatrixes have inverse matrix [30]. A vector $\mathbf{s} = [s_1, s_2, \cdots, s_n]^T$ belongs to zero space of $\mathbf{A}$. If more than $n - m$ elements converge to 0 in the vector, then the vector converges to 0.*

Proof. Suppose that $\mathbf{A}$ has been normalized by a column vector. For any $\gamma > 0$, I_γ is a subscript set that all the corresponding s_i are larger than γ. $\widehat{\mathbf{A}}$ is a submatrix which is composed of the columns corresponding to the subscript set.

Because $\mathbf{A}$ is a URP matrix, each column satisfies the linear independence. The matrix $\widehat{\mathbf{A}}$ has a left inverse matrix. $M = \max\{\|\mathbf{A}\wedge^{-1}\|\}$; $|I_\gamma|$ is defined as the number of

elements in the subscript set. From the known conditions,

$$\mathbf{As} = \sum_{i=1}^{m} s_i a_i = \sum_{i \in I_\gamma} s_i a_i + \sum_{i \notin I_\gamma} s_i a_i. \tag{15}$$

So

$$\sum_{i \in I_\gamma} s_i a_i = -\sum_{i \notin I_\gamma} s_i a_i. \tag{16}$$

There are 2 norms for the upper formulas:

$$X = [x_1, x_2, \cdots, x_L] \in R^{N \times L} = \sum_{i \notin I_\gamma} \gamma = \left(n - |I_\gamma|\right) \gamma \leq n\gamma. \tag{17}$$

$\widehat{\mathbf{A}} = [a_i], i \in I_\gamma$; rewrite the upper formula to a matrix form:

$$\sum_{i \in I_\gamma} s_i a_i = \mathbf{As}_{i \in I_\gamma} \Rightarrow \left\| \mathbf{s}_{i \in I_\gamma} \right\|_2 = \left\| \mathbf{A} \wedge^{-1} \sum_{i \in I_\gamma} s_i a_i \right\|_2 \leq \left\| \mathbf{A} \wedge^{-1} \right\|_2 \left\| \sum_{i \in I_\gamma} s_i a_i \right\|_2$$
$$\leq \left\| \mathbf{A} \wedge^{-1} \right\|_2 n\gamma. \tag{18}$$

And $\|\mathbf{s}\|_2 \leq \left\| s_{i \in I_\gamma} \right\|_2 + \left\| s_{i \notin I_\gamma} \right\|_2 \leq \left(\left\| \mathbf{A} \wedge^{-1} \right\|_2 + 1 \right) n\gamma$, so

$$\|\mathbf{s}\|_2 \leq \left(\left\| \mathbf{A} \wedge^{-1} \right\|_2 + 1 \right) n\gamma \leq (M+1) n\gamma. \tag{19}$$

Proof of lemma 1 is over.

Corollary 2. *Suppose that* $\mathbf{A} \in R^{m \times n}$ *is a URP matrix and the equation* $\mathbf{y} = \mathbf{As}$ *has a sparse solution. If there is another solution* $\tilde{\mathbf{s}} = (\tilde{s}_1, \tilde{s}_2, \cdots, \tilde{s}_n)^T$ *and* $\|\tilde{\mathbf{s}}\|_0 = k < m/2$, *the vector converges to 0.*

The proof of corollary can be drawn from the lemma. Because $0 \leq \|\tilde{\mathbf{s}} - \mathbf{s}^0\|_0 \leq 2k < m$, *more than* $n - m$ *elements are 0, and* $\tilde{\mathbf{s}} - \mathbf{s}^0$ *belongs to the zero space of* $\mathbf{A}$. *The matrix is a URP matrix, so the difference vector* $\tilde{\mathbf{s}} - \mathbf{s}^0$ *converges to 0. That is to say, the sparse solution is unique.*

For the approximate function proposed in this paper, suppose that ζ *is a very small positive number. If* $\gamma = \pm(1 - \zeta)^{1/p}$, *for a given* $|\gamma| \leq |s_i| \leq 1$, *the following equations hold:*

$$\lim_{p \to 0} f_p(s_i) = \lim_{p \to 0} (s_i^2 + p^2)^{p/2} \geq \left((1 - \zeta)^{2/p} \right)^{p/2} \approx 1. \tag{20}$$

Because the range of p is (0, 1], so $|1 - \zeta|^{1/p} < |1 - \zeta| < 1$. *For a very small coefficient* s_i *in the vector, the value of the proposed function is approximately 1.*

For any given $\gamma = \pm(1 + \zeta)^{1/p}$, *the following equations hold:*

$$\lim_{p \to 0} f_p(s_i) = \lim_{p \to 0} (s_i^2 + p^2)^{p/2} = \lim_{p \to 0} \left((1 + \zeta)^{2/p} + p^2 \right)^{p/2}$$
$$= \lim_{p \to 0} (1 + \zeta) \left(1 + \frac{p^2}{(1 + \zeta)^{2/p}} \right)^{p/2} \approx 1. \tag{21}$$

For all coefficients which are greater than 1, the function values are approximately 1. Therefore, the proposed approximate function in this paper realizes the function of 0 norm.

The above analysis theoretically ensures that the approximate function can find the sparse solution of the signal. The constraint condition $k < m/2$ is a sufficient rather than a necessary condition, and the constraint condition is strong. If this condition is not satisfied in practical application, the algorithm can reconstruct the sparse signal also, but the probability of reconstruction is very low.

4. Simulation Experiment of Reconstruction Algorithm

Three experiments are designed to verify the feasibility and effectiveness of the algorithm. In experiment 1, the measurement matrix is fixed. The reconstruction performance is analyzed at different sparse numbers. Antinoise performance is analyzed in experiment 2. The running time of the algorithm is also an important standard to test the algorithm. Experiment 3 analyzes the running time of algorithm.

4.1. Different Sparse Numbers. According to the proposed approximate function, a CS reconstruction model is defined as

$$\min_s \lim_{p \to 0} \sum_{i=0}^{N} \left(s_i^2 + p^2 \right)^{p/2} \tag{22}$$
$$\text{s.t.} \quad \|\mathbf{y} - \mathbf{\Phi s}\|_2 < \varepsilon.$$

In order to compare the reconstruction effect, the proposed algorithm is compared with OMP, Donoho method (1 norm) [27], and the reference [18] algorithm (SL0). The noise type is zero mean Gaussian white noise, and the noise variance is 0.01 in experiment 1. The measurement matrix is $\Phi \in R^{200 \times 500}$. The length of measurement value and sparse signal is fixed at 200 and 500, respectively. The simulation is run 1000 times at each sparse number, and the total reconstruction times are counted if the signal-to-noise ratio (SNR) of the reconstructed signal is greater than 25 dB. The experimental results are shown in Table 1.

As can be seen from Table 1, each algorithm can reconstruct the signal completely when the sparse number is less than 50. When the sparse number is greater than 130, the reconstruction SNR of all algorithms are less than 25 dB. When the sparse number increases, the successful reconstruction times of the proposed algorithm in this paper are

TABLE 1: Successful times at different sparse numbers.

Sparse number	50	60	70	80	90	100	110	120	130
OMP	1000	998	981	944	729	228	25	0	0
1 norm	1000	1000	995	978	864	403	96	0	0
SL0	1000	1000	1000	996	947	429	185	3	0
This paper	1000	1000	1000	1000	981	506	227	7	0

greater than other algorithms. For example, when the number of nonzero data reaches 120, there are still 7 times reconstructed signals with SNR above 25 dB, but SL0 algorithm only has 3 times and 1 norm and the OMP algorithm can no longer reconstruct the signal above 25 dB. Comparing the data in Table 1, the reconstruction effect algorithm proposed in this paper is better.

4.2. Robust Analysis of the Proposed Algorithm. The robust analysis of the proposed algorithm is analyzed at different noise intensities in experiment 2. The noise variance is [0.01, 0.02, 0.03, 0.04]. The other parameters are consistent with experiment 1. The average value of the reconstructed SNR is obtained from 1000 experiments. The experimental results are shown in Table 2.

From Table 2, it can be seen that the reconstruction SNR is above 25 dB at low sparse number when the noise variance is 0.01 and 0.02. The proposed algorithm is effective. The reconstruction SNR can reach 25.272 dB when the sparse number is 100 and noise variance is 0.01. The proposed algorithm shows strong robustness. However, when the noise variance is 0.04, the reconstruction effect of this algorithm is not ideal.

4.3. Average CPU Time of Algorithm. The reconstruction time is an important index to estimate the algorithm. The experimental parameter is designed as follows: the measurement matrix $\Phi \in R^{m \times n}$, $m/n = 0.4$, and sparse number $m = 4k$. According to the previous experiments, this numerical setting can effectively reconstruct sparse signals in a noiseless environment. The length of the signal is constantly adjusted, and the CPU times for each of the algorithms averaged over 100 trials are tabulated in Table 3.

From Table 3, it can be found that the running time of several algorithms is close when the total length of the signal is less than 2000 sampling points. However, when the signal length reaches 4000, the reconstruction CPU time of OMP is almost twice that of the proposed algorithm. For high-dimensional matrix, the pseudoinverse calculation of the matrix in the OMP algorithm takes too much time. SL0 is solved by the optimal gradient method which appears "zigzag" when finding the optimal value. This reduces the efficiency of finding the optimal solution. In this paper, the modified Newton method is used to improve the reconstruction speed. According to Table 3, the algorithm is faster than the other algorithms, so the algorithm is more suitable for high-dimensional signal processing.

5. GPS Sparse Multipath Signal Estimation

The GPS sparse multipath model is given in formula (3). The sparse multipath signal is estimated by CS theory in this section. First, the GPS signal is measured by measurement matrix $\Phi \in R^{m \times n}$:

$$z = \Phi r = \Phi Sa + \Phi n. \tag{23}$$

Then, sparse multipath signal is estimated by using the 0 norm approximation function which is proposed in this paper:

$$\min_a f(a) = \frac{1}{2} \|\Phi Sa - z\|^2 + \lambda \sum_{i=0}^{N} \left(a_i^2 + p^2\right)^{p/2}. \tag{24}$$

The multipath model discussed in this paper is based on the following assumptions [31, 32]:

(1) The transmission delay of the nondirect signal is greater than that of the direct signal, which is in line with the actual model. The nondirect signal is often formed by refraction or reflection which takes more time

(2) The amplitude of nondirect signal is lower than that of the direct signal. In the shelter or indoor environment, this assumption is generally not true which needs to estimate multipath with an additional method. But this situation is beyond the scope of this paper which is not analyzed here

(3) In this paper, the multipath delay in on chip is only considered. That is to say, the multipath delay is limited to one chip

(4) Multiple sparse values may be estimated due to noise interference. Assuming that the number of sparse multipath is known, the maximum sparse value is selected as the multipath signal. The small sparse values is considered as noise and excluded

5.1. Comparison of Different Sparse Multipath Estimation Methods. The TK operator [21] and ML estimate method [22] are selected to estimate the multipath signal. Assuming that the GPS signal has two nondirect signals. The direct signal delay is 0, and the normalized amplitude is 1. The delay time of the nondirect signal changes randomly in a chip. The ranges of absolute amplitude vary randomly between 0.2 and 0.8. The GPS signal length is 1 ms and sampling rate is 51.15 MHz. The SNR of correlation result is 35 dB. The compression rate is $m/n = 0.4$. The estimation results of different methods are shown in Figure 1. The amplitude of two random nondirect signals is (0.6, 0.4), and the position is (0.2, 0.58) chip.

From Figure 1, three methods can correctly estimate the multipath signal delay with some deviation in magnitude. The TK operator and ML estimate method can correctly find multipath signals. Due to noise interference, the estimation amplitude deviation of multipath signals is large, and the

TABLE 2: Robust analysis of the proposed algorithm.

Noise variance	Sparse number							
	40	50	60	70	80	90	100	110
0.01	33.108	32.689	32.179	31.574	31.129	29.956	25.272	14.594
0.02	26.785	26.777	26.510	26.280	26.139	25.003	19.559	11.744
0.03	20.602	20.693	20.678	20.652	20.411	19.488	15.945	9.897
0.04	16.952	16.920	16.827	16.747	16.570	14.910	11.887	7.680

TABLE 3: Reconstruction CPU time of different algorithms (second).

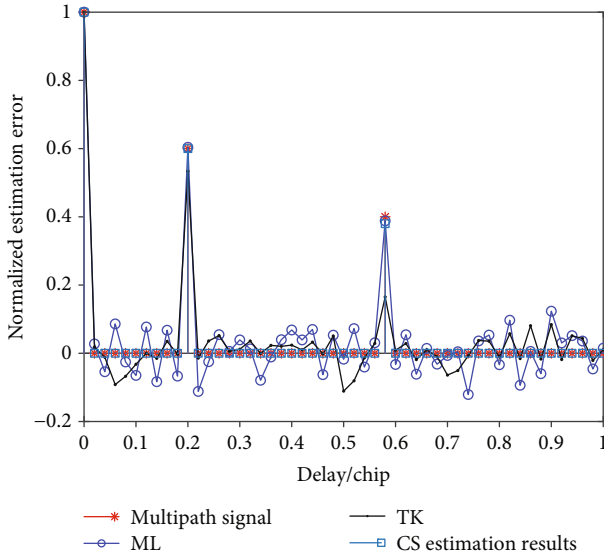| Algorithm	Signal length				
	500	1000	2000	3000	4000
OMP	0.160	0.262	1.816	9.095	21.696
1 norm	0.245	0.258	1.365	6.220	15.461
SL0	0.261	0.322	1.313	5.998	14.669
This paper	0.204	0.270	1.299	5.881	11.577

FIGURE 1: Estimation results of different methods.

estimated values of nonsparse positions fluctuate greatly which result in more fake multipath signals. The CS estimation method proposed in this paper adds sparse prior information, the estimated multipath signal is consistent with the simulation signal position, and the estimated multipath signal amplitudes are close to the real multipath signal.

ML and TK operators are not effective because of noise interference. After the multipath estimation of the signal, some detection algorithm is used to detect it, and the false multipath signal is excluded in practical application. For example, the GLRT criterion is used to analyze the estimated multipath signal in reference [23]. The proposed CS method adds sparse constraints and selects the largest signal as a sparse multipath signal. It can be seen from Figure 1 that the performance is better than that of ML and TK operators. It is unfair to compare the CS estimation method with the common methods, so it only compares the method of this

paper with the method based on the CS theory in the subsequent performance analysis.

5.2. Multipath Signal Estimation Based on CS. It is mainly affected by three factors: noise, sparse number (number of multipath), and measurement matrix dimension m to estimating multipath signal based on CS. This section analyzes the estimated effect from these aspects. First, the sparse estimation results based on CS are analyzed by an experiment. 1 ms GPS signal is selected with sampling rate 40.92 MHz. The sparse number is set to 4, and the delay of nondirect signal varies randomly in a chip. The absolute amplitude varies randomly between 0.2 and 0.8. The delay and amplitude are uniformly distributed. SNR is set to 0 dB. Compression ratio m/n is 0.5. In the simulation, assuming that the number of sparse multipath is known, the estimated results of the proposed method are shown in Figure 2. In a strong noise environment, the amplitude of the multipath signal and the position of the multipath signal may be deviated. Multipath signal estimation error is used to evaluate the estimation results, which is defined as

$$\text{Error} = \frac{1}{K} \sum_{i=1}^{N} |a_i - \widehat{a}_i|. \tag{25}$$

In the formula, N is the total number of sampling point of a chip. K is a sparse number. a_i and $\widehat{a}_i$ represent real and estimated multipath signals, respectively. All multipath signal amplitudes are normalized. It can be seen from Figure 2 that if the position and phase of the estimated signal are consistent with the real multipath signal, the estimation error is small. If the position or phase of estimated signal is different from the real signal, the real value and the estimated value are counted into error. Formula (25) can objectively evaluate the estimated performance.

The multipath estimation effect is analyzed at different multipath numbers and noise intensities in experiment 1. The measurement matrix is fixed and $m/n = 0.3$. The SNR and sparse number vary from -20 dB to 20 dB and 2-8, respectively. The other parameters are set as above. The experiment repeats 100 times at each sparse number. The average estimation error is shown in Figure 3. From the figure, it can be found that the estimation error increases when the sparse number increase. It is consistent with the CS theory that the reconstruction effect decreases with the increase of sparsity at the same measurement matrix. The best estimation effect is obtained when there is only one nondirect
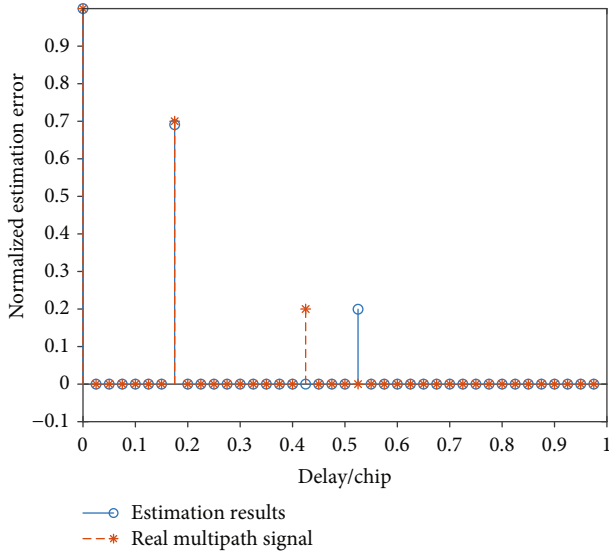
FIGURE 2: Estimation results of the proposed method.
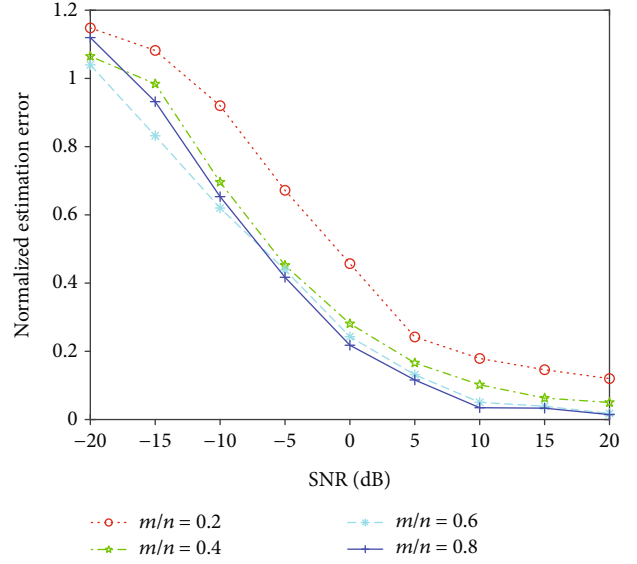
FIGURE 4: Estimation results of different compression ratios.
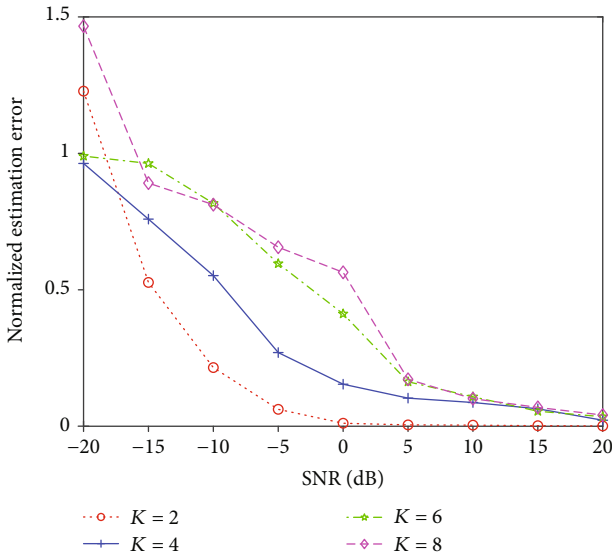
FIGURE 3: Estimation results of different sparse numbers.
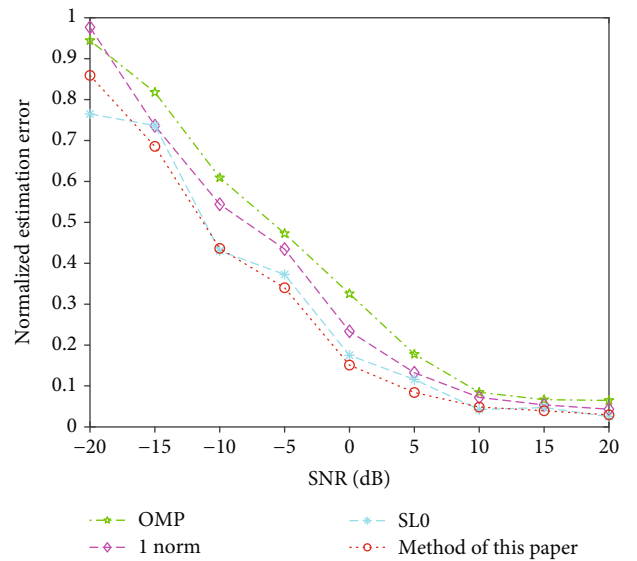
FIGURE 5: Estimation results of different reconstruction algorithms.

multipath signal. The reconstruction becomes unstable when $K = 8$.

When the multipath number is fixed, the influence of measurement matrix is analyzed in experiment 2. The multipath number is set to 6. The column of the matrix remains unchanged, and m/n range from 0.2 to 0.8. The other related parameters are set in the same experiment 1. The experiment repeats 100 times at each m/n. The average estimation error is shown in Figure 4. The estimation error will be relatively reduced when the number of random measurement is increased, which can retain more information of sparse signal. The estimation effect is not good when the SNR is less than 0 dB.

The reconstruction algorithm has some effect on sparse signal reconstruction in CS, which also has some influence on sparse multipath estimation. The OMP, 1 norm, and

SL0 are selected to estimate the multipath signal. The estimation effect of different algorithms is analyzed in experiment 3. The sparse number of signal is set to 5. The m/n of the measurement matrix is 0.5. The GPS signal length is 1 ms with 40.92 MHz sampling rate. The experiment repeats 100 times at each SNR, and the average estimation error is shown in Figure 5. The estimation error of all methods is large at low SNR. When the SNR is higher than 0 dB, the estimation effect is good. As the SNR increases, three multipath estimation errors based on optimization algorithms are slightly lower than OMP methods.

From the three experiments, it can be seen that the sparse multipath signal can be estimated by the CS method effectively. When the SNR is low, the estimation effect is poor. For example, the CS method cannot estimate the multipath

signal at -20 dB strong noise interference. As the SNR increases, the estimation error decreases. When the SNR is higher than 5 dB, the estimated multipath signal is basically consistent with the added multipath signal. In this paper, the length of GPS signal is 1 ms. In practical application, the SNR can be improved by increasing the GPS signal length because the time of navigation message is 20 ms, which can obtain lower estimation error than 1 ms.

6. Conclusions

In this paper, we proposed a new GPS multipath signal estimation method based on CS. A new 0 norm approximate function is proposed. The sparse signal can be reconstructed by using the modified Newton method. A large number of simulation results show that the new method can reconstruct the sparse signal effectively at different sparse numbers and SNR. The reconstruction performance of the proposed algorithm is better than OMP and other optimization algorithms. The GPS signal sparse multipath model is established. The GPS multipath signal can be estimated by the proposed method in this paper. The estimation result is better than the TK and ML method and other CS reconstruction methods. Because there are 20 periods of spread spectrum code in a navigation message, the SNR can be improved by accumulating for a long time, and then, the estimation effect can be improved. Our research team will continue to delve into how to improve the estimation effect of this method.

Data Availability

The simulation data used to support the research of this paper are available from the corresponding author.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] E. D. Kaplan, *Understanding GPS: Principles and Application*, Artech House Mobile Communications, 2005.

[2] L. Cheng, K. Wang, M. Ren, and G. Yan, "Adaptive Filter Approach for GPS Multipath Estimation Under Correntropy Criterion in Dynamic Multipath Environment," *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5798–5810, 2019.

[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[4] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] D. Meng, X. Wang, M. Huang, L. Wan, and B. Zhang, "Robust weighted subspace fitting for DOA estimation via block sparse recovery," *IEEE Communications Letters*, vol. 24, no. 3, pp. 563–567, 2020.

[6] X. Wang, M. Huang, and L. Wan, "Joint 2D-DOD and 2D-DOA estimation for coprime EMVS–MIMO radar," *Circuits Systems and Signal Processing*, pp. 1–17, 2021.

[7] G. He, M. Song, X. He, and Y. Hu, "GPS signal acquisition based on compressive sensing and modified greedy acquisition algorithm," *IEEE Access*, vol. 7, pp. 40445–40453, 2019.

[8] G. He, M. Song, S. Zhang, P. Song, and X. Shu, "Sparse GLONASS signal acquisition based on compressive sensing and multiple measurement vectors," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9654120, 11 pages, 2020.

[9] A. Osman, M. M. E. Moussa, M. Tamazin, M. J. Korenberg, and A. Noureldin, "DOA elevation and azimuth angles estimation of GPS jamming signals using fast orthogonal search," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 5, pp. 3812–3821, 2020.

[10] F. Wen and J. Shi, "Fast direction finding for bistatic EMVS-MIMO radar without pairing," *Signal Processing*, vol. 173, article 107512, 2020.

[11] Q. Cheng, A. A. Ihalage, Y. Liu, Y. Hao, and A. Noureldin, "Compressive Sensing Radar Imaging With Convolutional Neural Networks," *IEEE Access*, vol. 8, no. 5, pp. 212917–212926, 2020.

[12] J. Shi, F. Wen, and T. Liu, "Nested MIMO radar: coarrays, tensor modeling and angle estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 1, pp. 573–585, 2021.

[13] J. He, L. Li, and T. Shu, "Sparse nested arrays with spatially spread square acoustic vector sensors for high accuracy underdetermined direction finding," *IEEE Transactions on Aerospace and Electronic Systems*, p. 1, 2021.

[14] X. Wang, L. Wang, M. Li, and G. Bi, "Nuclear norm minimization framework for DOA estimation in MIMO radar," *Signal Processing*, vol. 153, pp. 147–152, 2017.

[15] E. Michael, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[16] J. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[17] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[18] H. Mohimani, M. Babie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, 2009.

[19] F.-Y. Wu, K. Yang, and R. Duan, "Compressed sensing of underwater acoustic signals via structured approximation l_0-norm," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8504–8513, 2018.

[20] B. Kai, D. M. Akos, and N. Bertelsen, *A Software-Defined GPS and Galileo Receiver*, Birkhäuser Boston, 2007.

[21] R. Hamila, E. S. Lohan, and M. Renfors, "Subchip multipath delay estimation for downlink WCDMA system based on Teager-Kaiser operator," *IEEE Communications Letters*, vol. 7, no. 1, pp. 1–3, 2003.

[22] E. Ertin, U. Mitra, and S. Siwamogsatham, "Maximum-likelihood-based multipath channel estimation for code-division multiple-access systems," *IEEE Transactions on Communications*, vol. 49, no. 2, pp. 290–302, 2001.

[23] N. Sokhandan, J. T. Curran, A. Broumandan, and G. Lachapelle, "An advanced GNSS code multipath detection and estimation algorithm," *GPS Solutions*, vol. 20, no. 4, pp. 627–640, 2016.

[24] B. R. Townsend, P. C. Fenton, K. J. Van Dierendonck, and D. J. R. Van Nee, "Performance evaluation of the multipath estimating delay lock loop," *Navigation*, vol. 42, no. 3, pp. 502–514, 1995.

[25] L. Cheng, K. Wang, M. Ren, and G. Yan, "Adaptive filter approach for GPS multipath estimation under correntropy criterion in dynamic multipath environment," *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5798–5810, 2019.

[26] P. Zhong, X. L. Ding, D. W. Zheng, W. Chen, and D. F. Huang, "Adaptive wavelet transform based on cross-validation method and its application to GPS multipath mitigation," *GPS Solutions*, vol. 12, no. 2, pp. 109–117, 2008.

[27] M. Lustig, D. Donoho, J. M. Pauly, and M. R. I. Sparse, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[28] L. Wan, X. Kong, and F. Xia, "Joint range-Doppler-angle estimation for intelligent tracking of moving aerial targets," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1625–1636, 2018.

[29] A. Antoniou and W. S. Lu, *Practical Optimization: Algorithms and Engineering Applications*, Springer Publishing Company, Incorporated, 2010.

[30] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[31] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performance*, Ganga-Jamuna Press, 2011.

[32] Y. Hu, M. Z. Song, X. Y. Dang, and H. L. Yan, "Interference mitigation for the GPS receiver utilizing the cyclic spectral analysis and RR-MSWF algorithm," *Radioengineering*, vol. 26, no. 3, pp. 798–807, 2017.

WILEY | Hindawi

Research Article

An Efficient Tensor Completion Method Combining Matrix Factorization and Smoothness

Leiming Tang [ID]**, Xunjie Cao, Weiyang Chen** [ID]**, and Changbo Ye** [ID]

Nanjing University of Aeronautics and Astronautics, Nanjing, China

Correspondence should be addressed to Weiyang Chen; weiweigenes@nuaa.edu.cn

In this paper, the low-complexity tensor completion (LTC) scheme is proposed to improve the efficiency of tensor completion. On one hand, the matrix factorization model is established for complexity reduction, which adopts the matrix factorization into the model of low-rank tensor completion. On the other hand, we introduce the smoothness by total variation regularization and framelet regularization to guarantee the completion performance. Accordingly, given the proposed smooth matrix factorization (SMF) model, an alternating direction method of multiple- (ADMM-) based solution is further proposed to realize the efficient and effective tensor completion. Additionally, we employ a novel tensor initialization approach to accelerate convergence speed. Finally, simulation results are presented to confirm the system gain of the proposed LTC scheme in both efficiency and effectiveness.

1. Introduction

As a high-order generalization of vector and matrix, tensor can show complicated structures of high-order data more clearly. With the tensor form, we can understand the internal connection of data from a higher level perspective. Thus, tensor is widely applied in several fields like signal reconstruction [1], signal processing [2–5], image recovery [6], and video inpainting [7].

The purpose of low-rank tensor completion (LRTC) is to estimate missing entries and recover the incomplete tensor data, which takes advantage of the low rank prior to estimate the missing data. Optimization problem LRTC can be ordinarily formulated as follows:

$$\min_{\mathscr{X}} \ \mathrm{rank}\,(\mathscr{X}),$$
$$\mathrm{s.t.} \mathscr{X}_{\Omega} = \mathscr{T}_{\Omega}. \tag{1}$$

Here, $\mathscr{X}$ is the target tensor, and $\mathscr{T}$ is the available data; Ω is observed entries index. The constraint $\mathscr{X}_{\Omega} = \mathscr{T}_{\Omega}$ is to keep the entries of $\mathscr{X}$ in Ω consistent with $\mathscr{T}$.

It is obvious that minimizing the tensor rank is the fundamental problem of LRTC. However, there is no uniform definition of the tensor rank. CP-rank [8] and Tucker-rank [9]

are two commonly used definitions, which lie in the corresponding decompositions, CP decomposition [10], and Tucker decomposition [11]. Moreover, the problem of calculating CP-rank is NP-hard, and it has no relevant relaxation.

Although minimizing Tucker-rank is still NP-hard, it can be relaxed by sum of nuclear norm (SNN) since it is the tightest surrogate of matrix rank. Based on this property of Tucker-rank, Liu et al. [12] developed a theoretical framework for LRTC, and the definition of the nuclear norm for tensors is formed as follows:

$$\|\mathscr{X}\|_* = \sum_{i=1}^{N} \omega_i \left\| \mathbf{X}_{(i)} \right\|_*. \tag{2}$$

Here, $\mathbf{X}_{(i)}$ is the mode-i unfolding matrix of $\mathscr{X}$, $\|\mathscr{X}\|_*$ presents the nuclear norm of $X_{(i)}$, and $\omega_i \geq 0$ ($\sum_{i=1}^{N} \omega_i = 1$, $i = 1, 2, \cdots, N$) denotes the corresponding weight value of $\mathbf{X}_{(i)}$. Then, the LRTC model can be written as

$$\min_{\mathscr{X}} \ \sum_{i=1}^{N} \omega_i \left\| \mathbf{X}_{(i)} \right\|_*,$$
$$\mathrm{s.t.} \mathscr{X}_{\Omega} = \mathscr{T}_{\Omega}. \tag{3}$$

As shown in (3), the LRTC problem is divided into a series of matrix nuclear norm minimization. Through unfolding the target tensor, the matrices are formed to be optimized. In order to resolve (3), SiLRTC, HaLRTC, and FaLRTC were proposed by Liu as the solution for LRTC [12]. Unfortunately, all these algorithms require to compute singular value decomposition (SVD) iteratively, which is complexity exhausted. Considering this difficulty, Liu and Shang [13] applied matrix factorization to the SNN model to reduce the scale of matrices requiring SVD (see details in Section 2).

$$\min_{L_i, R_i, \mathcal{X}} \sum_{i=1}^{N} \omega_i \|R_i\|_*, \quad i = 1, \cdots, N,$$
$$\text{s.t.} X_{(i)} = L_i R_i, \mathcal{X}_\Omega = \mathcal{T}_\Omega, \tag{4}$$

where $L_i \in St(I_i, s_i)$, $R_i \in \mathbb{R}^{s_i \times \prod_{j \neq i} I_j}$, and $St(I_i, s_i)$ denote the Stiefel manifold, and $s_i > r_i$ is a given upper bound of rank. In this paper, we use SNN combined with matrix factorization to utilize the low-rankness feature of tensors and improve the efficiency of LRTC.

Additionally, only considering low-rankness factor is not adequate for LRTC, which results in inevitable performance degradation with the decrement of the sampling rate [14, 15]. Apart from the low rankness of tensor, smoothness is also a significant attribute in real world data. For instance, spectral signals [16] and natural pictures and videos [17] generally have this property. Total variation (TV) [18] is commonly applied as constraints to ensure smoothness. The application of TV in image processing has been proved that it can effectively encourage the smooth prior [4, 19, 20].

For the purpose of improving the quality of tensor completion, Yokota and Hontani [21] proposed LRTV-PDS to minimize SNN and TV simultaneously as:

$$\min_{\mathcal{X}} \rho \|\mathcal{X}\|_* + (1 - \rho) \|\mathcal{X}\|_{\text{TV}},$$
$$\text{s.t.} v_{\min} \leq \mathcal{X} \leq v_{\max}, \|\mathcal{X}_\Omega - \mathcal{T}_\Omega\|_F^2 \leq \delta, \tag{5}$$

where ρ is the trade-off connecting the SNN and TV term and δ is the noise threshold parameter. The first constraint restricts the recovered data in a range, $[v_{\min}, v_{\max}]$. The second constraint means that the output tensor and original tensor are usually the same at observed entries. Considering the effect of noise, some deviation within a certain range is allowed.

As another commonly used definition of tensor rank, CP-rank is also widely utilized in LRTC. On the basis of CP decomposition, Yokota et al. [22] considered TV as the smoothness constraint along with the CP-rank and proposed the smooth PARAFAC tensor completion (SPC) method to achieve better result of tensor completion when the sampling rate is exceedingly low.

However, the TV regularization usually leads to staircase effect [23], which may cause the possible loss of information and geometric features in practice. For this reason, the framelet regularization can be further applied to avoid the possible performance degradation. More precisely, as a generation of the orthogonal basis, framelet effectively relaxes the restriction of the orthogonality and linear independence so that the valuable information and geometric features can be well preserved by the introduced redundancy [24]. In this paper, the framelet is applied to exploit the smoothness and preserve more details.

Considering SNN, matrix factorization, framelet, and TV simultaneously, we propose a novel model for tensor completion, which is named smooth matrix factorization (SMF). The SMF model is formulated as

$$\min_{\mathcal{X}} \sum_{i=1}^{N} \omega_i \|R_{(i)}\|_* + \lambda_1 \|WX_{(3)}^T\|_{1,1} + \lambda_2 \|D_S X_{(3)}\|_{1,1},$$
$$\text{s.t.} \mathcal{X}_\Omega = \mathcal{T}_\Omega, X_{(i)} = L_i R_i, L_i \in St(I_i, s_i), i = 1, \cdots, N. \tag{6}$$

Here, λ_1 and λ_2 are adjustable parameters to balance SNN, framelet, and TV. Besides W indicates framelet transformation, D_S denotes the difference matrix, and $l_{1,1}$-norm is the sum of absolute values of the matrix elements.

In addition, we notice that how to initialize the observed tensor can affect the efficiency and effectiveness of tensor completion. Most of these existing works only focus on optimizing the model and ignore initializing the tensor before iteration algorithm. There are two commonly used initialization methods: one is to set the unknown values as zero; the other is to set them as the average of all observed values. Different from the existing simple methods, our proposed initialization method considers the location of the known data and expands the known data around until the entire tensor is completed. Through the rough estimation of unknown data, we get a more accurate initial tensor, which can accelerate the convergence speed of solution.

In summary, the low-complexity tensor completion (LTC) scheme is proposed to improve the efficiency of LRTC. The proposed LTC mainly consists of two parts—the SMF model and the alternating direction method of multiple- (ADMM-) based solution. Specifically, the matrix factorization is introduced into the SNN to exploit the low rankness of the entire tensor. By factorizing matrices, the computational complexity in calculating SVD can be significantly reduced, which leads to efficiency improvement. Then, to guarantee the effectiveness of tensor completion, the TV is used to exploit the smoothness globally. After that, the framelet is further applied to further ensure the smoothness and preserve the information due to its redundancy. With respect to the proposed SMF model, we also give an effective ADMM-based algorithm to solve it. Additionally, we propose a novel initialization approach to accelerate convergence speed. On the basis of natural data, such as color image and grayscale video, the experimental results show our LTC scheme can achieve a better trade-off between performance and complexity for LRTC.

Compared with the existing works, the contributions of this paper are mainly three folds:

(i) By concurrently utilizing the low-rank and smooth properties, we propose an advanced model for low-

rank tensor completion. Besides, matrix factorization is introduced to this model to save calculation cost

(ii) We propose an effective tensor initialization approach, which can get a better start point for tensor completion to accelerate the convergence speed

(iii) An ADMM-based algorithm is developed to resolve the new SMF model. As shown in the result of numerical experiments, our method can clearly improve the effectiveness of LRTC

The rest of this paper is structured as follows: Firstly, we present some notations that this work needs in Section 2. Then, we propose the LTC scheme which consists of the SMF model and the corresponding ADMM-based solution in Section 3. Afterwards, the performance of LTC is evaluated and compared with other competing methods in Section 4. Finally, the conclusions are given in Section 5.

2. Preliminary

In this section, several basic notations and relevant definitions this work needs are shown [25].

2.1. Tensor Basics. Here, we use different fonts to distinguish data formats. For instance, we write vectors as $\boldsymbol{x}$, matrices as $\boldsymbol{X}$, and tensors as $\mathcal{X}$. The $(i_1, i_2, \cdots, i_N)$ – th component of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is denoted as $x_{i_1, i_2, \cdots, i_N}$.

The inner product of two tensors $\mathcal{A}$ and $\mathcal{B}$ is formulated as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1 i_2 \cdots i_N} a_{i_1 i_2 \cdots i_N} b_{i_1 i_2 \cdots i_N}. \tag{7}$$

As the corresponding norm to the inner product, the definition of Frobenius norm of tensor $\mathcal{A}$ is then given as

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \sqrt{\sum_{i_1 i_2 \cdots i_N} |a_{i_1 i_2 \cdots i_N}|^2}. \tag{8}$$

The fiber of the tensor $\mathcal{X}$ is a vector obtained by fixing each index except one. The mode-n fiber is denoted as $\mathcal{X}(i_1, \cdots, i_{n-1}, :, i_{n+1}, \cdots, i_N)$. The mode-$n$ unfolding of tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is represented as $\boldsymbol{X}_{(n)} \in \mathbb{R}^{I_n \times \prod_{i \neq n} I_i}$. Additionally, we denote the inverse operator of unfolding as "fold." For example, the operation of fold a matrix into a tensor is written as $\mathcal{X} = \text{fold}_n(\boldsymbol{X}_{(n)})$.

2.2. Matrix Factorization. Given a low-rank tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with rank $(r_1, r_2, \cdots, r_N)$, the mode-i unfolding $\boldsymbol{X}_{(i)} \in \mathbb{R}^{I_i \times \prod_{j \neq i} I_j}$ can be factorized into the form $\mathcal{X} = \boldsymbol{L}_i \boldsymbol{R}_i$, $i = 1, \cdots, N$, where $\boldsymbol{L}_i \in St(I_i, s_i)$, $\boldsymbol{R}_{(i)} \in \mathbb{R}^{s_i \times \prod_{j \neq i} I_j}$, and $St(I_i, s_i) = \{\boldsymbol{B} \in \mathbb{R}^{I_i \times s_i} : \boldsymbol{B}_T \boldsymbol{B} = \boldsymbol{I}\}$ represent the Stiefel manifold, the collection of all orthonormal $I_i \times s_i$ matrices, and $s_i > r_i$ is a designated constraint of rank, $i = 1, \cdots, N$. By introducing matrix factorization, we have the following property of tensor nuclear norm:

$$\left\| \boldsymbol{X}_{(i)} \right\|_* = \|\boldsymbol{L}_i \boldsymbol{R}_i\|_* = \|\boldsymbol{R}_i\|_*, \quad i = 1, \cdots, N. \tag{9}$$

Thus, the SNN minimization problem (3) can be rewritten with smaller scale matrices as follows to reduce computational complexity:

$$\min_{\boldsymbol{L}_i, \boldsymbol{R}_i, \mathcal{X}} \sum_{i=1}^{N} \omega_i \|\boldsymbol{R}_i\|_*,$$
$$\text{s.t.} \quad \boldsymbol{X}_{(i)} = \boldsymbol{L}_i \boldsymbol{R}_i, \boldsymbol{L}_i \in St(I_i, s_i), \tag{10}$$
$$\mathcal{X}_\Omega = \mathcal{T}_\Omega, \quad i = 1, \cdots, N.$$

2.3. Framelet. In the discrete setting, the framelet transform is denoted as a linear operator $\boldsymbol{W} \in \mathbb{R}^{k \times mn}$. For instance, $\boldsymbol{Wf}$ means that the framelet transform operator $\boldsymbol{W}$ is applied to the image data, which rearranged as a vector $\boldsymbol{f} \in \mathbb{R}^{mn}$. On the basis of the unitary extension principle, $\boldsymbol{W}$ denotes the framelet transform matrix satisfying $\boldsymbol{W}^T \boldsymbol{W} = \boldsymbol{I}$. In this paper, the piecewise linear B-spline framelets constructed by [26] is applied to exploit the smoothness and preserve details in tensor completion.

3. The Proposed LTC Scheme

In this section, the LTC scheme is proposed, which consists of the SMF model, the tensor initialization method, and the ADMM-based algorithm.

3.1. The Proposed SMF Model. Considering a 3rd-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the proposed SMF model is as follows:

$$\min_{\mathcal{X}} \sum_{i=1}^{3} \omega_i \left\| \boldsymbol{R}_{(i)} \right\|_* + \lambda_1 \left\| \boldsymbol{W} \boldsymbol{X}_{(3)}^T \right\|_{1,1} + \lambda_2 \left\| \boldsymbol{D}_S \boldsymbol{X}_{(3)} \right\|_{1,1},$$
$$\text{s.t.} \mathcal{X}_\Omega = \mathcal{T}_\Omega, \boldsymbol{X}_{(i)} = \boldsymbol{L}_i \boldsymbol{R}_i, \boldsymbol{L}_i \in St(I_i, s_i), i = 1, 2, 3. \tag{11}$$

Here, λ_1 and λ_2 are regularization parameters, $\mathcal{X}$ is the object tensor, and $\mathcal{T}_\Omega$ is the incomplete input tensor, and Ω is the set of indices of available data in $\mathcal{T}$.

Typically, the SMF model contains two main terms—SNN with matrix factorization and smoothness constraints.

The first term, SNN with matrix factorization, can exploit the low-rank property. The goal of introducing SNN is to exploit the globally multidimensional structure, which is the basic of LRTC. Based on the introduced SNN, matrix factorization is further applied to save calculation cost. The purpose of introducing matrix factorization in SNN is to improve the efficiency. Given a low-rank tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with rank $(r_1, r_2, \cdots, r_N)$, the mode-i unfolding $\boldsymbol{X}_{(i)} \in \mathbb{R}^{I_i \times \prod_{j \neq i} I_j}$ can be factorized into the form $\mathcal{X} = \boldsymbol{L}_i \boldsymbol{R}_i$, $i = 1, \cdots, N$. Then, the SNN problem can be rewritten with smaller scale matrices to reduce computational complexity. For example, the computational complexity of SVD of $\boldsymbol{X}$ $(\boldsymbol{X} \in \mathbb{R}^{m \times n}, \text{ rank } (\boldsymbol{X}) = r)$ is $O(m^2 n + mn^2)$. By

introducing matrix factorization, the cost of computing SVD can be reduced to $O(m^2 s + m s^2), r < s \ll n$ [13].

The second term, smoothness constraints, contains the total variation regularization and the framelet regularization, which are used for a better performance of tensor completion.

The total variation regularization $\|\boldsymbol{D}_S \boldsymbol{X}_{(3)}\|_{1,1}$ is used to exploit piecewise smoothness along the mode-3 unfolding of $\mathcal{X}$, where $\boldsymbol{D}_s$ is the difference matrix

$$D_s = \begin{pmatrix} -1 & 1 & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \\ 1 & 0 & \cdots & 0 & -1 \end{pmatrix}. \quad (12)$$

The TV regularization is used to make the third dimension of the recovered tensor smooth to improve the efficiency of tensor completion.

The framelet regularization $\|\boldsymbol{W} \boldsymbol{X}_{(3)}^T\|_{1,1}$ retains details in spatial domain. Here, W denotes the framelet transform matrix satisfying $\boldsymbol{W}^T \boldsymbol{W} = \boldsymbol{I}$. The framelet regularization can further ensure the output tensor smooth and preserve the details due to its redundancy.

In summary, SMF exploits the global low-rank property and the smoothness prior along both the spatial and third mode. As shown in experiments, SNN with matrix factorization term, total variation regularization, and framelet regularization, the three parts of SMF take advantage of these two properties to improve the efficiency and effectiveness of tensor completion.

3.2. The Tensor Initialization Method.
In order to improve the efficiency of tensor completion, we initialize the incomplete tensor to get a better start point for subsequent data processing. Considering the location of known data, the main idea of our tensor initialization method is expanding the known data around until the entire tensor is completed.

Here, we use a specific example to illustrate our tensor initialization method. Considering an incomplete color image which is denoted as $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we make a rough estimate of missing data by assigning the value of each known pixel to the unknown pixels around it. In other words, by averaging the known values adjacent each unknown pixel, the value of the pixel is initialized. For example, if $x_{i,j,k}$ is unknown and two adjacent pixels, $x_{i-1,j,k}$ and $x_{i,j+1,k}$, are observed, we can set $x_{i,j,k}$ as $0.5(x_{i-1,j,k} + x_{i,j+1,k})$. All unknown pixels observed with adjacent pixels can be initialized in this way.

Here, we introduce the operator $\mathcal{H} = \mathrm{P}(\mathcal{X})$, which means

$$\mathcal{H}(i,j,k) = \begin{cases} 1, & \text{if } \mathcal{X}(i,j,k) \neq 0, \\ 0, & \text{if } \mathcal{X}(i,j,k) = 0, \end{cases} \quad (13)$$

which is used to project the original tensor to the binarized tensor.

In this step, we have completed an expansion of the known pixels, and the pixels initialized this time will be considered as known in next expansion. Then, we repeat this expansion process until the entire tensor is filled. The proposed tensor initialization method is detailed in Algorithm 1.

3.3. The Proposed ADMM-Based Algorithm.
Here, we design an effective ADMM-based algorithm to resolve the convex problem. In particular, we introduce two additional variables M, N, and the model (11) can be equivalently transformed to the following formulation:

$$\min_{L_i, R_i, \mathcal{X}} \quad \sum_{i=1}^{3} \omega_i \|\boldsymbol{R}_i\|_* + \lambda_1 \|\boldsymbol{M}\|_{1,1} + \lambda_2 \|\boldsymbol{N}\|_{1,1},$$
$$\text{s.t.} \quad \boldsymbol{X}_{(i)} = \boldsymbol{L}_i \boldsymbol{R}_i, \boldsymbol{L}_i \in St(I_i, s_i), i = 1, 2, 3, \quad (14)$$
$$\boldsymbol{M} = \boldsymbol{W} \boldsymbol{X}_{(3)}^T, \boldsymbol{N} = \boldsymbol{D}_S \boldsymbol{X}_{(3)}, \mathcal{X}_\Omega = \mathcal{T}_\Omega,$$

where $\boldsymbol{M} \in \mathbb{R}^{n_1 n_2 \times n_3}$ and $\boldsymbol{N} \in \mathbb{R}^{n_3 \times n_1 n_2}$. By introducing matrices $\boldsymbol{M}$ and $\boldsymbol{N}$, we separate blocks of variables and the augmented Lagrangian function of (14) becomes:

$$L(\boldsymbol{L}_i, \boldsymbol{R}_i, \mathcal{X}, \boldsymbol{M}, \boldsymbol{N}) = \sum_{i=1}^{3} \left(\omega_i \|\boldsymbol{R}_i\|_* + \frac{\beta_1}{2} \left\| \boldsymbol{X}_{(i)} - \boldsymbol{L}_i \boldsymbol{R}_i \right\|_F^2 \right)$$
$$+ \lambda_1 \|\boldsymbol{M}\|_{1,1} + \left\langle \boldsymbol{W} \boldsymbol{X}_{(3)}^T - \boldsymbol{M}, \Psi \right\rangle$$
$$+ \frac{\beta_2}{2} \left\| \boldsymbol{W} \boldsymbol{X}_{(3)}^T - \boldsymbol{M} \right\|_F^2 + \lambda_2 \|\boldsymbol{N}\|_{1,1}$$
$$+ \left\langle \boldsymbol{D}_S \boldsymbol{X}_{(3)} - \boldsymbol{N}, \Theta \right\rangle + \frac{\beta_3}{2} \left\| \boldsymbol{D}_S \boldsymbol{X}_{(3)} - \boldsymbol{N} \right\|_F^2,$$
$$\text{s.t.} \boldsymbol{L}_i \in St(I_i, s_i), i = 1, 2, 3, \mathcal{X}_\Omega = \mathcal{T}_\Omega, \quad (15)$$

where Ψ and Θ are the Lagrange multipliers and β_1, β_2, and β_3 are the penalty parameters. Based on ADMM, we can divide the problem (15) into subproblems which are easier to deal with in smaller sizes.

For the purpose of facilitating the analysis of time complexity in the proposed ADMM-based algorithm, we assume that $s = \max(s_1, s_2, s_3)$, $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and $n_1 \geq n_2 \geq n_3$.

The first subproblem optimizes the variable $\boldsymbol{L}_i$, which is presented as:

$$\boldsymbol{L}_i^{k+1} = \arg \min_{L_i} \left\| \boldsymbol{X}_{(i)}^k - \boldsymbol{L}_i \boldsymbol{R}_i^k \right\|_F^2,$$
$$\text{s.t.} \boldsymbol{L}_i \in St(I_i, s_i). \quad (16)$$

Through solving this subproblem with the orthogonality constraint, the optimal $\boldsymbol{L}_i$ is calculated. Following [27, 28], the optimal solution is given by:

$$L_i^{k+1} = QR\left(X_{(i)}^k \left(R_i^k \right)^T \right). \quad (17)$$

The cost of computing L is $O(s^2 n_1)$.

Input: The observed tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.
Output: The initialized tensor $\mathcal{X}$.
1: $\mathcal{X} = \mathcal{T}, \mathcal{X}_{next} = \mathcal{X}$
2: **while** 0 exists in $\mathcal{X}$ **do**
3:    $\mathcal{H} = \mathrm{P}(\mathcal{X})$
4:    **for** $k = 1$ to n_3 **do**
5:      **for** $i = 1$ to n_1 **do**
6:        **for** $j = 1$ to n_2 **do**
7:          **if** $\mathcal{X}(i, j, k) = 0$ **then**
8:            $\alpha_{sum} = \mathcal{H}(i-1, j, k) + \mathcal{H}(i, j+1, k) + \mathcal{H}(i+1, j, k) + \mathcal{H}(i, j-1, k)$
9:            **if** $\alpha_{sum} \neq 0$ **then**
              $\mathcal{X}_{next}(i, j, k) = (h_{i-1,j,k}/\alpha_{sum})\mathcal{X}(i-1, j, k) +$
10:            $(h_{i,j+1,k}/\alpha_{sum})\mathcal{X}(i, j+1, k) + (h_{i+1,j,k}/\alpha_{sum})\mathcal{X}(i+1, j, k) +$
             $(h_{i,j-1,k}/\alpha_{sum})\mathcal{X}(i, j-1, k)$
11:         **end if**
12:        **end if**
13:       **end for**
14:      **end for**
15:    **end for**
16:    $\mathcal{X} = \mathcal{X}_{next}$
17:**end while**
18:**return** $\mathcal{X}$

ALGORITHM 1: The Proposed Tensor Initialization Method.

The second subproblem optimizing the variable $\boldsymbol{R}_i$ can be calculated as follows:

$$\boldsymbol{R}_i^{k+1} = \arg \min_{\boldsymbol{R}_i} \omega_i \|\boldsymbol{R}_i\|_* + \frac{\beta_1}{2} \left\| \boldsymbol{X}_{(i)}^k - \boldsymbol{L}_i^{k+1} \boldsymbol{R}_i \right\|_F^2. \tag{18}$$

Then, $\boldsymbol{R}_i^{k+1}$ is the optimal solution to the subproblem (18) if and only if

$$0 \in \omega_i \partial \left\| \boldsymbol{R}_i^{k+1} \right\|_* + \beta_1 \left(\boldsymbol{L}_i^{k+1} \right)^T \left(\boldsymbol{L}_i^{k+1} \boldsymbol{R}_i^{k+1} - \boldsymbol{X}_{(i)}^k \right), \tag{19}$$

where $\partial \|\cdot\|_*$ is the subdifferential of $\|\cdot\|_*$. Considering the property of $\boldsymbol{L}_i^{k+1}$, $\left(\boldsymbol{L}_i^{k+1} \right)^T \boldsymbol{L}_i^{k+1} = \boldsymbol{I}_s$, the formula can be transformed as

$$0 \in \omega_i \partial \left\| \boldsymbol{R}_i^{k+1} \right\|_* + \beta_1 \left(\boldsymbol{R}_i^{k+1} - \left(\boldsymbol{L}_i^{k+1} \right)^T \boldsymbol{X}_{(i)}^k \right). \tag{20}$$

Obviously, the formula (20) also needs to be satisfied by the optimal solution $\boldsymbol{R}_i$ for the following convex problem:

$$\arg \min_{\boldsymbol{R}_i} \omega_i \|\boldsymbol{R}_i\|_* + \frac{\beta_1}{2} \left\| \boldsymbol{R}_i - \left(\boldsymbol{L}_i^{k+1} \right)^T \boldsymbol{X}_{(i)}^k \right\|_F^2. \tag{21}$$

Therefore, $\boldsymbol{R}_i^{k+1}$ is also the solution to the subproblem (21), which has an explicit solution.

$$\boldsymbol{R}_i^{k+1} = \mathrm{SVT}_{\omega_i/\beta_1} \left(\left(\boldsymbol{L}_i^{k+1} \right)^T \boldsymbol{X}_{(i)}^k \right). \tag{22}$$

Here, $\mathrm{SVT}_\tau(\cdot)$ is a singular value thresholding operator defined by $\mathrm{SVT}_\tau(\boldsymbol{X}) = \boldsymbol{U} \operatorname{diag}[\max(\sigma - \tau, 0)] \boldsymbol{V}^T$, the SVD of $\boldsymbol{X}$ is given by $\boldsymbol{Q} = \boldsymbol{U} \operatorname{diag}(\{\sigma_i\}_{1 \le i \le r}) \boldsymbol{V}^T$. The complexity of computing $\boldsymbol{R}$ is $O(sn_1^2 n_2^2)$.

The third subproblem about $\boldsymbol{M}$ related to framelet regularization can be written as:

$$\begin{aligned}
\boldsymbol{M}^{k+1} &= \arg \min_{\boldsymbol{M}} \lambda_1 \|\boldsymbol{M}\|_{1,1} + \left\langle \boldsymbol{W} \left(\boldsymbol{X}_{(3)}^k \right)^T - \boldsymbol{M}, \Psi \right\rangle \\
&\quad + \frac{\beta_2}{2} \left\| \boldsymbol{W} \left(\boldsymbol{X}_{(3)}^k \right)^T - \boldsymbol{M} \right\|_F^2 \\
&= \arg \min_{\boldsymbol{M}} \lambda_1 \|\boldsymbol{M}\|_{1,1} \\
&\quad + \frac{\beta_2}{2} \left\| \boldsymbol{W} \left(\boldsymbol{X}_{(3)}^k \right)^T - \boldsymbol{M} + \frac{\Psi}{\beta_2} \right\|_F^2.
\end{aligned} \tag{23}$$

This problem has an explicit optimal solution

$$\boldsymbol{M}^{k+1} = \mathrm{S}_{\lambda_1/\beta_2} \left(\boldsymbol{W} \left(\boldsymbol{X}_{(3)}^k \right)^T + \frac{\Psi}{\beta_2} \right), \tag{24}$$

where $\mathrm{S}_\mu(\cdot)$ is a soft-thresholding operator written as:

$$S_\mu(x) = \begin{cases} 0, & \text{if } |x| \le \mu, \\ \operatorname{sign}(x)(|x| - \mu), & \text{if } |x| > \mu. \end{cases} \tag{25}$$

The cost of calculating $\boldsymbol{W} \boldsymbol{X}_{(3)}^T$ is $O(la^2 n_1 n_2 n_3)$. Here, l denotes the framelet level and a denotes filter number, then the cost of calculating $\boldsymbol{M}$ is $O(lm^2 n_1 n_2 n_3)$.

Input: The initialized tensor $\mathscr{T}$, index set Ω, parameters $\lambda_1, \lambda_2, \beta_1, \beta_2,$ and β_3.
Output: The completed tensor $\mathscr{X}$.
$$\mathscr{X}^0 = \mathscr{T}, L_i^0 = \text{eye}(I_i, s_i), R_i^0 = \text{rand}(s_i, \Pi_{i \neq j} I_j),$$
1: **initialize:** $i = 1, 2, 3, M^0, N^0, \psi^0$ and
                  Θ^0 initialized to 0, k = 0, k$_{\max}$ = 300.
2: **while** not converged and $k < k_{\max}$ **do**
3:      **for** $i = 1$ to 3 **do**
4:          L_i^{k+1} via (17)
5:          R_i^{k+1} via (22)
6:      end for
7:      M^{k+1} via (24)
8:      N^{k+1} via (27)
9:      $\mathscr{X}^{k+1}$ via (29)
10:      ψ^{k+1}, Θ^{k+1} via (30)
11: **end while**
12: **return** $\mathscr{X}^{k+1}$

ALGORITHM 2: The Proposed ADMM-based Algorithm for Solving the SMF Model in (11).

The fourth subproblem about N concerning total variation regularization can be calculated as follows:

$$N^{k+1} = \arg\min_N \lambda_2 \|N\|_{1,1} + \left\langle D_S X_{(3)}^k - N, \Theta \right\rangle + \frac{\beta_3}{2} \left\| D_S X_{(3)}^k - N \right\|_F^2$$
$$= \arg\min_N \lambda_2 \|N\|_{1,1} + \frac{\beta_3}{2} \left\| D_S X_{(3)}^k - N + \frac{\Theta}{\beta_3} \right\|_F^2,$$

$$(26)$$

which has an explicit solution

$$N^{k+1} = S_{\lambda_2/\beta_3} \left(D_S X_{(3)}^k + \frac{\Theta}{\beta_3} \right). \tag{27}$$

The cost of computing N is $O(n_1 n_2 n_3^2)$.
The final subproblem optimizing $\mathscr{X}$ can be formulated as:

$$\mathscr{X}_{\Omega^c}^{k+1} = \arg\min_{\mathscr{X}} \sum_{i=1}^3 \left(\frac{\beta_1}{2} \left\| X_{(i)} - L_i R_i \right\|_F^2 \right)$$
$$+ \frac{\beta_2}{2} \left\| W X_{(3)}^{kT} - M + \frac{\Psi}{\beta_2} \right\|_F^2 \tag{28}$$
$$+ \frac{\beta_3}{2} \left\| D_S X_{(3)}^k - N + \frac{\Theta}{\beta_3} \right\|_F^2.$$

We can update $\mathscr{X}$ as follows:

$$\mathscr{X}_{\Omega^c}^{k+1} = \left(\sum_i \text{fold}_i \left(\text{inv}\left(\beta_1 I + \beta_2 I + \beta_3 D_s^T D_s \right) * \left(\beta_1 L_i^{k+1} R_i^{k+1} \right) \right. \right.$$
$$\left. \left. + \beta_2 \left[W^T \left(M - \frac{\Psi}{\beta_2} \right) \right]^T + \frac{\beta_3 D_s^T (N - (\Theta/\beta_3))}{N} \right) \right)_{\Omega^c},$$
$$\mathscr{X}_\Omega^{k+1} = \mathscr{T}_\Omega. \tag{29}$$
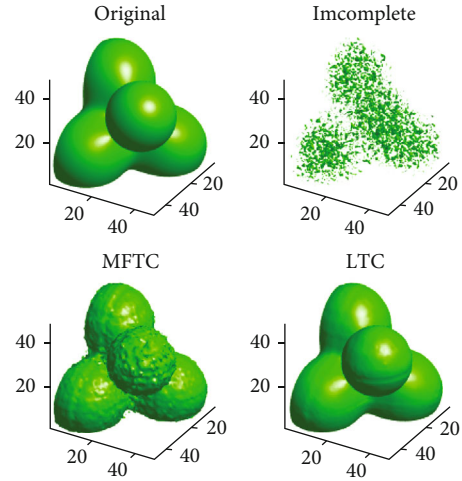
FIGURE 1: Isosurface of synthetic data and completed results of MFTC and LTC.

The cost of calculating $\mathscr{X}$ is $O(sn_1 n_2 n_3 + lm^2 n_1 n_2 n_3 + n_1 n_2 n_3^2 + n_1^3)$.

According to ADMM, the multipliers Ψ and Θ are updated as

$$\Psi^{k+1} = \Psi^k + \beta_2 \left(W \left(X_{(3)}^{k+1} \right)^T - M^{k+1} \right),$$
$$\Theta^{k+1} = \Theta^k + \beta_3 \left(D_S X_{(3)}^{k+1} - N^{k+1} \right). \tag{30}$$

To summarize, the proposed iterative ADMM-based algorithm for solving the SMF model in (11) is outlined in Algorithm 2. The computing complexity of $L, R, M, N,$ and $\mathscr{X}$ at each iteration is $O(sn_1^2 n_2^2 + lm^2 n_1 n_2 n_3 + n_1^3)$.

TABLE 1: The PSNR and SSIM values of five LRTC algorithms on image data.

| Image | MR | PSNR | | | | | SSIM | | | | |
		LTC	MFTC	HaLRTC	PDS	SPC	LTC	MFTC	HaLRTC	PDS	SPC
Lena	70%	28.36	22.46	24.96	28.11	28.83	0.8784	0.6738	0.7523	0.8656	0.8880
	80%	26.82	20.17	21.51	26.06	27.46	0.8274	0.5603	0.5834	0.8095	0.8473
	90%	24.21	16.90	17.54	23.11	25.44	0.7329	0.4043	0.3875	0.7083	0.7744
Giant	70%	24.16	20.59	20.36	23.09	25.26	0.8145	0.6949	0.7123	0.7762	0.8476
	80%	22.59	18.48	17.30	21.31	23.55	0.7359	0.5612	0.5780	0.6810	0.7781
	90%	20.41	15.26	13.91	19.22	21.52	0.5934	0.3578	0.3722	0.5340	0.6580
Pepper	70%	29.26	24.73	24.47	30.87	30.38	0.9426	0.8633	0.8440	0.9682	0.9521
	80%	27.64	22.05	19.80	28.74	29.08	0.9109	0.7709	0.6721	0.9445	0.9338
	90%	25.09	18.10	15.85	25.32	27.00	0.8420	0.5807	0.4811	0.8765	0.8944
Airplane	70%	27.55	25.03	26.16	27.53	28.10	0.9276	0.8630	0.9010	0.9476	0.9278
	80%	25.99	22.71	22.61	25.70	27.10	0.8857	0.7787	0.7938	0.9064	0.9042
	90%	23.59	19.51	19.11	23.07	25.49	0.8059	0.6256	0.6261	0.8069	0.8584

Observed   MFTC   HaLRTC   PDS   SPC   LTC
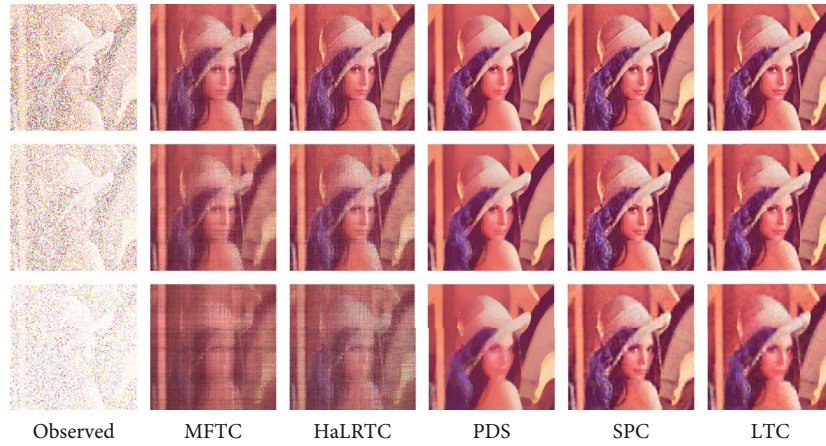
FIGURE 2: The recovered color image by five tensor completion methods with MR = 70%, 80%, 90%.

Observed   MFTC   HaLRTC   PDS   SPC   LTC
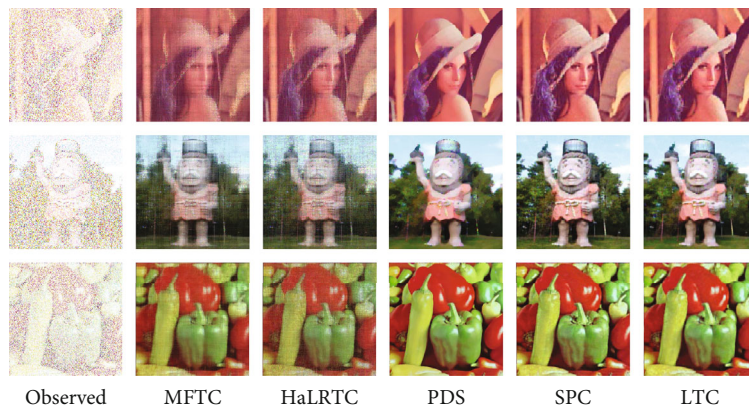
FIGURE 3: The different recovered color images by five tensor completion methods with MR = 80%.

4. Numerical Experiments

In this part, several experiments are conducted on one synthetic and some visual data to demonstrate the performance of our proposed LTC scheme. We also compare it with another four state-of-the-art algorithms: MFTC [13], HaLRTC [12], LRTV-PDS [21], and SPC [22]. Here, we notice that MFTC and HaLRTC only use Tucker-rank to

(a)

(b)

(c)

FIGURE 4: Comparison of performances by five LRTC methods on the color image. (a) Comparison of PSNR, (b) comparison of SSIM, and (c) comparison of running time.

exploit the low-rank prior, while LRTV-PDS and SPC use Tucker-rank and CP-rank, respectively. Besides, both of them are expected to achieve better results on visual data for considering TV term as the smoothness constraint. To measure the estimated accuracy of completed tensor data, peak signal-to-noi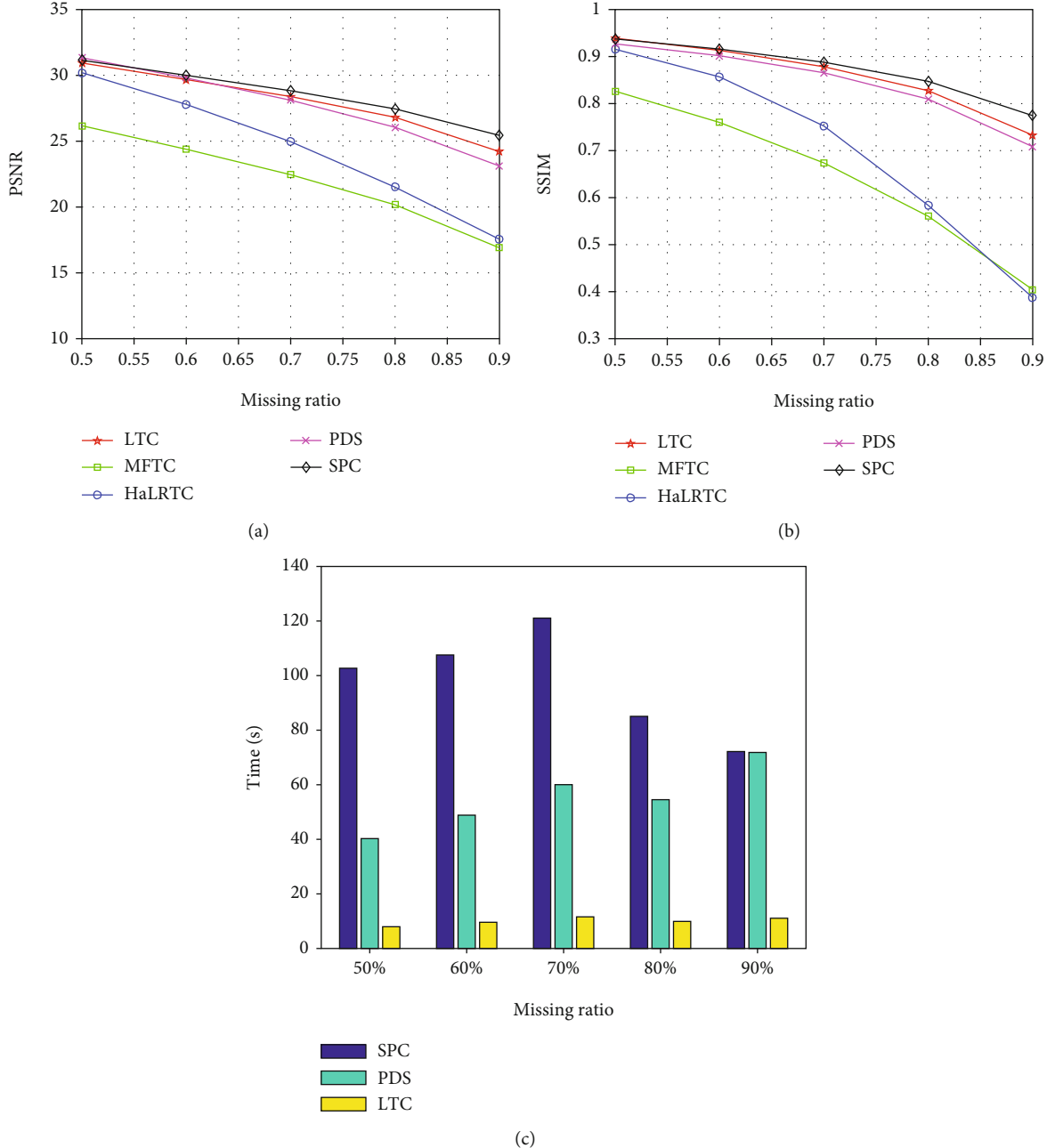se ratio (PSNR) and structural similarity index (SSIM) are employed as quality metrics. The PSNR is defined as $10 \log_{10}(\text{MAX}^2/\text{MSE})$. Here, MAX is the max of whole data, and MSE is formulated as $\|\widehat{\mathcal{X}} - \mathcal{X}_0\|_F^2/N$. The SSIM estimates the resemblance of images in luminance, contrast, and structure. The relative change is defined as RelCha = $\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_F/\|\mathcal{X}\|_F$, which is set as stopping cri-

terion of all five algorithms. In the following experiments, if the relative change is smaller than the tolerance 10^{-5} during the iteration, we terminate the iterative process and output $\mathcal{X}^k$ as the recovered tensor.

4.1. Synthetic Data Completion.
Here, we apply our LTC scheme and MFTC on synthetic data.

Figure 1 displays a visualization of a synthetic tensor and the corresponding results of tensor completion achieved by MFTC and LTC. We use four Gaussian functions to get the synthetic 3rd-order tensor. By randomly removing 80% of the voxels, we get the incomplete tensor. Compared to the
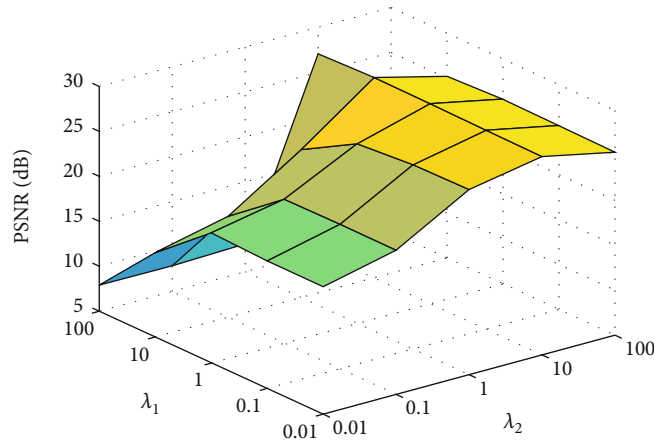
FIGURE 5: The PSNR values with different values of parameters λ_1 and λ_2.

results of MFTC and LTC, it is obvious that only low-rank assumption is not enough for highly missing data, and smoothness constraints can effectively improve the performance of tensor completion.

4.2. Color Images Completion. In this part, we compare the performance of MFTC, HaLRTC, PDS, SPC, and the proposed LTC on four color images "Lena," "Giant," "Pepper," and "Airplane." The incomplete images are generated by randomly deleting elements. In color image data completion, the missing ratios (MRs) are set as 70%, 80%, and 90%. The quality metrics of recovered images completed by five methods are listed in Table 1. From Table 1, it is obvious that LTC achieves high quality metrics values and performs better than other LRTC algorithms except SPC.

Figure 2 presents the results of color image "Lena" recovered by MFTC, HaLRTC, PDS, SPC, and LTC with different missing ratios. The MRs are set as 70%, 80%, and 90%. From Figure 2, it can be observed that the quality of the completed images by MFTC and HaLRTC becomes worse as the missing ratio increases, while the results obtained by PDS, SPC, and LTC still contain most information of the original image.

Figure 3 presents the recovered color images "Lena," "Giant," "Pepper," and "Airplane" completed by MFTC, HaLRTC, PDS, SPC, and LTC with MR = 80%. Obviously, we can see that methods with smoothness constraints (PDS, SPC, and LTC) perform better than those only considering low-rank prior (MFTC, HALRTC) for different color images.

Figure 4 shows performance comparison of color image "Lena" by MFTC, HaLRTC, PDS, SPC, and LTC in PSNR, SSIM, and running time. The MRs are set as 50%, 60%, 70%, 80%, and 90%, respectively. We can see that PDS, SPC, and LTC are superior to MFTC and HaLRTC in terms of PSNR and SSIM. With MR increasing, the advantage of methods with smoothness constraints becomes more prominent. While ensuring the performance, LTC achieves about 80% running time reduction compared to PDS and SPC, which implies that introducing matrix factorization indeed improves the efficiency of tensor completion.

In order to analyze the effect of parameters λ_1 and λ_2, we evaluate the performance of the recovered color image
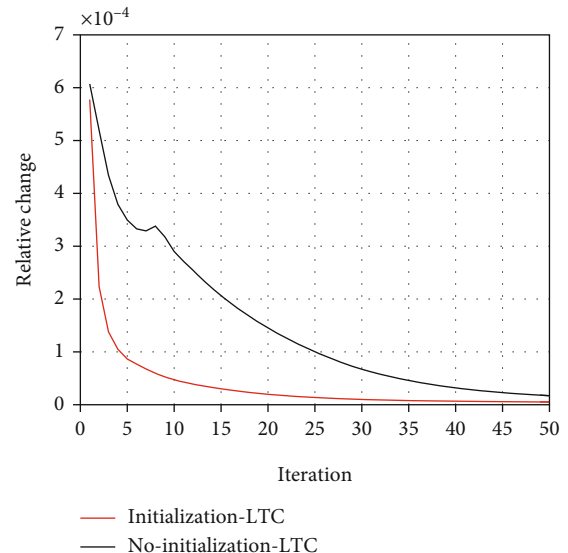
FIGURE 6: The convergence behavior of LTC with or without tensor initialization to show the effect of smoothness constraints in tensor completion.

"Lena" by LTC with MR = 70%. Figure 5 shows the change of the PSNR values for different values of λ_1 and λ_2. It is observed that both λ_1 and λ_2 evidently effect the tensor completed results of our proposed LTC scheme. Additionally, the LTC achieves higher PSNR values when $\lambda_1 = 10$, $\lambda_2 = 10$.

To compare the convergence behavior of LTC with or without tensor initialization, we display the relative change (RelCha) values of the recovered color image "Lena" with MR = 90% in Figure 6. It is observed that the RelCha value of LTC with tensor initialization decreases faster than LTC without tensor initialization, which proves the effectiveness of our proposed tensor initialization method to accelerate the convergence speed.

4.3. Video Completion. Here, two videos: "suzie" and "hall" are tested, and only the first 30 frames of each video are used

| Observed | MFTC | HaLRTC | PDS | SPC | LTC |
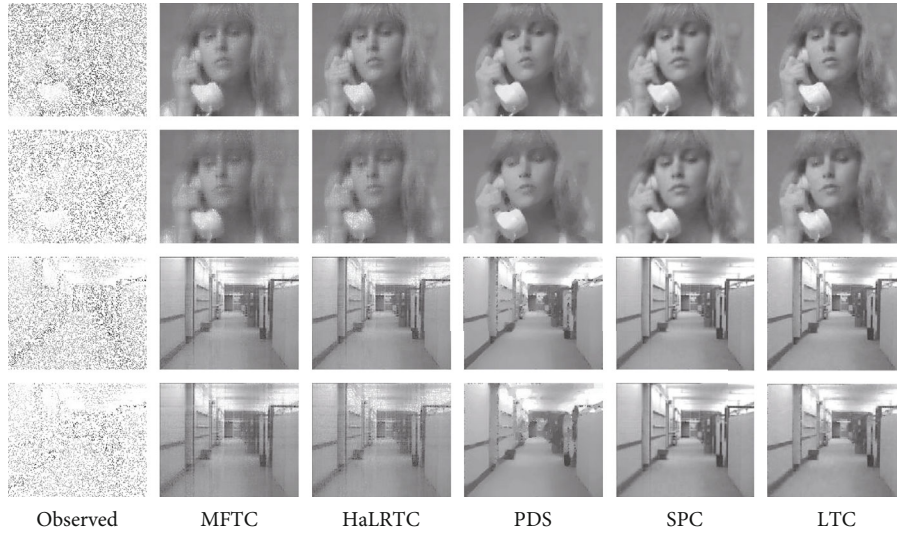
FIGURE 7: The single frame of recovered videos by five tensor completion methods with MR = 70%, 80%.
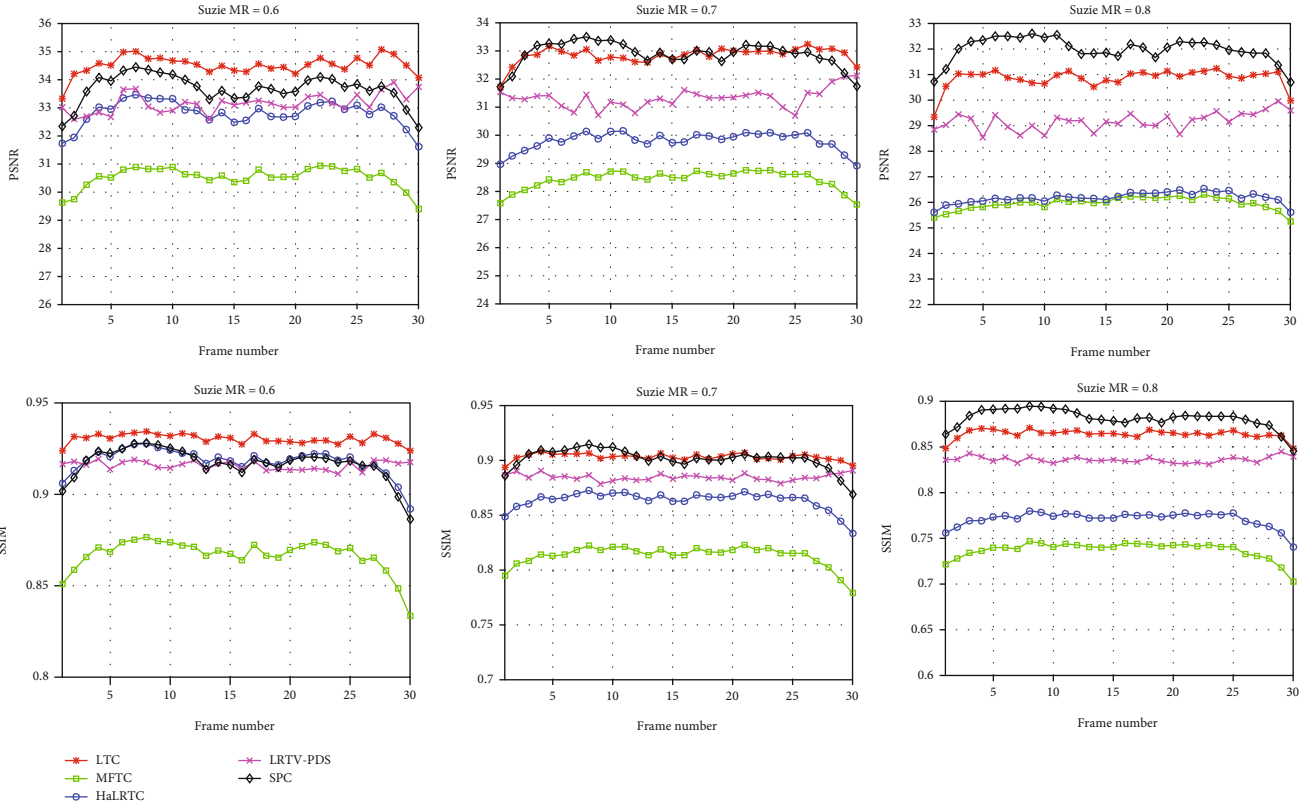
FIGURE 8: The PSNR and SSIM values by five tensor completion methods with MR = 60%, 70%, 80%.

for tensor completion. The size of videos for LRTC is 144 × 176 × 30.

Figure 7 shows the 10th frame in two videos recovered by five tensor completion methods with MRs = 70% and 80%, which presents the visual results of recovered videos. Obviously, the videos recovered by PDS, SPC, and LTC are visually superior to MFTC and HaLRTC. Definitely, SPC and LTC has a better performance to complete the video for pre-

serving details well, while some parts of recovered video by PDS are still blurry.

Figure 8 presents the PSNR and SSIM values of each frame of completed video "suzie" by five algorithms with MRs = 60%, 70% and 80%, which shows the performance comparison more straightforward. From curves in Figure 8, we can conclude that SPC and LTC are better than other algorithms in both stability and effectiveness.

(a)

(b)

(c)
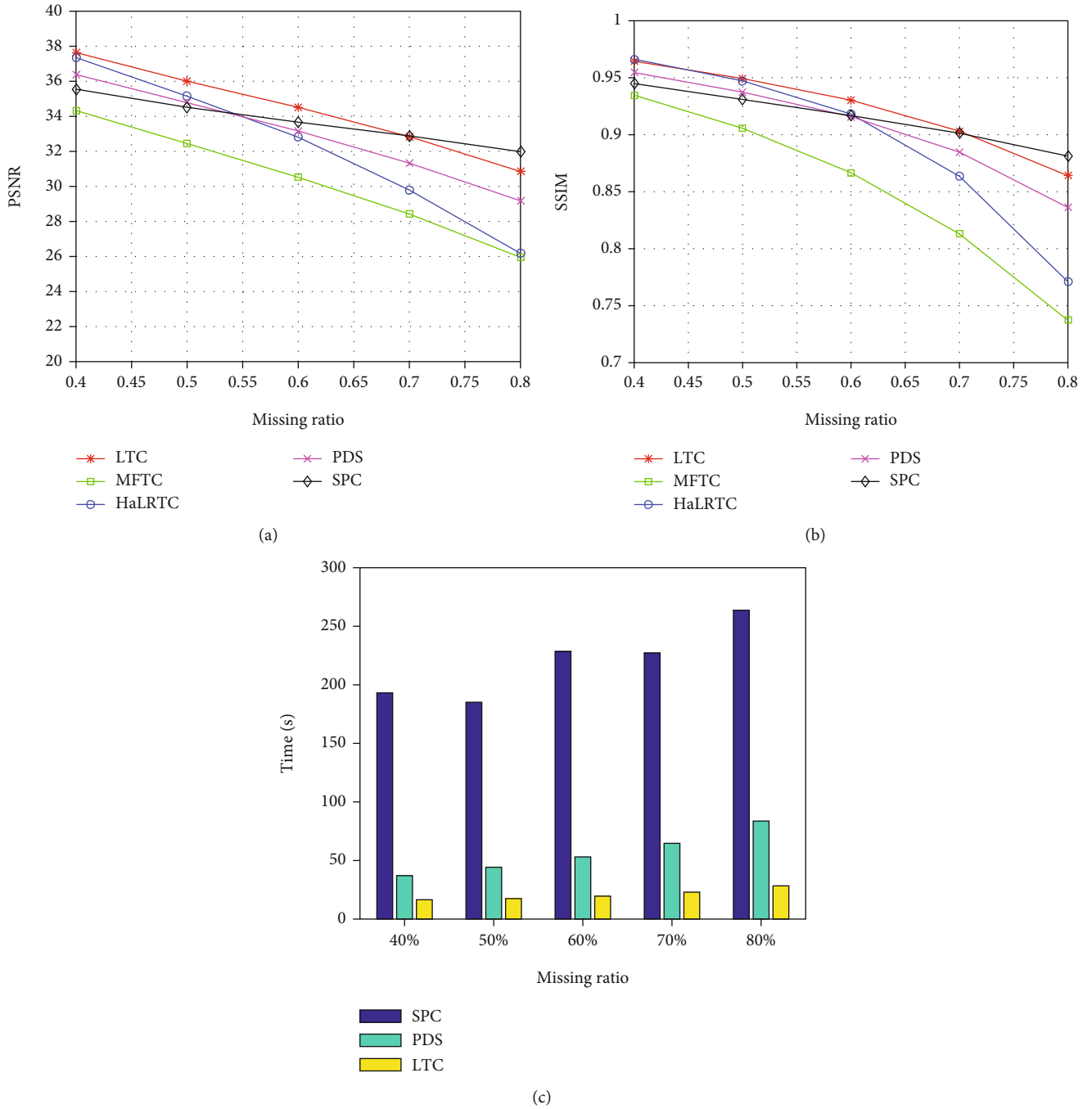
Figure 9: Comparison of performances by five LRTC methods on the video. (a) Comparison of PSNR, (b) comparison of SSIM, and (c) comparison of running time.

Figure 9 shows performance comparison of tensor completion for video "suzie" by our proposed LTC and the comparing four algorithms in PSNR, SSIM, and running time. The MRs are set as 40%, 50%, 60%, 70%, and 80%, respectively. Apparently, PDS, SPC, and LTC perform better than MFTC and HaLRTC when the MR is high. While ensuring the performance, LTC achieves about 90% and 50% running time reduction compared to SPC and PDS, respectively. It means that introducing matrix factorization can effectively reduce the running time of recovering incomplete videos.

5. Conclusion

In this paper, we propose a low-complexity tensor completion scheme. Our model takes advantage of SNN to exploit the low-rankness, TV and framelet to recover details and characterize the smoothness, and matrix factorization to improve the efficiency. Besides, a novel tensor initialization method is proposed to accelerate convergence speed. Moreover, an efficient ADMM-based algorithm is developed to solve the SMF model. The numerical experiments on

synthetic and real-world data demonstrate both the efficiency and effectiveness of proposed LTC scheme for tensor completion. Smoothing constraints for higher-dimensional data will be the research direction of tensor completion in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] N. D. Sidiropoulos, L. de Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.

[2] J. He, L. Li, and T. Shu, "Sparse nested arrays with spatially spread orthogonal dipoles: high accuracy passive direction finding with less mutual coupling," *IEEE Transactions on Aerospace and Electronic Systems*, 2021.

[3] X. Wang, M. Huang, and L. Wan, "Joint 2D-DOD and 2D-DOA estimation for coprime EMVS–MIMO radar," *Circuits, Systems, and Signal Processing*, pp. 1–17, 2021.

[4] X. Wang, L. T. Yang, D. Meng, M. Dong, K. Ota, and H. Wang, "Multi-UAV cooperative localization for marine targets based on weighted subspace fitting in SAGIN environment," *IEEE Internet of Things Journal*, 2021.

[5] F. Wen, J. Shi, and Z. Zhang, "Closed-form estimation algorithm for EMVS-MIMO radar with arbitrary sensor geometry," *Signal Processing*, vol. 186, article 108117, 2021.

[6] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1227–1243, 2018.

[7] T. Yokota, N. Lee, and A. Cichocki, "Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria," *IEEE Transactions on Signal Processing*, vol. 65, no. 5, pp. 1196–1206, 2017.

[8] M. Zhou, Y. Liu, Z. Long, L. Chen, and C. Zhu, "Tensor rank learning in CP decomposition via convolutional neural network," *Signal Processing: Image Communication*, vol. 73, pp. 12–21, 2019.

[9] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, no. 2, article 025010, 2011.

[10] J. Douglas Carroll, S. Pruzansky, and J. B. Kruskal, "Candelinc: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters," *Psychometrika*, vol. 45, no. 1, pp. 3–24, 1980.

[11] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[12] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.

[13] Y. Liu and F. Shang, "An efficient matrix factorization method for tensor completion," *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 307–310, 2013.

[14] Z. Wang and C. Ling, "Lattice Gaussian sampling by Markov chain Monte Carlo: bounded distance decoding and trapdoor sampling," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3630–3645, 2019.

[15] Z. Wang and C. Ling, "On the geometric ergodicity of Metropolis-Hastings algorithms for lattice Gaussian sampling," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 738–751, 2018.

[16] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3660–3677, 2012.

[17] Y. Liu, S. Wu, X. Huang, B. Chen, and C. Zhu, "Hybrid CS-DMRI: periodic time-variant subsampling and omnidirectional total variation based reconstruction," *IEEE Transactions on Medical Imaging*, vol. 36, no. 10, pp. 2148–2159, 2017.

[18] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[19] X. Li, Y. Ye, and X. Xu, "Low-rank tensor completion with total variation for visual data inpainting," in *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, California, USA, 2017.

[20] Y. Chen, S. Wang, and Y. Zhou, "Tensor nuclear norm-based low-rank approximation with total variation regularization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1364–1377, 2018.

[21] T. Yokota and H. Hontani, "Simultaneous tensor completion and denoising by noise inequality constrained convex optimization," *IEEE Access*, vol. 7, pp. 15669–15682, 2019.

[22] T. Yokota, Q. Zhao, and A. Cichocki, "Smooth PARAFAC decomposition for tensor completion," *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5423–5436, 2016.

[23] D. C. Dobson and F. Santosa, "Recovery of blocky images from noisy and blurred data," *SIAM Journal on Applied Mathematics*, vol. 56, no. 4, pp. 1181–1198, 1996.

[24] J. F. Cai, R. H. Chan, and Z. Shen, "A framelet-based image inpainting algorithm," *Applied and Computational Harmonic Analysis*, vol. 24, no. 2, pp. 131–149, 2008.

[25] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[26] A. Ron and Z. Shen, "A ne systems in L2 (IRd): the analysis of the analysis operator," *J. Funct. Anal.*, 1995.

[27] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Mathematical Programming Computation*, vol. 4, no. 4, pp. 333–361, 2012.

[28] Y. Shen, Z. Wen, and Y. Zhang, "Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization," *Optimization Methods and Software*, vol. 29, no. 2, pp. 239–263, 2014.

WILEY | Hindawi

Research Article

Using DTMB-Based Passive Radar for Small Unmanned Aerial Vehicle Detection

Huijie Zhu [iD],[1,2] **Lijun Wang,**[1,2] **and Mingqian Liu**[3]

[1]*Science and Technology on Communication Information Security Control Laboratory, Jiaxing 314033, China*
[2]*The 36th Research Institute of China Electronics Technology Group Corporation, Jiaxing 314033, China*
[3]*State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China*

Correspondence should be addressed to Huijie Zhu; zhuhuijie@zju.edu.cn

There is inevitable polarization angle deviation between the target echo signal and the direct path signal of illuminator of opportunity (IO) in passive radar. In order to investigate the potential performance loss in target detection induced by the random deviation, small unmanned aerial vehicle (UAV) detection experiments with digital television terrestrial multimedia broadcasting- (DTMB-) based passive radar are conducted in this paper. Experimental results show that the polarization angles of the clutter signal and target echo signal are inconsistent. When the polarization diversity technology is used to suppress the clutter signal, the processing performance of the target echo signal may be reduced. On the premise that clutter is effectively suppressed by the processing algorithm, polarization synthesis can maximize the target echo signal processing gain. The effectiveness of the target localization algorithm combining time difference of arrival (TDOA) and direction of arrival (DOA) is also verified with polarization diversity reception in this paper.

1. Introduction

Passive radar is a type of special bistatic or multistatic radar which exploits the existing electromagnetic signals in space as illuminators of opportunity (IOs) and realizes target detection and localization through passive reception. As its features of no dedicated transmitter and no frequency assignment, passive radar has attracted wide attention from academia and industry. FM radio [1, 2], digital TV [3, 4], mobile communications [5, 6], Wi-Fi [7, 8], WiMAX [9–11], satellite communications [12, 13], and many other IOs have been exploited for passive radar, which makes it achieve remarkable development over the years.

The IOs are not purposely designed for target detection; as a result, passive radar needs complicated signal processing to ensure its performance. However, the cost of signal processing has been greatly reduced with the development of high-performance computing technology. Furthermore, the development of software-defined radio technology makes passive radar be developed towards software-defined radar. The passive radar researchers only need to focus on algo-

rithms rather than hardware platforms, which greatly promotes the iterative evolution of passive radar processing algorithms. The signal processing of passive radar has approached its ultimate performance at present, and it is urgent to find new ways to improve the performance of passive radar.

In recent years, polarization diversity is generally considered a feasible way to further improve the passive radar performance [1, 14–18]. However, most studies focused on realizing clutter suppression by taking advantage of polarization diversity technology. While the clutter signal is effectively suppressed, the detection performance of the target echo signal may be limited by the polarization angle deviation between the clutter signal and the echo signal.

In order to verify the performance of the polarization diversity-based echo signal detection, field experiments with dual-polarization reception are carried out in this paper. The digital television terrestrial multimedia broadcasting (DTMB) signal is exploited as the IO signal while the unmanned aerial vehicle (UAV) of DJI Mavic 2 is selected as the target. DTMB is an international terrestrial high-

definition television broadcasting standard led by China. The bandwidth of the DTMB signal is about 8 MHz while the transmission power is usually between 300 and 1000 watts.

The rest of this paper is organized as follows. Section 2 introduces the signal model for passive radar in detail. Section 3 analyzes the polarization diversity for clutter suppression and target detection, together with the polarization-induced problem. Section 4 discusses the experimental process and results. Finally, conclusions and future work are summarized in Section 5.

2. Signal Model for Passive Radar

2.1. Signal Model for Receiving Channels. A passive radar system generally consists of two synchronous receiving channels, which are called the reference channel and surveillance channel, respectively. The received signal $x_r(t)$ of the reference channel can be described as

$$x_r(t) = C_r \cdot s(t) + \sum_{i=1}^{L} C_{r,i} \cdot s(t - \tau_i) + w_r(t), \qquad (1)$$

where $s(t)$ is the pure signal transmitted by the IO source, C_r is the amplitude, L is the number of multipath reflection, $C_{r,i}$ and τ_i are the amplitude and time delay of the i_{th} multipath, respectively, and $w_r(t)$ is the noise in the reference channel. $x_r(t)$ usually has a high signal-to-noise ratio (SNR) as the reference antenna is in the direction of the IO when the passive radar is in operation. Especially for the digital modulated signal, we can reconstruct the pure signal $s(t)$ from $x_r(t)$ by means of demodulation and remodulation [4]; (1) can then be simply rewritten as

$$x_r(t) \approx C_r \cdot s(t). \qquad (2)$$

The received signal $x_s(t)$ of the surveillance channel can be written as

$$x_s(t) = \sum_{i=0}^{N_c} C_{s,i} \cdot s(t - \tau_{s,i}) + \sum_{i=N_c+1}^{N_c+N_s} C_{s,i} \cdot s(t - \tau_{s,i}) + \sum_{i=1}^{N_m} C_{m,i} \cdot s$$
$$\cdot (t - \tau_{m,i}) e^{j2\pi f_i t} + w_s(t); \qquad (3)$$

in (3), $C_{s,i}$ and $\tau_{s,i}$ are the amplitude and time delay of the i_{th} stationary object echo (including multipath), respectively, and $i = 0$ is for the direct path signal (DPS) received from the side lobe of the surveillance antenna. $C_{m,i}$, $\tau_{m,i}$, and f_i are the amplitude, time delay, and Doppler frequency of the i_{th} moving target echo, respectively. N_c is the number of clutter echoes (excluding the DPS), N_s is the number of stationary targets, N_m is the number of moving targets, and $w_s(t)$ is the noise in the surveillance channel.

2.2. Signal Model for Clutter Suppression. The energy of the target echo is far weaker than that of clutter in passive radar. Clutter suppression is crucial for the operation of passive radar as it may mask target echoes. The least-squares (LS)

matrix solution, sometimes referred to as the extensive cancelation algorithm (ECA) or direct matrix inversion (DMI), is a widely used method as its inherently parallel algorithm structure that can be perfectly accelerated by hardware implementation [19, 20]. According to (3), the signal model for the surveillance channel can be written in a discrete form after sampling as

$$x_s(n) = x_c(n) + x_t(n) + w_s(n)$$
$$= \sum_{i=0}^{K_c} C_{s,i} \cdot s(n - i) + x_t(n) + w_s(n), \qquad (4)$$

where $x_c(n)$, $x_t(n)$, and $w_s(n)$ are the discrete form of clutter echoes, target echoes (both stationary and moving), and noise, respectively. $s(n)$ is the discrete form of $s(t)$, K_c is the maximum integer number corresponding to a signal delay in samples of the clutter echoes, and (4) can be rewritten in a compact matrix form as

$$\mathrm{X}_s = \mathrm{S} \cdot \mathrm{C}_s + \mathrm{X}_t + \mathrm{W}_s; \qquad (5)$$

in (5),

$$\mathrm{X}_s = [x_s(n)\, x_s(n+1) \,\cdots\, x_s(n+N)]^T, \qquad (6)$$

$$\mathrm{S} = \begin{bmatrix} s(n) & s(n-1) & \cdots & s(n-K_c) \\ s(n+1) & s(n) & \cdots & s(n-K_c+1) \\ \vdots & \vdots & \ddots & \vdots \\ s(n+N) & s(n+N-1) & \cdots & s(n+N-K_c) \end{bmatrix}, \qquad (7)$$

$$\mathrm{C}_s = \begin{bmatrix} C_{s,0} & C_{s,1} & \cdots & C_{s,K_c} \end{bmatrix}^T, \qquad (8)$$

$$\mathrm{X}_t = [x_t(n) \quad x_t(n+1) \quad \cdots \quad x_t(n+N)]^T, \qquad (9)$$

$$\mathrm{W}_s = [w_s(n) \quad w_s(n+1) \quad \cdots \quad w_s(n+N)]^T, \qquad (10)$$

where N is the number of available samples and $[\cdot]^T$ indicates matrix transposition.

The ECA-based solution is first to solve the following optimization problem:

$$\min_{C} J = \|\mathrm{X}_s - \mathrm{S} \cdot \mathrm{C}\|^2. \qquad (11)$$

The solution of (11) can be found by calculating the pseudoinverse of **S**:

$$\mathrm{C} = \left(\mathrm{S}^H \mathrm{S}\right)^{-1} \mathrm{S}^H \mathrm{X}_s, \qquad (12)$$

where $[\cdot]^H$ indicates matrix conjugate transposition. The calculated coefficient vector **C** represents estimated values of the clutter amplitudes for successive delays of the reference signal. Then, the clutter suppression can be expressed as

$$\mathrm{X}_{\text{rem}} = \mathrm{X}_s - \mathrm{S} \cdot \mathrm{C}; \qquad (13)$$

in (13),

$$X_{\mathrm{rem}} = \begin{bmatrix} x_{\mathrm{rem}}(n) & x_{\mathrm{rem}}(n+1) & \cdots & x_{\mathrm{rem}}(n+N) \end{bmatrix}^T \tag{14}$$

is the residual target echo signal vector after clutter cancellation.

For the operation of passive radar, a substitution of $x_r(t)$ for $s(t)$ is usually valid due to the high SNR. As a result, the clutter suppression in practice is performed as

$$X_{\mathrm{rem}} = X_s - X_r \left(X_r^H X_r \right)^{-1} X_r^H X_s\,; \tag{15}$$

in (15),

$$X_r = \begin{bmatrix} x_r(n) & x_r(n-1) & \cdots & x_r(n-K_c) \\ x_r(n+1) & x_r(n) & \cdots & x_r(n-K_c+1) \\ \vdots & \vdots & \ddots & \vdots \\ x_r(n+N) & x_r(n+N-1) & \cdots & x_r(n+N-K_c) \end{bmatrix}, \tag{16}$$

where $x_r(n)$ is the discrete form of $x_r(t)$.

2.3. Signal Model for Target Detection and Localization.
Target detection in passive radar is achieved by applying a cross-ambiguity function (CAF) to the residual target echo and the time-Doppler shifted visions of the reference signal. The calculation process is defined as

$$\psi(l, k) = \sum_{n=0}^{N-1} x_{\mathrm{rem}}(n) x_r^*(n-l) \mathrm{e}^{-j2\pi nk/N}\,; \tag{17}$$

by maximizing the CAF, the bistatic parameter pair of the strongest target echo signal can be estimated as

$$(l_t, k_t) = \arg\max_{l,k} |\psi(l, k)|\,; \tag{18}$$

then, the estimated time delay τ_t and Doppler-shift f_t can be derived as

$$\begin{aligned} \tau_t &= \frac{l_t}{f_s}, \\ f_t &= k_t \frac{f_s}{N}. \end{aligned} \tag{19}$$

The time difference of arrival (TDOA) measurement method for finding a target position has been widely used in the presence of multiple transmitters or multiple receivers [21–24]. The main target localization method used for bistatic passive radar is based on the measurements of bistatic range and direction of arrival (DOA) [25–31]. Different from the beamforming-based DOA estimation technique, a 2-D interferometric approach is exploited to estimate the DOA of the target echo in this paper. Though the beamforming technique can enhance the signal-to-disturbance ratio, the
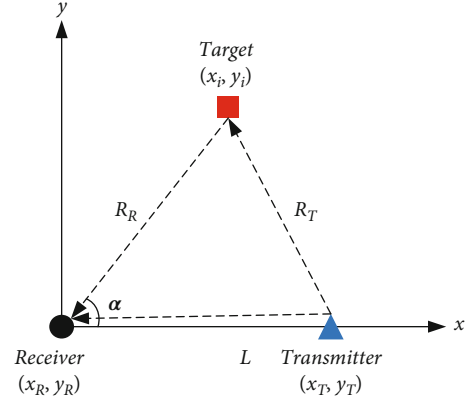

FIGURE 1: Bistatic plane within a 2-D coordinate system.

interferometric approach is simple and sufficient to verify the superiority of polarization diversity.

According to (17) and (18), the CAF results of the M surveillance channels can be described as

$$\begin{aligned} \psi(l_t, k_t) &= \begin{bmatrix} \psi_0(l_t, k_t) & \psi_1(l_t, k_t) & \cdots & \psi_{M-1}(l_t, k_t) \end{bmatrix}^T \\ &= \mathrm{a}(\mathbf{d}, \alpha) \sum_{n=0}^{N-1} x_{\mathrm{rem}}(n) x_r^*(n-l_t) \mathrm{e}^{-j2\pi nk_t/N}, \end{aligned} \tag{20}$$

where $\mathbf{a}(\mathbf{d}, \alpha) = [a_0(d_0, \alpha), a_1(d_1, \alpha), \cdots, a_{M-1}(d_{M-1}, \alpha)]^T$ is the steering vector of the antenna array, d_m ($m = 0, 1, \cdots, M-1$) is the distance difference between the m_{th} and the reference antenna elements, and α is the DOA of the target echo. Therefore, DOA estimation based on the CAF results is completely equivalent to the interferometer direction finding.

Figure 1 shows the bistatic plane within a 2-D coordinate system, where the coordinates of receiver, transmitter, and target are (x_R, y_R), (x_T, y_T), and (x_i, y_i), respectively. Without loss of generality, we can set (x_R, y_R) as $(0, 0)$ for simplification. Then, the baseline L can be calculated from the known positions of both receiver and transmitter as

$$L = \sqrt{x_T^2 + y_T^2}\,; \tag{21}$$

the range between the transmitter and the target R_T can be derived from the unknown position of the target and the known transmitter as

$$R_T = \sqrt{(x_T - x_i)^2 + (y_T - y_i)^2}\,; \tag{22}$$

similarly, R_R shown in Figure 1 can be described as

$$R_R = \sqrt{x_i^2 + y_i^2}\,; \tag{23}$$

with the estimations of time delay τ_t and target echo DOA α, additional equations can be obtained as

$$\begin{cases} c \cdot \tau_t = R_T + R_R - L, \\ \tan \alpha = \dfrac{y_i}{x_i}, \end{cases} \tag{24}$$

where c is the speed of propagation.

The closed-form localization result can be realized by solving (22), (23), and (24) as

$$\begin{cases} x_i = \dfrac{c \cdot \tau_t \cdot (2L + c \cdot \tau_t) \cos \alpha}{2(L + c \cdot \tau_t) - 2x_T - 2y_T \sin \alpha}, \\ y_i = \dfrac{c \cdot \tau_t \cdot (2L + c \cdot \tau_t) \sin \alpha}{2(L + c \cdot \tau_t) - 2x_T - 2y_T \sin \alpha}. \end{cases} \tag{25}$$

3. Polarization Diversity and Problem Analysis

There is inevitable polarization angle deviation between the echo signal and the incident signal. The effects of polarization have not been taken into account in the signal model above. When the polarization angle is present, (3) can be rewritten as

$$\begin{aligned} x_s(t) = &\sum_{i=0}^{N_c} C_{s,i} \cdot s(t - \tau_{s,i}) \cdot \cos \theta_{s,i} + \sum_{i=N_c+1}^{N_c+N_s} C_{s,i} \cdot s(t - \tau_{s,i}) \cdot \cos \theta_{s,i} \\ &+ \sum_{i=1}^{N_m} C_{m,i} \cdot s(t - \tau_{m,i}) e^{j2\pi f_i t} \cdot \cos \theta_{m,i} + w_s(t), \end{aligned} \tag{26}$$

where $\theta_{s,i}$ and $\theta_{m,i}$ are the polarization angles of the i_{th} stationary object echo and the i_{th} moving object echo, respectively, and they are parameters that we can control under polarization diversity reception.

Most of the IOs for passive radar are transmitted by horizontal or vertical polarization antennas. Assume we use an orthogonal dual-polarized antenna as the surveillance antenna, the received signal vector can be described as

$$X(n) = \begin{bmatrix} x_{0^\circ}(n) \\ x_{90^\circ}(n) \end{bmatrix} = \begin{bmatrix} \cos \theta \\ \sin \theta \cdot e^{j\varphi} \end{bmatrix} x_s(n), \tag{27}$$

where $x_{0^\circ}(n)$ is one output of the dual-polarized antenna and $x_{90^\circ}(n)$ is the other, θ is the polarization angle determined by the antenna, and φ is the phase difference between the two orthogonal channels. After polarization diversity reception, the arbitrary polarization angle can be realized by polarization synthesis as follows:

$$\begin{aligned} \widehat{x}_s(n, \vartheta, \phi) &= \begin{bmatrix} \cos \vartheta & \sin \vartheta \cdot e^{j\phi} \end{bmatrix}^* \begin{bmatrix} x_{0^\circ}(n) \\ x_{90^\circ}(n) \end{bmatrix} \\ &= \begin{bmatrix} \cos \vartheta & \sin \vartheta \cdot e^{-j\phi} \end{bmatrix} \begin{bmatrix} \cos \theta \\ \sin \theta \cdot e^{j\varphi} \end{bmatrix} x_s(n) \\ &= \left(\cos \vartheta \cos \theta + \sin \vartheta \sin \theta \cdot e^{j(\varphi-\phi)} \right) \cdot x_s(n); \end{aligned} \tag{28}$$

in (28), $[\cdot]^*$ indicates the conjugate operation. By substituting $\phi = \varphi$ in (28), we obtain

$$\widehat{x}_s(n, \vartheta) = \cos (\theta - \vartheta) \cdot x_s(n), \tag{29}$$

which means that the polarization angle has changed. We can then rewrite (15) as

$$X_{\text{rem}}(\vartheta, \phi) = \widehat{X}_s - X_r \left(X_r^H X_r \right)^{-1} X_r^H \widehat{X}_s. \tag{30}$$

The essence of clutter suppression by taking advantage of polarization diversity technology can be expressed as

$$\min_{\vartheta, \phi} J = \| X_{\text{rem}}(\vartheta, \phi) \|^2; \tag{31}$$

the optimal polarization synthesis parameter pair for clutter suppression can be described as

$$(\theta_c, \varphi_c) = \arg \min_{\vartheta, \phi} \| X_{\text{rem}}(\vartheta, \phi) \|^2. \tag{32}$$

On the premise that clutter can be effectively suppressed by a processing algorithm, target detection in passive radar will only depend on the CAF operation. The optimal polarization synthesis parameter pair for the (l_t, k_t) target detection can be obtained as

$$\begin{aligned} (\theta_t, \varphi_t) &= \arg \max_{\vartheta, \phi} \left| \widehat{\psi}(l_t, k_t, \vartheta, \phi) \right| \\ &= \arg \max_{\vartheta, \phi} \left| \sum_{n=0}^{N-1} x_{\text{rem}}(n, \vartheta, \phi) x_r^*(n - l_t) e^{-j2\pi n k_t / N} \right|, \end{aligned} \tag{33}$$

where $x_{\text{rem}}(n, \vartheta, \phi)$ is the element of $X_{\text{rem}}(\vartheta, \phi)$.

Therefore, when the parameter pairs of (θ_c, φ_c) and (θ_t, φ_t) are inconsistent, clutter suppression based on the polarization diversity technology will result in performance loss for the target detection.

4. Experimental Analysis

In order to verify the theoretical analysis above, we conduct an experimental system which consists of a reference antenna, surveillance antenna, software-defined radio (SDR) receiver, and heterogeneous computing platform. The experimental antenna system is shown in Figure 2. It is mainly composed of a vertical polarization antenna and a dual-polarized antenna.

4.1. Reference Signal Reconstruction. It can be clearly seen from Figure 2 that the DTMB transmitting antenna is visible to the reference antenna. Therefore, the reference signal has a high SNR. Further, reference signal reconstruction is adopted to remove the inevitable multipath and noise. The channel decoding and coding operations are omitted in our processing due to the high SNR, and the whole reference signal reconstruction process is given in Figure 3.
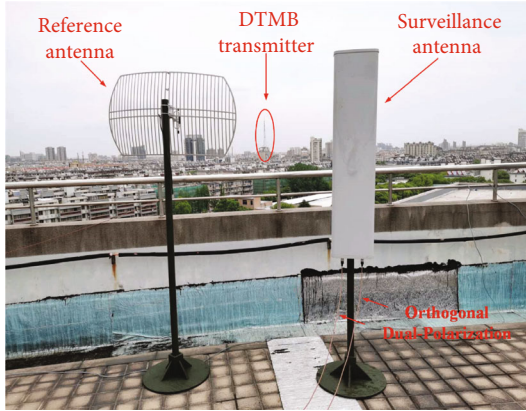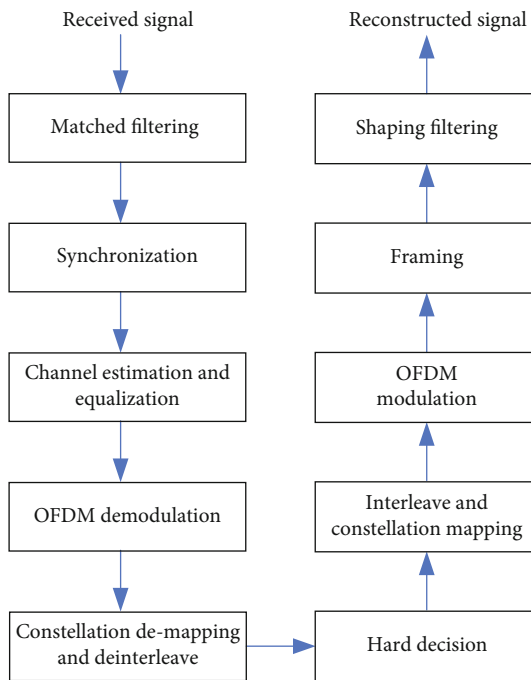
Figure 2: The experimental antenna system.

Figure 3: Reference signal reconstruction process.

Matched filtering is first performed to reduce intersymbol interference for the received signal. Then, synchronization, channel estimation and equalization, OFDM demodulation, and other operations are carried out in turn.

A multichannel synchronous SDR receiver is used for signal acquisition, which can turn the radio frequency and sampling rate flexibly. The carrier frequency of the DTMB signal exploited in our experiments is 754 MHz, and the used sampling rate is 30.24 MHz, which means that the oversampling rate is 4. The duration of the reference signal used for demodulation is 0.1 s. Figure 4 shows the obtained one-frame constellation after OFDM demodulation. According to the DTMB standard, the frame body of the DTMB signal consists of two different types of symbols, the system information (SI) symbol and the streaming data symbol. The SI symbol is mapped to the BPSK constellation fixedly while the streaming data symbol can be mapped to one of the
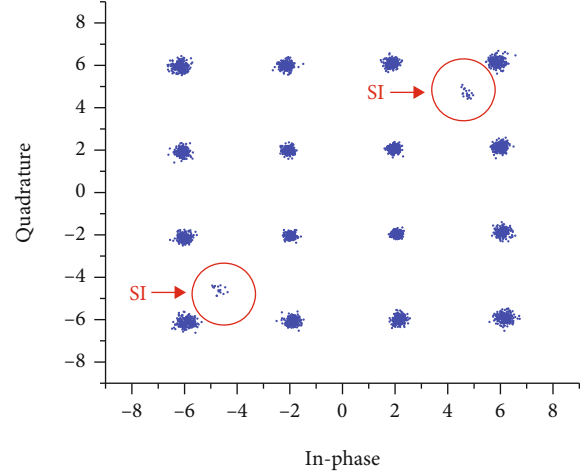

Figure 4: The constellation of reference signal after demodulation (one frame).

4QAM, 16QAM, 32QAM, and 64QAM constellations. Both the BPSK and 16QAM constellations can be seen in Figure 4, in which all the constellation points are focused.

After constellation demapping and deinterleave, we make a hard decision to recover the transmitted bit sequence. Finally, the recovered bit sequence is processed by the operations of interleave, constellation mapping, OFDM modulation, framing, and shaping filtering, which are the same as the transmitter, to reconstruct the pure reference signal.

4.2. Target Detection. The DJI Mavic 2 UAV, which has unfold dimensions of 322 mm × 242 mm × 84 mm is used as the target to be detected in our experiments. The clutter suppression is first conducted by using the ECA-based solution with the reconstructed reference signal and one of the dual-polarized echo signals, and the result is given in Figure 5. It shows that a suppression capacity of about 30 dB is achieved. Since the reconstructed reference signal cannot be consistent with the clutter components in terms of frequency deviation and time delay variation, the residual signal still has some noise components.

The decimation filter and fast Fourier transform (FFT) are used after clutter suppression to realize the CAF operation as (17). The detailed schematic is shown in Figure 6.

The obtained bistatic Rang-Doppler (RD) map after the CAF processing is shown in Figure 7. As can be seen, there is an obvious target that the bistatic range equals 119.05 m and the bistatic Doppler frequency equals 35 Hz.

4.3. Polarization Diversity for Clutter Suppression. In the clutter signal, the polarization angle of the DPS component is different from that of the multipath components. Therefore, we cannot suppress every component by the way of polarization diversity.

We evaluate the power of the polarization synthesized signal as (28). When $-180° \le \theta \le 180°$ and $-180° \le \varphi \le 180°$, the obtained signal power is shown in Figure 8. For different polarization parameter pairs, the power of the synthesized signal can differ by larger than 10.5 dB.
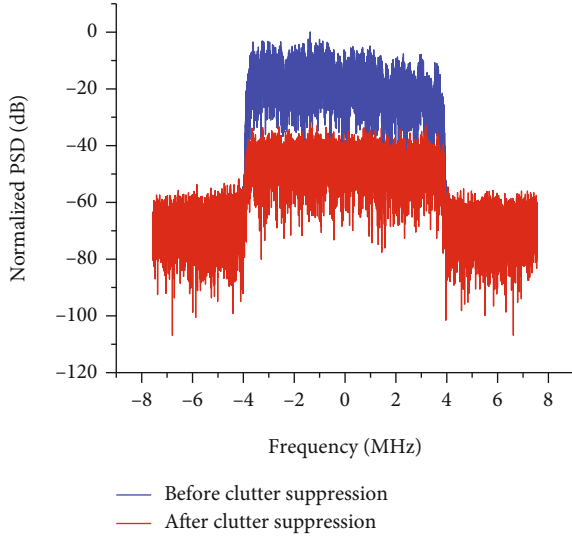
FIGURE 5: PSD comparison: before and after clutter suppression.

FIGURE 7: RD map after CAF processing.

FIGURE 6: CAF processing schematic.

FIGURE 8: Normalized power (dB) of the polarization synthesized signal for $-180° \leq \theta \leq 180°$ and $-180° \leq \varphi \leq 180°$.

mum value is at $(-62°, -32°)$. As a result, the value of -32° is the estimation of the phase difference between the two orthogonal channels, and the estimated value of 28° is treated as the equivalent polarization angle of the composite clutter signal. The optimal polarization angle used for clutter suppression has a value of -62° in our experiments.

4.4. Polarization Diversity for Target Detection. The target detection performance under different polarization angles is also evaluated in our experiments. Since different polarization angles may result in different residual noise components by the ECA-based solution, the SNR of the RD map defined is used in the evaluation:

$$SNR_t = \frac{|\psi(\tau_t, f_t)|}{\underset{\tau,f}{\mathrm{mean}} |\psi(\tau, f)|}. \tag{34}$$

The obtained normalized SNR_t map is shown in Figure 10.

There are four maximums and four minimums in the range of [-180°, 180°] due to the symmetry of trigonometric functions. Then, we narrow the range to [-90°, 90°] and obtain the partial but complete results shown in Figure 9. In this range, the synthesized signal gets its maximum value with the parameter pair (θ, φ) at $(28°, -32°)$ while its mini-
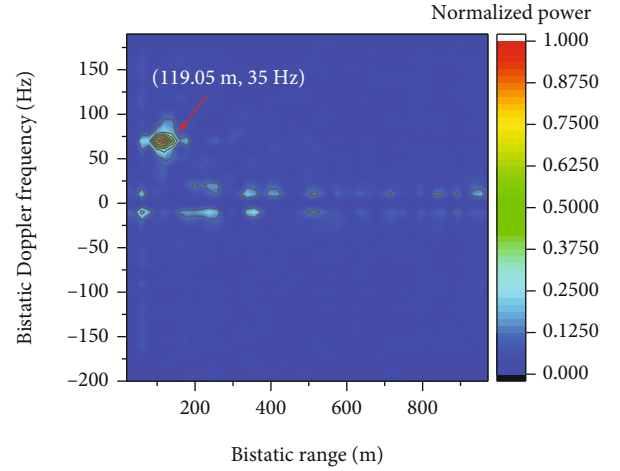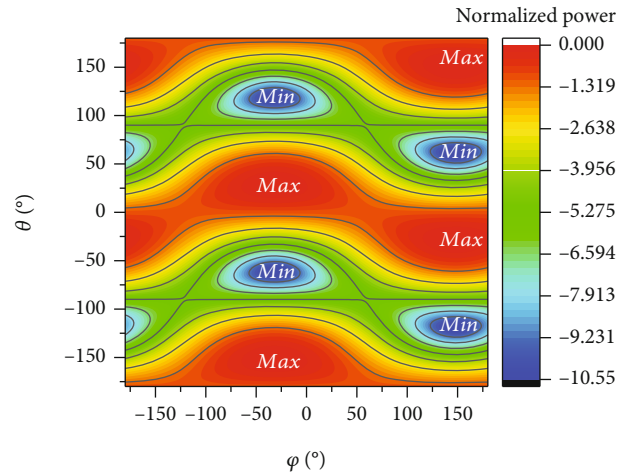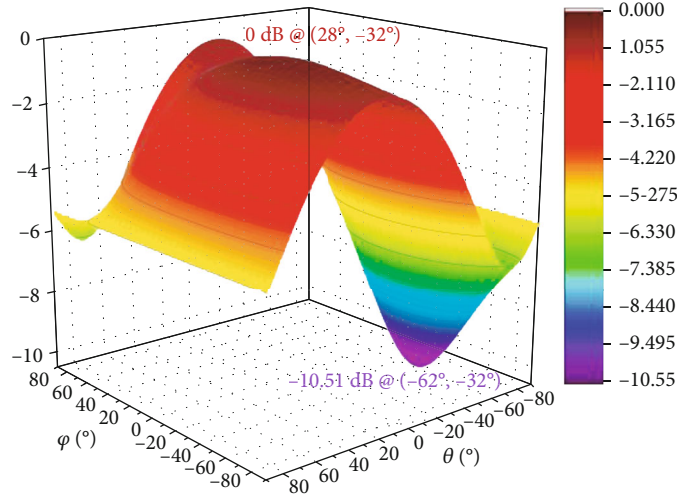
FIGURE 9: Normalized power (dB) of the polarization synthesized signal for $-90 \leq \theta \leq 90°$ and $-90° \leq \varphi \leq 90°$.
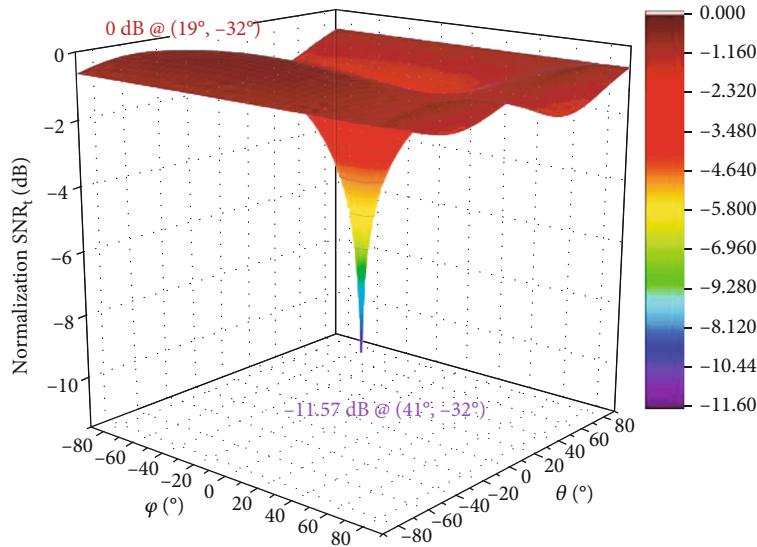
FIGURE 10: Normalized SNR_t (dB) map for $-90° \leq \theta \leq 90°$ and $-90° \leq \varphi \leq 90°$.

It shows that SNR_t gets its maximum value with the parameter pair (θ, φ) at (-49°, -32°) while its minimum value at (41°, -32°), and the difference between the maximum and minimum is larger than 11.5 dB. As the phase difference between the two orthogonal channels is only related to the receiver, it can be seen that the value of φ for the extremum in Figure 10 is the same as that in Figure 9. However, the polarization angle of -49° for the maximum of SNR_t is inconsistent with the optimal value of -62° for clutter suppression. As a result, it proves that clutter suppression based on the polarization diversity technology may result in performance loss for the target detection due to the randomness of the polarization angle difference.

Table 1 shows the obtained normalized SNR_t with several polarization angles we are interested in. We only use one of the dual-polarization echo signals in the case of $\theta = 0°$. As the noises in the two orthogonal polarized channels are not

TABLE 1: Comparison of SNR_t.

θ	φ	Normalized SNR_t (dB)
0°	0°	-1.22287
28°	-32°	-3.19481
-62°	-32°	-0.58790
-49°	-32°	0

correlated with each other, the random noise components will be superimposed after polarization synthesis. In addition to the random noise, the clutter signal and target echo signal in the synthesized signal are affected dramatically by changing the polarization angle parameter. Only when the polarization angle adopted in the polarization synthesis matches the actual signal polarization angle, the SNR of the synthesized signal can be maximized. Table 1 shows that

FIGURE 11: Array configuration of the 2-D interferometer.

polarization matching has a performance improvement of about 1.22 dB compared with single-polarization reception in our experiments.

As the stability of the ECA-based clutter suppression algorithm is related to the calculating precision of matrix operation, the effect of clutter suppression will not change with the amplitude of the clutter signal in a certain range. The algorithm always works within the stable range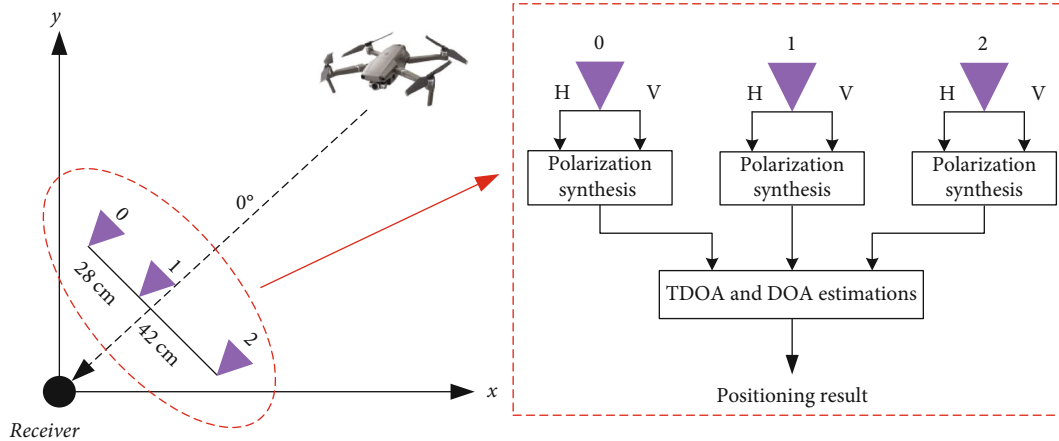 in our experiments, and the clutter signal can be well suppressed at all the polarization angles. Although there is a performance improvement of up to 2.61 dB compared to 28° at -62°, we need to point out that this is not the reason for better clutter suppression; the theoretical reason is that this polarization angle is closer to the actual polarization angle of -49°. When the clutter signal is too strong for the ECA algorithm to work normally, polarization diversity is an effective way to increase the additional capacity for clutter suppression. Though there may be performance loss for target detection, it is worthful for the normal operation of the system.

4.5. Target Localization. A small antenna array which consists of three elements is exploited to estimate the DOA of the UAV echo. Figure 11 shows the configuration of the antenna array, which consists of three antenna elements. Each antenna element is orthogonally dual-polarized as shown in the figure, where H represents horizontal polarization and V represents vertical polarization. The spacing is 28 cm for Element 0 and Element 1, while it is 42 cm for Element 1 and Element 2. Since a virtual spacing of 14 cm can be constructed by this nonuniform linear array, which is below the half-wavelength of about 20 cm in our experiments, it will not lead to the cyclically ambiguous problem in DOA estimation. For each dual-polarized element, we conduct polarization synthesis, clutter suppression, and TDOA estimation separately. The three processing results are further used for DOA estimation. The positioning result of the target is finally obtained after both TDOA and DOA estimations.

The essence of the 2-D interferometer direction finding is to determine the direction of the signal source by the phase difference formed by the received signal on different antenna elements. The wave-path difference between the two antenna
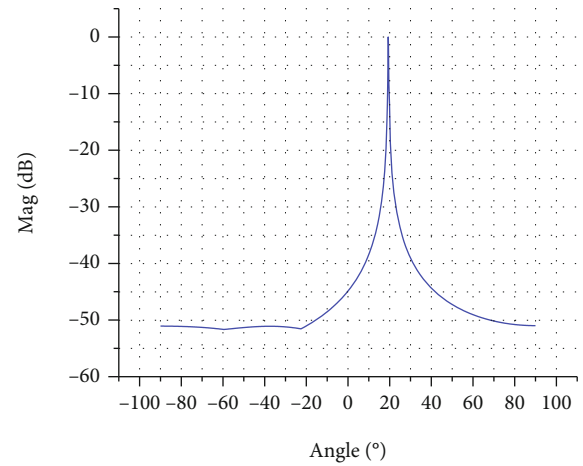

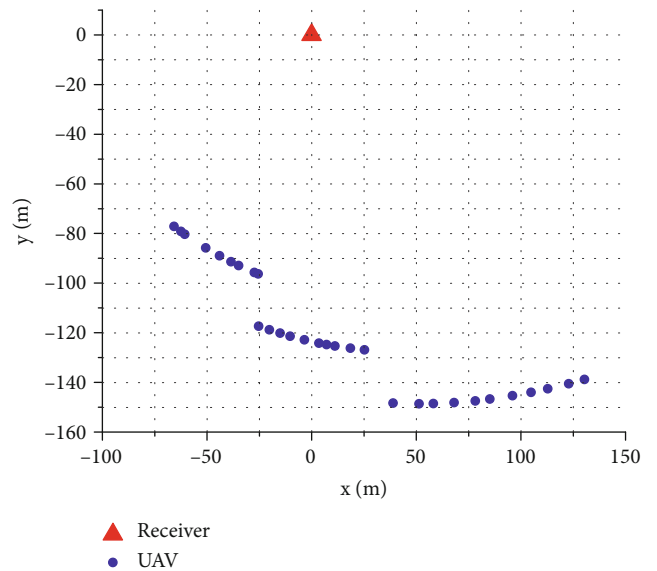FIGURE 12: DOA output of the 2-D interferometer.

FIGURE 13: Positioning result of the UAV with the combination of TDOA and DOA.

elements will cause the phase difference. Though the polarization angle is random, it is not instantaneous. As a result, we use the same value for the parameter of θ in polarization synthesis of all the three dual-polarized antenna elements. Since the parameter of φ is mainly related to different channels of the receiver, we use the premeasured values for it. We search optimal DOA estimation between -90° and 90°. The result of the direction finding at a certain moment is shown in Figure 12, which shows an optimal estimation of about 20°.

The combination of TDOA and DOA as (25) is adopted to locate the UAV. Figure 13 shows the positioning result of the UAV in a certain continuous time, which is consistent with its actual flight path in the experiments. The two obvious discontinuous sudden changes in Figure 13 are caused by the range resolution of the signal, which is about 39.68 m.

Figure 13 shows that the direction finding and localization of the echo signal can still be realized after polarization synthesis, which verifies the validity of the localization algorithm in passive radar with polarization diversity reception. The positioning distance may increase significantly at some polarization angles through polarization synthesis. However, since there is only 1.22 dB performance improvement in target detection compared with single-polarization reception in our experiments, the positioning distance will not increase obviously.

5. Conclusion

We have established an experimental system to evaluate the adverse effect of polarization angle on passive radar. By analyzing the experimental data, the necessity of polarization diversity to ensure the target detection performance is revealed. The effect of polarization diversity on clutter suppression is significant; especially when the clutter signal is strong, it is necessary to increase the additional suppression capacity through polarization diversity for the normal operation of the system. However, when the clutter suppression algorithm is sufficient to suppress the clutter, the significance of the polarization diversity will lie in matching the polarization angle of the target echo signal. When the two perfectly match, the maximum SNR of the target echo signal will be achieved, and the system will have the optimal target detection performance. The target localization capacity under polarization synthesis for multichannel is also investigated. We will further study the adaptive polarization synthesis strategy to achieve a compromise between clutter suppression and target detection.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

H.Z. proposed the method, performed the experiments, and wrote the paper; L.W. designed the experiments and analyzed the data; H.Z. and M.L. made the discussion; L.W. and M.L. gave the instructional suggestions.

Acknowledgments

References

[1] J. You, X. Wan, Y. Fu, and G. Fang, "Experimental study of polarisation technique on multi-FM-based passive radar," *IET Radar, Sonar & Navigation*, vol. 9, no. 7, pp. 763–771, 2015.

[2] G.-H. Park, Y.-K. Seo, and H.-N. Kim, "Range-Doppler domain-based DOA estimation method for FM-band passive bistatic radar," *IEEE Access*, vol. 8, pp. 56880–56891, 2020.

[3] T. Martelli, F. Colone, and R. Cardinali, "DVB-T based passive radar for simultaneous counter-drone operations and civil air traffic surveillance," *IET Radar, Sonar & Navigation*, vol. 14, no. 4, pp. 505–515, 2020.

[4] X. Zhang, J. Yi, X. Wan, and Y. Liu, "Reference signal reconstruction under oversampling for DTMB-based passive radar," *IEEE Access*, vol. 8, pp. 74024–74038, 2020.

[5] M. Wielgo, P. Krysik, K. Klincewicz, L. Maslikowski, S. Rzewuski, and K. Kulpa, "Doppler only localization in GSM-based passive radar," in *in Proc. IEEE Int. Radar Conf*, pp. 1–6, 2014.

[6] R. S. A. R. Abdullah, A. A. Salah, N. H. A. Aziz, and N. E. A. Rasid, "Vehicle recognition analysis in LTE based forward scattering radar," in *in Proc. IEEE Radar Conf.*, pp. 1–5, 2016.

[7] T. Martelli, F. Colone, and P. Lombardo, "First experimental results for a WiFi-based passive forward scatter radar," in *in Proc. IEEE Radar Conf.*, pp. 1–6, 2016.

[8] W. Li, R. J. Piechocki, K. Woodbridge, C. Tang, and K. Chetty, "Passive WiFi radar for human sensing using a stand-alone access point," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1986–1998.

[9] Q. Wang, C. Hou, and Y. Lu, "An experimental study of WiMAX-based passive radar," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 12, pp. 3502–3510, 2010.

[10] T. Higgins, T. Webster, and E. L. Mokole, "Passive multistatic radar experiment using WiMAX signals of opportunity. Part 1: signal processing," *IET Radar, Sonar & Navigation*, vol. 10, no. 2, pp. 238–247, 2016.

[11] T. Webster, T. Higgins, and E. L. Mokole, "Passive multistatic radar experiment using WiMAX signals of opportunity. Part 2: multistatic velocity backprojection," *IET Radar, Sonar & Navigation*, vol. 10, no. 2, pp. 248–255, 2016.

[12] F. Santi, D. Pastina, and M. Bucciarelli, "Experimental demonstration of ship target detection in GNSS-based passive radar combining target motion compensation and track-before-detect strategies," *Sensors*, vol. 20, no. 3, p. 599, 2020.

[13] O. Cabrera, C. Bongioanni, F. Filippini, O. Sarabakha, F. Colone, and P. Lombardo, "Detecting drones and human beings with DVB-S based COTS passive radar for short-range surveillance," in *in Proc. IEEE Int. Radar Conf*, pp. 37–42, 2020.

[14] M. Conti, C. Moscardini, and A. Capria, "Dual-polarization DVB-T passive radar: experimental results," in *in Proc. IEEE Radar Conf*, pp. 1–5, 2016.

[15] F. Colone and P. Lombardo, "Polarimetric passive coherent location," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 2, pp. 1079–1097, 2015.

[16] F. Colone and P. Lombardo, "Non-coherent adaptive detection in passive radar exploiting polarimetric and frequency diversity," *IET Radar, Sonar & Navigation*, vol. 10, no. 1, pp. 15–23, 2016.

[17] Y. Yi, X. Wan, J. Yi, and X. Cao, "Polarization diversity technology research in passive radar based on subcarrier processing," *IEEE Sensors Journal*, vol. 19, no. 5, pp. 1710–1719, 2019.

[18] L. Wan, K. Liu, Y.-C. Liang, and T. Zhu, "DOA and polarization estimation for non-circular signals in 3-D millimeter wave polarized massive MIMO systems," *IEEE Trans. Wirel. Commun*, 2021.

[19] M. Malanowski, *Signal Processing for Passive Bistatic Radar*, Artech House, Norwood, MA, USA, 2019.

[20] P. Zhang, Y. Wu, J. Wang, and J. Qiao, "Real-time signal processing for FM-based passive bistatic radar using GPUs," in *in Proc. IEEE Int. Conf. Digital Signal Processing*, pp. 536–540, 2014.

[21] A. Noroozi and M. A. Sebt, "Target localization from bistatic range measurements in multi-transmitter multi-receiver passive radar," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2445–2449, 2015.

[22] M. Malanowski and K. Kulpa, "Two methods for target localization in multistatic passive radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 572–580, 2012.

[23] M. Liu, G. Liao, N. Zhao, H. Song, and F. Gong, "Data-driven deep learning for signal classification in industrial cognitive radio networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3412–3421, 2021.

[24] Y. Zhao, D. Hu, Y. Zhao, Z. Liu, and C. Zhao, "Refining inaccurate transmitter and receiver positions using calibration targets for target localization in multi-static passive radar," *Sensors*, vol. 19, no. 15, 2019.

[25] M. Ummenhofer, M. Kohler, J. Schell, and D. W. O'Hagan, "Direction of arrival estimation techniques for passive radar based 3D target localization," in *in Proc. IEEE Radar Conf*, pp. 1–6, 2019.

[26] L. Wan, Y. Sun, L. Sun, Z. Ning, and J. J. P. C. Rodrigues, "Deep learning based autonomous vehicle super resolution DOA estimation for safety driving," in *IEEE Trans. Intell. Transp. Syst*, 2020.

[27] G. Bournaka, J. Heckenbach, A. Baruzzi, D. Cristallini, and H. Kuschel, "A two stage beamforming approach for low complexity CFAR detection and localization for passive radar," in *in Proc. IEEE Radar Conf.,*, pp. 1–4, 2016.

[28] M. Liu, K. Yang, N. Zhao, Y. Chen, H. Song, and F. Gong, "Intelligent signal classification in industrial distributed wireless sensor networks based industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4946–4956, 2021.

[29] J. H. Huang, J. L. Garry, G. E. Smith, and C. J. Baker, "Array based passive radar target localization," in *in Proc. IEEE Radar Conf*, pp. 1–6, 2016.

[30] L. Wan, L. Sun, K. Liu, X. Wang, Q. Lin, and T. Zhu, "Autonomous vehicle source enumeration exploiting non-cooperative UAV in software defined Internet of vehicles," in *IEEE Trans. Intell. Transport. Syst*, 2020.

[31] P. Falcone, F. Colone, A. Macera, and P. Lombardo, "Two-dimensional location of moving targets within local areas using WiFi-based multistatic passive radar," *IET Radar, Sonar & Navigation*, vol. 8, no. 2, pp. 123–131, 2014.

WILEY | Hindawi

Research Article

Channel Estimation Approach with Low Pilot Overhead in FBMC/OQAM Systems

Jun Sun,[1,2] Xiaomin Mu,[1] Dejin Kong [ID],[3] Qian Wang,[3] Xinmin Li,[4] and Xing Cheng[5]

[1]School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
[2]School of Electronic and Information, Zhongyuan University of Technology, Zhengzhou 450007, China
[3]State Key Laboratory of New Textile Materials and Advanced Processing Technologies, The School of Electronic and
 Electrical Engineering, Wuhan Textile University, Wuhan 430200, China
[4]School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China
[5]School of Information and Communication Engineering, Beijing Information Science and Technology University,
 Beijing 100101, China

Correspondence should be addressed to Dejin Kong; djkou@wtu.edu.cn

In filter bank multicarrier with offset quadrature amplitude modulation (FBMC/OQAM) systems, a large pilot overhead is required due to the existence of the imaginary interference. In this paper, we present an approach to reduce the pilot overhead of channel estimation. A part of pilot overhead is used for transmitting data, and compensating symbols are required and designed to remove the imaginary interference. It is worthwhile to point out that the power of compensating symbols can be helpful for data recovery; hence, the proposed approach decreases the overhead of pilots significantly without the cost of additional pilot energy. In addition, the proposed scheme is extended into multiple input multiple output systems without the performance loss. Compared with the conventional preamble consisting of 3 columns symbols, the pilot overhead is equivalent to 2 column symbols in the proposed preamble. To verify the proposed preamble, numerical simulations are carried out with respects to bit error ratio.

1. Introduction

Currently, filter bank multicarrier with offset quadrature amplitude modulation (FBMC/OQAM) has been considered as a potential alternative [1–4] to the conventional orthogonal frequency division multiplexing (OFDM). In the past several years, FBMC/OQAM and other filter-based waveforms have been studied to overcome the disadvantages of OFDM. This paper focuses on the FBMC/OQAM technique and presents an effective solution to the channel estimation with low pilot overhead.

Unlike in OFDM, channel estimation in FBMC/OQAM is not a straightforward mission, which is related to that the waveform synthesis and analysis of FBMC/OQAM are not same as that of OFDM. Since the orthogonality condition only is met in real field [2, 5], FBMC/OQAM systems transmit real-valued symbols obtained by the real and imaginary parts of complex-valued QAM symbols, and there exists imaginary interferences among the transmitted real-valued symbols, called the intrinsic imaginary interference [6]. For the channel estimation, the imaginary interference has a crucial effect on the pilot design. The imaginary interference has to be eliminated to ensure the good system performance [7]. In particular, while in OFDM receiver data/pilot symbols that are perfectly separated after applying the FFT, the signal samples from the output of the analysis filter in an FBMC/OQAM receiver are subject to intersymbol interference (ISI), both along the time and frequency/subcarriers. ISI-free symbols could be achieved after the channel equalization and the operation of taking the real parts. In the absence of the channel, ISI appears as an imaginary component that adds to the real-valued data symbols which are carried by the FBMC/OQAM waveform, which has to be considered in the pilot design.

To avert the imaginary interference, a direct method is to disable the symbols surrounding the pilots, which results in a reduced spectral efficiency. By applying the following method, this spectral efficiency loss could be avoid [8]. One dummy symbol is set adjacent to the pilot to eliminate the imaginary interference of the FBMC/OQAM system. This dummy symbol was then called auxiliary pilot (AP) in [9]. Although it is helpful to remove the imaginary interference to pilots, the AP method effectively increases the transmit power. A more complex approach, but more efficient in terms of transmit power, was proposed in [10]. The data symbols surrounding pilots are performed by a linear coding so that the data symbols do not produce any imaginary interference. Nevertheless, to completely remove the imaginary interference, a long code length is required in the coding scheme, which in turn increases a large complexity at both of the transmitter and the receiver. In addition, it also introduces other complications, which are discussed in [11]. In [12], the authors presented the interference approximation method (IAM), in which it was revealed that the channel estimation performance is decided by the so-called pseudopilot power consisting of pilot and the imaginary interference from adjacent data. Afterwards, the modified versions of IAM were presented to improve the pseudopilot power by employing imaginary-valued pilots, i.e., IAM-C [13] and IAM-I [14]. However, it should be noted that the existing IAM-based methods require 3 columns real-valued symbols as pilot overhead to achieve the channel estimation. To reduce the pilot overhead of IAM-based schemes, the iterative algorithm was proposed to eliminate the imaginary interference for channel estimation [15]. Nevertheless, it suffers from the problems of high computational complexity. In [16], the authors proposed the pairs of the real pilots (POP) method, with only 2 columns pilots. However, by reason of the bad ability of against noise, the POP method exhibits poor channel estimation performance compared with the conventional IAM-based methods.

In this paper, we present an approach to reduce the pilot overhead of channel estimation in FBMC/OQAM systems. Compared with the conventional pilot structures, a part of pilot overhead is used for transmitting data, and compensating symbols are required and designed to eliminate the imaginary interference from data. Note that it is proven that the power of compensating symbols can be helpful for data recovery; hence, the proposed approach decreases the overhead of pilots significantly without the cost of additional pilot energy. Compared with the conventional methods with 3 columns pilots, the proposed scheme only requires 2 columns pilots for the channel estimation. In addition, the proposed approach is extended into multiple input multiple output- (MIMO-) based FBMC (MIMO-FBMC) systems.

The rest of this paper is organized as follows. The IAM method is briefly introduced in Section 2. The proposed approach is presented in Section 3, followed by the corresponding algorithm. Then, the proposed scheme is extended into MIMO-FBMC systems in Section 4. Section 5 shows the simulations, and Section 6 is the conclusions.

2. Channel Estimation with the IAM Method in FBMC/OQAM Systems

2.1. System Model.
As depicted in Figure 1, the diagram block of FBMC/OQAM transceiver is presented, in which M subcarriers are considered in the FBMC/OQAM system with the subcarrier spacing $1/M$. Note that the complex-valued data symbols have the interval of M samples in time, and by partitioning each complex-valued symbol, a pair of PAM symbols are obtained. The PAM symbols are denoted by $d_{m,n}$, with m the subchannel index and n the time index. In addition, $d_{m,2\bar{n}}$ and $d_{m,2\bar{n}+1}$, i.e., the real and imaginary parts of a QAM symbol, have the interval of $M/2$ samples in time. $g[l]$ is the prototype filter and spans over the time interval $0 \leq l \leq KM - 1$, where K is the overlapping factor and is supposed to be a positive integer. It is further assumed that the filter $g[l]$ is even and symmetric around its center, hence, $g[l] = g[KM - 1 - l]$ for $0 \leq l \leq KM - 1$.

Following Figure 1, the transmitted signal is [17]

$$s[l] = \sum_{m=0}^{M-1} \sum_{n\in\mathbb{Z}} d_{m,n} g\left[l - n\frac{M}{2}\right] e^{j2\pi ml/M} e^{j\pi(m+n)/2}, \quad (1)$$

where M stands for the subcarrier number. $d_{m,n}$ represents one transmitted symbol of position (m, n), with only real value.

Then, the signal at the receive antenna is obtained

$$r[k] = h[l] * s[l] + \eta[l], \quad (2)$$

where the sign $*$ is the convolution operator. $h[l]$ represents the multipath channel, and there exists a channel noise $\eta[l]$, satisfying the Gaussian distribution with variance σ^2 [18–20].

Then, demodulations at the receiver can be written as

$$\widehat{d}_{m,n} = \sum_{l=-\infty}^{\infty} r[l] g\left[l - n\frac{M}{2}\right] e^{-j2\pi ml/M} e^{-j\pi(m+n)/2}. \quad (3)$$

Then, an operator of taking real part is required.

$$\Re\left\{\widehat{d}_{m,n}\right\} = d_{m,n}. \quad (4)$$

Note that channel estimation is necessary in the FBMC/OQAM system under the multipath channel.

2.2. The IAM Method.
In [12], the IAM method with 3 column pilots has been presented for the channel estimation in FBMC/OQAM systems. The estimation model of IAM can be obtained [12]:

$$\widehat{d}_{m,n} \approx h_{m,n}\left(d_{m,n} + d_{m,n}^{(*)}\right) + \eta_{m,n}, \quad (5)$$

where $\widehat{d}_{m,n}$ is the demodulation at the receiver of FBMC/OQAM. $h_{m,n}$ stands for frequency-domain channel at m-th subcarrier, which is supposed quasi-invariant in the time domain in this paper, i.e.,
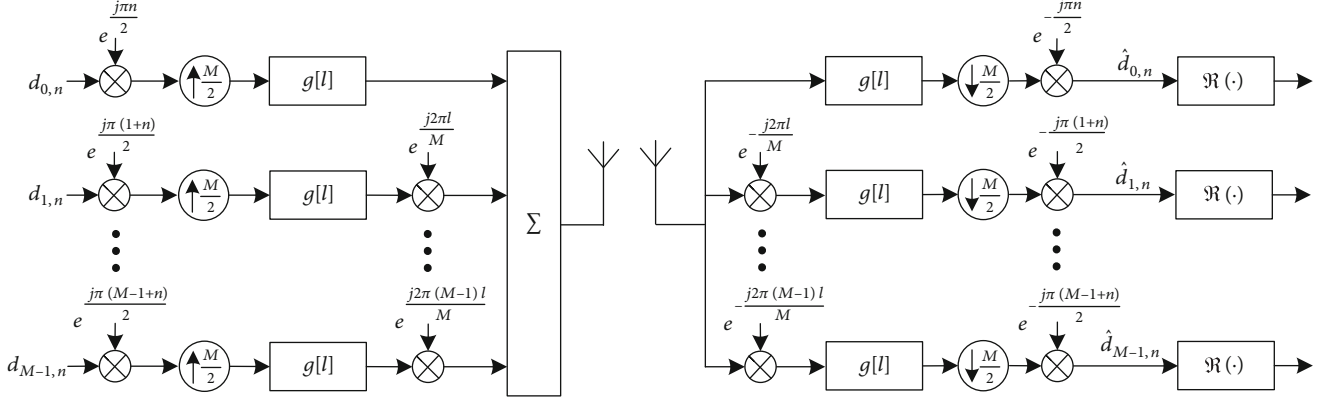
Figure 1: The FBMC/OQAM system diagram.
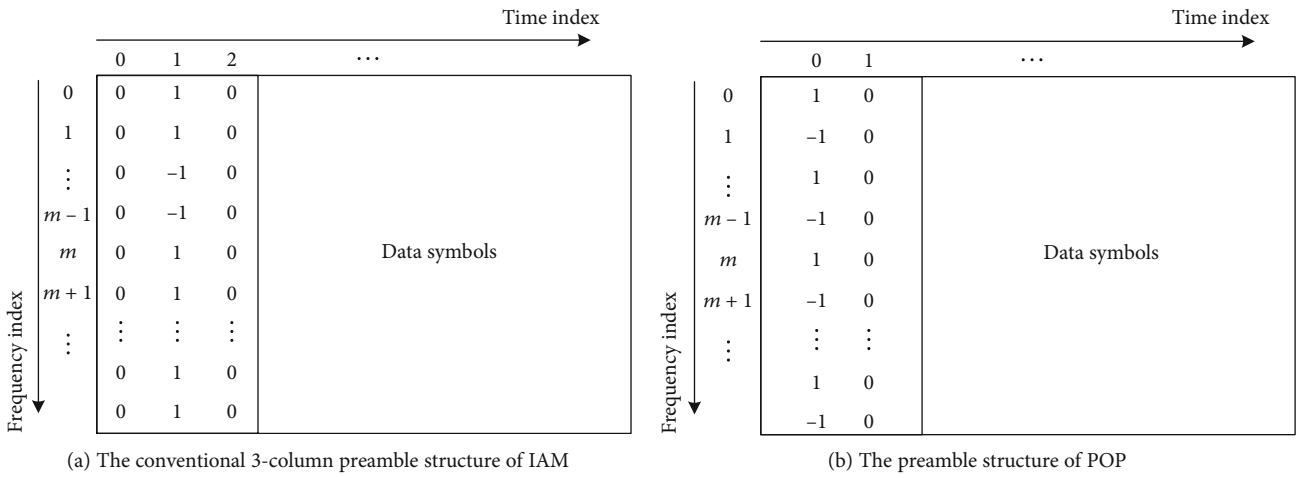
(a) The conventional 3-column preamble structure of IAM

(b) The preamble structure of POP

Figure 2: The conventional preamble and the proposed preamble in FBMC/OQAM.

$$h_{m,n} \approx h_{m,0} = \sum_{k=0}^{L-1} h[k] e^{-2j\pi mk/M}, \qquad (6)$$

with the maximum channel delay spread of L. The imaginary interference $d_{m,n}^{(*)}$ is written as

$$d_{m,n}^{(*)} = \sum_{\Omega} d_{m,n} \xi_{m+p,n+q}^{m,n}, \qquad (7)$$

with $\Omega = \{(p,q) \mid |p|, |q| \leq 1 \text{ and } (p,q) \neq (0,0)\}$, and the imaginary interference factor, $\xi_{m+p,n+q}^{m,n}$, is defined as

$$\xi_{m+p,n+q}^{m,n} = \xi_{p,q}^{0,0} = \sum_{l=-\infty}^{\infty} g[l] g\left[l + q\frac{M}{2}\right] e^{j2\pi pl/M} e^{j\pi(p+q)/2}. \qquad (8)$$

And the noise term $\eta_{m,n}$ is

$$\eta_{m,n} = \sum_{l=-\infty}^{\infty} \eta[l] g\left[l - n\frac{M}{2}\right] e^{-j2\pi ml/M} e^{-j\pi(m+n)/2}. \qquad (9)$$

Then, the IAM channel estimation is written as [12]

$$\widehat{h}_{m,n} = \frac{\widehat{d}_{m,n}}{d_{m,n} + d_{m,n}^{(*)}} = h_{m,n} + \frac{\eta_{m,n}}{d_{m,n} + d_{m,n}^{(*)}}. \qquad (10)$$

Figure 2(a) depicts the existing preamble of IAM, i.e., $d_{m,0} = d_{m,2} = 0$ with $m = 0, 1, \cdots, M-1$, $d_{4l,1} = d_{4l+1,1} = 1$, and $d_{4l+2,1} = d_{4l+3,1} = -1$ with $l = 0, 1, \cdots, M/4 - 1$. Note that it is well known that the interval of an FBMC/OQAM symbol is only half of that of an OFDM symbol. Therefore, the pilot overhead in FBMC/OQAM systems is 1.5 times of that classical OFDM systems.

2.3. POP Method. The POP method is another preamble-based channel estimation method in [16]. As shown in Figure 2(b), only two columns of real-valued pilots are required in the POP method, i.e., $d_{2k,0} = 1$, $d_{2k+1,1} = -1$ with $k = 0, 1, \cdots, M/2 - 1$, and $d_{m,1} = 0$ with $m = 0, 1, \cdots, M-1$.

Suppose $(m_1, n_1), (m_2, n_2)$ is the time-frequency positions of two symbols. The channel estimation by POP can be written as
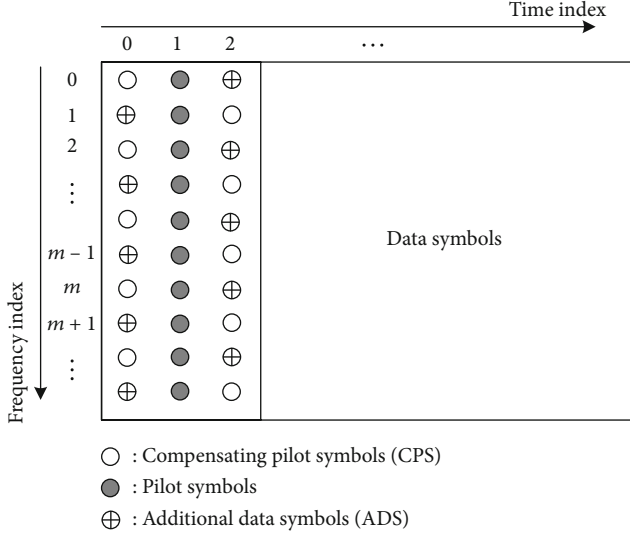
FIGURE 3: The proposed pilot structure in FBMC/OQAM.

$$
\begin{cases}
\Re\left(h_{m_1,n_1}\right) = \dfrac{\Re\left(\widehat{d}_{m_1,n_1}\right) + C \cdot \Im\left(\widehat{d}_{m_1,n_1}\right)}{d_{m_1,n_1}}, \\[3mm]
\Im\left(h_{m_1,n_1}\right) = C \cdot \Re\left(h_{m_1,n_1}\right),
\end{cases} \tag{11}
$$

where $\Re(\cdot)$ and $\Im(\cdot)$ are the operators of taking the real part and the imaginary part, respectively. C is the ratio between the imaginary part and the real part of h_{m_1,n_1},

$$
C = \frac{d_{m_2,n_2} \cdot \Re\left(\widehat{d}_{m_1,n_1}\right) - d_{m_1,n_1} \cdot \Re\left(\widehat{d}_{m_2,n_2}\right)}{d_{m_1,n_1} \cdot \Im\left(\widehat{d}_{m_2,n_2}\right) - d_{m_2,n_2} \cdot \Im\left(\widehat{d}_{m_1,n_1}\right)}. \tag{12}
$$

Then, the channel coefficients have been estimated. This approach does not require any knowledge of the prototype function and could be adopted as preamble-based method and also as a scattered-based channel estimation method.

3. Proposed Channel Estimation Scheme for FBMC/OQAM Systems

3.1. Preamble Structure of the Proposed Scheme. Figure 3 depicts the proposed preamble structure, where $d_{m,1}, m = 0, 1, \cdots, M-1$ are pilot symbols. Different from Figure 2(a), a part time-frequency resources of the first and the third columns are used to transmit additional data symbols (ADS), i.e., $\mathbf{u} = [u_0, x_1, \cdots, u_{M-1}]^T$ with $u_{2m} = a_{2m,2}$ and $u_{2m+1} = a_{2m+1,0}$. To eliminate the imaginary interference, compensating pilot symbols (CPS) are required, i.e., $\mathbf{v} = [v_0, v_1, \cdots, v_{M-1}]^T$ with $v_{2m} = a_{2m,0}$ and $v_{2m+1} = a_{2m+1,2}$. According to the criteria that the imaginary interferences from ADS and CPS should be mutually canceling, $\mathbf{v}$ can be designed by

$$
\mathbf{\Phi} H u + \mathbf{\Omega} H v = 0, \tag{13}
$$

where $\mathbf{H}$ represents the diagonal matrix with m-th diagonal element $h_{m,0}$, and

$$
\mathbf{\Phi} = \begin{pmatrix}
\xi_{0,2}^{0,1} & \xi_{1,0}^{0,1} & \xi_{2,2}^{0,1} & \cdots & \xi_{M-1,0}^{0,1} \\
\xi_{0,2}^{1,1} & \xi_{1,0}^{1,1} & \xi_{2,2}^{1,1} & \cdots & \xi_{M-1,0}^{1,1} \\
\xi_{0,2}^{2,1} & \xi_{1,0}^{2,1} & \xi_{2,2}^{2,1} & \cdots & \xi_{M-1,0}^{2,1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\xi_{0,2}^{M-1,1} & \xi_{1,0}^{M-1,1} & \xi_{2,2}^{M-1,1} & \cdots & \xi_{M-1,0}^{M-1,1}
\end{pmatrix}, \tag{14}
$$

$$
\mathbf{\Omega} = \begin{pmatrix}
\xi_{0,0}^{0,1} & \xi_{1,2}^{0,1} & \xi_{0,2}^{0,1} & \cdots & \xi_{M-1,2}^{0,1} \\
\xi_{0,0}^{1,1} & \xi_{1,2}^{1,1} & \xi_{2,0}^{1,1} & \cdots & \xi_{M-1,2}^{1,1} \\
\xi_{0,0}^{2,1} & \xi_{1,2}^{2,1} & \xi_{2,0}^{2,1} & \cdots & \xi_{M-1,2}^{2,1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\xi_{0,0}^{M-1,1} & \xi_{1,2}^{M-1,1} & \xi_{2,0}^{M-1,1} & \cdots & \xi_{M-1,2}^{M-1,1}
\end{pmatrix}. \tag{15}
$$

Then, it is obtained as

$$
v = -H^{-1}\mathbf{\Omega}^{-1}\mathbf{\Phi} H u. \tag{16}
$$

It should be noted that (16) cannot be used to design CPS directly due to the fact that it is difficult to obtain H at the transmitter.

It is proven that $\mathbf{\Omega}^{-1}\mathbf{\Phi}$ is a unitary matrix in the appendix. Let α_{mn} be the (m, n)-th entry of $\mathbf{\Omega}^{-1}\mathbf{\Phi}$, and it can be obtained as

$$
\begin{cases}
\sum_{n=0}^{M-1} |\alpha_{mn}|^2 = 1, \\[3mm]
\sum_{n} |\alpha_{mn}|^2 \approx 0.9991, |m-n| \leq 3, |m-n| \geq M-3.
\end{cases} \tag{17}
$$

Thus, it can be concluded that $\alpha_{mn} \approx 0$ for $M - 3 > |m - n| > 3$.

In addition, define $C = H^{-1}\mathbf{\Omega}^{-1}\mathbf{\Phi} H$, and its (m, n)-th element is $C_{mn} = (h_{m,0}/h_{n,0})\beta_{mn}$, which is close to zero for $M - 3 > |m - n| > 3$. Then, it can be assumed that $H_{m,0}/H_{n,0} \approx 1$ for $|m - n| \leq 3$ or $|m - n| \geq M - 3$. Therefore, $C_{mn} \approx \beta_{mn}$ and v can be obtained by

$$
v = -\mathbf{\Omega}^{-1}\mathbf{\Phi} u. \tag{18}
$$

3.2. Data Recovery of ADS. In this subsection, it is proven that the CPS is helpful for data recovery of ADS. Therefore, the ADS has the similar ability to fight against the noise compared with data symbols as we can see below.
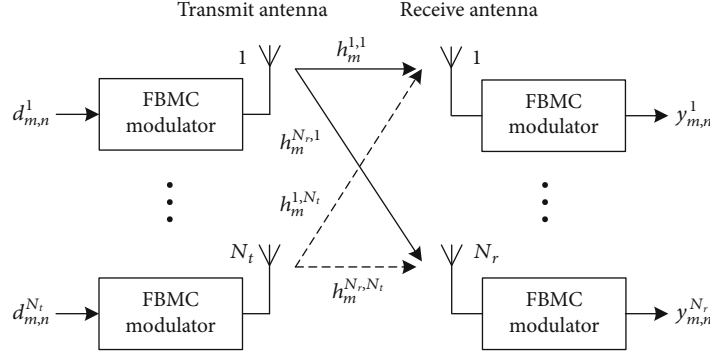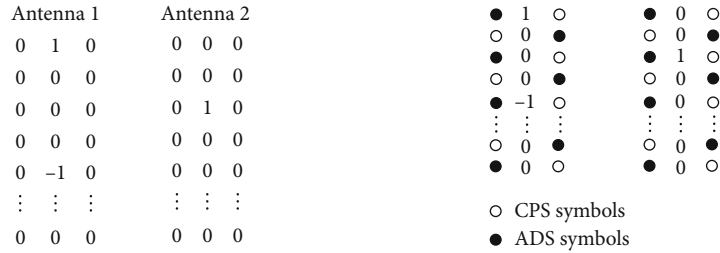
FIGURE 4: The MIMO-FBMC system model.

(a) Conventional preamble structure in MIMO-FBMC (b) Proposed preamble structure in MIMO-FBMC

FIGURE 5: Proposed preamble in MIMO-FBMC systems.

According to (5), the demodulation of v and u is

$$
\begin{cases}
u = H\left(u + u^{(*)}\right) + \boldsymbol{\eta}^u, \\
v = H\left(v + v^{(*)}\right) + \boldsymbol{\eta}^v,
\end{cases}
\tag{19}
$$

where $u^{(*)}$ and $v^{(*)}$ are the imaginary interference, respectively, i.e., $\mathbf{u}^{(*)} = [u_0^{(*)}, u_1^{(*)}, \cdots, u_{M-1}^{(*)}]^T$, in which $u_{2k}^{(*)} = d_{2k,2}^{(*)}$ and $u_{2k+1}^{(*)} = d_{2k+1,0}^{(*)}$, $\mathbf{v}^{(*)} = [v_0^{(*)}, v_1^{(*)}, \cdots, v_{M-1}^{(*)}]^T$, in which $v_{2k}^{(*)} = d_{2k,0}^{(*)}$ and $v_{2k+1}^{(*)} = d_{2k+1,2}^{(*)}$. $\boldsymbol{\eta}^u = [\eta_0^u, \eta_1^u, \eta_3^u, \cdots, \eta_{M-1}^u]^T$ with $\eta_{2i}^u = \eta_{2i,2}$ and $\eta_{2i+1}^u = \eta_{2i+1,0}$. $\boldsymbol{\eta}^v = [\eta_0^v, \eta_1^v, \eta_2^v, \cdots, \eta_{M-1}^v]^T$ with $\eta_{2i}^v = \eta_{2i,0}$ and $\eta_{2i+1}^v = \eta_{2i+1,1}$.

According to (18), we have

$$
u = \left(-\boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}\right)^{-1} v.
\tag{20}
$$

Let $u = \left(-\boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}\right)^{-1} v$, and it is obtained

$$
\begin{aligned}
u &= \left(-\boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}\right)^{-1} H\left(v + v^{(*)}\right) + \left(-\boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\eta}^v \\
&= \left(-\boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}\right)^{-1} H\left(v + v^{(*)}\right) + \tilde{\boldsymbol{\eta}}.
\end{aligned}
\tag{21}
$$

When the channel noise vanishes, we have $\Re(H^{-1}u) \approx \left(-\boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}\right)^{-1} v = u = \Re(H^{-1}u)$. Then, $1/2u + 1/2u$ will be the input of channel equalizer instead of u, since the noise variance will be reduced half after the linear combination. By this

way, the ADS has similar capability to fight against the noise compared with the data symbols.

4. Channel Estimation in MIMO-FBMC

In this section, the proposed scheme in Section 3 can be easily extended into MIMO-FBMC systems with N_t transmit antennas and N_r receive antennas, as shown in Figure 4. Denote the symbol of the t-th transmit antenna as $d_{m,n}^t$, and let $h_m^{r,t}$ be the channel frequency response between the r-th receive antenna and the t-th transmit antenna at the m-th subcarrier.

Without loss of generality, N_t is set to 2 in this section for simplicity. As is well known, the imaginary interference exists among the FBMC/OQAM symbols [12]. When there exists the imaginary interference between different antennas, it is difficult for one user to perform the channel estimation since the imaginary interference from other users is not available. Therefore, the key of the preamble design in MIMO-FBMC systems is the imaginary interference cancelation between antennas. Figure 5(a) depicts the conventional preambles on the two transmit antennas in MIMO-FBMC systems. Zeros are placed in the first and third columns to avert imaginary interference from data. It should be noted that nonzero pilots only locate in a part of subcarriers to ensure no imaginary interference between antennas, i.e., d_{4p}, $p = 0, 1, \cdots, M/4 - 1$ is a nonzero pilot for the first transmit antenna, and d_{4p+2}, $p = 0, 1, \cdots, M/4 - 1$ is a nonzero pilot for the second transmit antenna. In our proposed preambles, in Figure 5(b), the nonzero pilots in the second column are
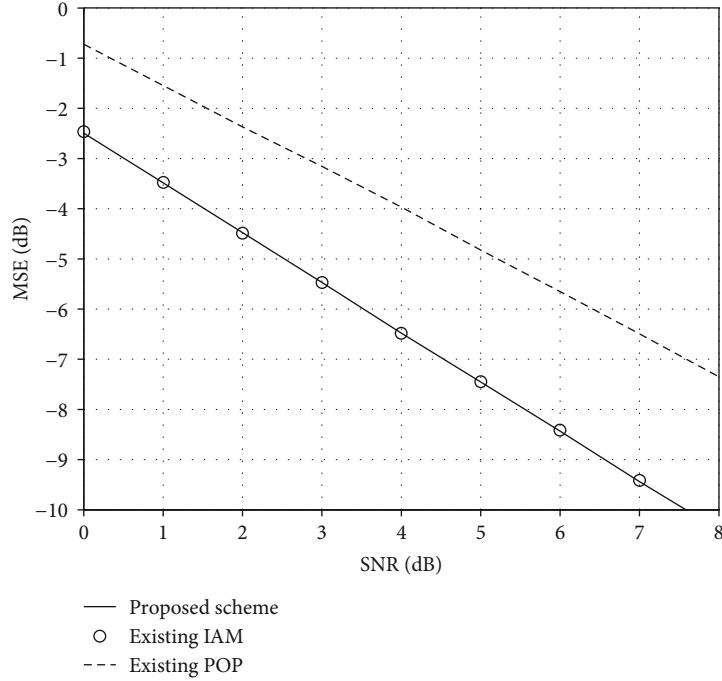
FIGURE 6: MSE of the proposed scheme.

the same as the conventional preambles. Differently, half of subcarriers in the first and third columns are placed by data symbols, improving the spectral efficiency. To eliminate the imaginary interference, compensating pilot symbols are required and designed according to (16) in Section 3. Therefore, compared with the conventional preambles, the pilot overhead of the proposed preamble is reduced by 1/3.

At the receiver, the demodulation of the r-th receive antenna is obtained

$$y^r_{m_0,n_0} = \sum_{t=1}^{N_t} \left[h^{r,t}_{m_0} d^t_{m_0,n_0} + \underbrace{\sum h^{r,t}_m d^t_{m,n} \xi^{m_0,n_0}_{m,n}}_{I^t_{m_0,n_0}} \right] + \eta^r_{m_0,n_0}, \quad (22)$$

where noise $\eta^r_{m_0,n_0}$ satisfies the Gaussian distribution with mean 0 and variance σ^2. $I^t_{m_0,n_0}$ is the imaginary interference term to the transmit symbol $d^t_{m_0,n_0}$.

As mentioned above, the imaginary interference between antennas can be removed completely. Then, for the proposed preambles in Figure 5(b), equation (22) can be rewritten as

$$y^r_{4p,1} = h^{r,1}_{4p} d^1_{4p,1} + \eta^r_{4p,1},$$
$$y^r_{4p+2,1} = h^{r,2}_{4p+2} d^2_{4p+2,1} + \eta^r_{4p+2,1}, \quad (23)$$

$$p = 0, 1, \cdots, M/4 - 1. \quad (24)$$

Accordingly, the channel estimation in MIMO-FBMC systems can be obtained as

$$h^{r,1}_{4p} = \frac{y^r_{4p,1}}{d^1_{4p,1}}, \quad (25)$$

$$h^{r,2}_{4p+2} = \frac{y^r_{4p+2,1}}{d^2_{4p+2,1}}, p = 0, 1, \cdots, M/4 - 1. \quad (26)$$

Then, the simple linear interpolation is performed on $h^{r,1}_{4p}$ and $h^{r,2}_{4p+2}$ to obtain the channel estimation of all subcarriers, respectively. It should be noted that our proposed approach could decrease the overhead of pilots significantly without the cost of additional pilot energy. Although the CPS symbols consume energy, it will be completely used for symbol recovery as presented in Subsection 3.2.

5. Simulation Results

In this section, we evaluate the performance of the proposed channel estimation approaches that are presented in this paper through computer simulations. The FBMC/OQAM system employs the PHYDYAS filter [17] and the overlap parameter $K = 4$. The multipath channel model is simulated, i.e., SUI proposed by the IEEE 802.16 broadband wireless access working group [21]. In simulations, the following parameters are considered.

    (i) Subcarrier number: 2048

    (ii) Sampling rate (MHz): 30.72

    (iii) Path number: 3

    (iv) Delay of path (μs): 0, 0,4, 0.9
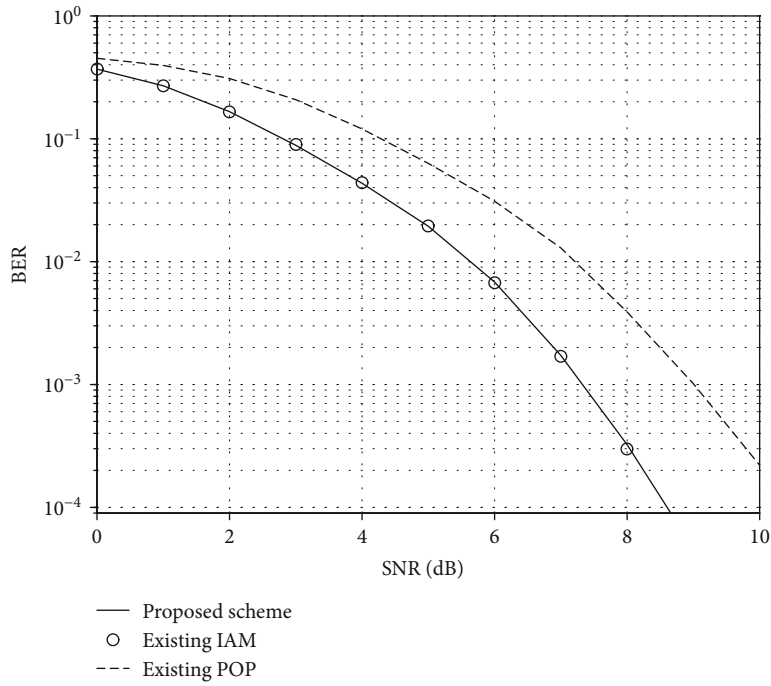
FIGURE 7: BER of the proposed scheme.
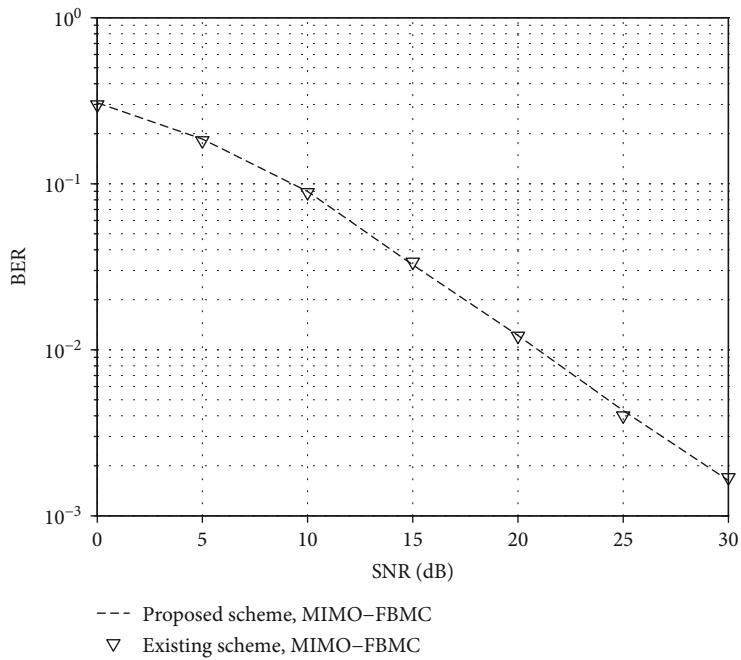
FIGURE 8: BER of the proposed scheme in MIMO-FBMC with 2 transmit antennas and 2 receive antennas.

(v) Power delay profile (dB): 0, -5, -10

(vi) Modulation: 4QAM

(vii) Channel coding: convolutional coding [12]

For comparison, we also give the performance of POP [16], which only requires two columns of pilots.

Figures 6 and 7 show the MSE and BER performances of the proposed approach. It can be easily seen that obvious performance loss is observed in the conventional POP method. As presented in [16], the performance of POP depends on a random power that could be close to zero sometimes, leading to a poor performance. In addition, the proposed scheme can achieve the same MSE and BER

performances as the IAM method, which indicates that the imaginary interference between pilots and data symbols can be eliminated completely in the proposed scheme. It should be noted that only two columns pilots are needed in the proposed scheme, while the conventional preamble requires three columns of pilots. Therefore, better spectral efficiency can be achieved by our proposed channel estimation approach.

Figure 8 shows BER of the proposed approach in MIMO-FBMC systems, in which both of N_t and N_r are set to 2. The proposed 2-column preamble can achieve similar BER compared to the conventional 3-column preamble, which demonstrates the effectiveness of the proposed scheme. It should be noted that our proposed scheme can reduce the pilot overhead significantly without the cost of additional pilot energy. Although the CPS symbols consume energy as shown in Figures 3 and 5, it will be completely used for symbol recovery as presented in Subsection 3.2.

6. Conclusions

In this paper, an approach was presented for the pilot overhead reduction in the channel estimation of FBMC/OQAM systems. Compared with the conventional methods with 3 columns pilots, the proposed scheme only requires 2 column pilots. In addition, the proposed approach was also extended into MIMO-FBMC systems. It was also proven that our proposed approach could decrease the overhead of pilots significantly without the cost of additional pilot energy and performance loss. Simulations have been done to verify the effectiveness of the proposed approach.

Appendix

From (8), we have $\zeta_{0,1}^{0,0} = -\zeta_{0,-1}^{0,0} = g$ and $\zeta_{1,1}^{0,0} = \zeta_{1,-1}^{0,0} = \zeta_{-1,1}^{0,0} = \zeta_{-1,-1}^{0,0} = \zeta_{M-1,-1}^{0,0} = \zeta_{-M+1,1}^{0,0} = f$ where g, f are imaginary-valued and constant. Furthermore, when $M - 1 > |p| > 1$ or $|q| > 1$, $\zeta_{m+p,n+q}^{m,n} \approx 0$ [12]. Thus, (14) and (15) can be rewritten as

$$\Phi = \begin{pmatrix} g & f & 0 & 0 & 0 & \cdots & 0 & f \\ f & -g & f & 0 & 0 & \cdots & 0 & 0 \\ 0 & f & g & f & 0 & \cdots & 0 & 0 \\ 0 & 0 & f & -g & f & \cdots & 0 & 0 \\ 0 & 0 & 0 & f & g & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & g & f \\ f & 0 & 0 & 0 & 0 & \cdots & f & -g \end{pmatrix}, \quad (A.1)$$

$$\Omega = \begin{pmatrix} -g & f & 0 & 0 & 0 & \cdots & 0 & f \\ f & g & f & 0 & 0 & \cdots & 0 & 0 \\ 0 & f & -g & f & 0 & \cdots & 0 & 0 \\ 0 & 0 & f & g & f & \cdots & 0 & 0 \\ 0 & 0 & 0 & f & -g & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -g & f \\ f & 0 & 0 & 0 & 0 & \cdots & f & g \end{pmatrix}, \quad (A.2)$$

respectively.

In addition, let

$$G = \begin{pmatrix} g & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -g & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & g & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -g & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & g & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & g & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -g \end{pmatrix}, \quad (A.3)$$

denote an $M \times M$ diagonal matrix and $F = \Phi - G$. Therefore, we have

$$\begin{cases} \Phi & = F + G, \\ \Omega & = F - G. \end{cases} \quad (A.4)$$

Note that $G^{\dagger} = -G$ and $F^{\dagger} = -F$ since both of g and f are imaginary-valued, where $(\cdot)^{\dagger}$ is the Hermitian transpose operation. For the even subcarrier M, it can be easily obtained $FG + GF = 0$ where 0 is the zero matrix. Then, we have

$$FF + GG = FF + GG,$$
$$FF + GG + FG + GF = FF + GG - FG - GF,$$
$$(F + G)(F + G) = (F - G)(F - G), \quad (A.5)$$
$$(F + G)(F - G)^{-1} = (F + G)^{-1}(F - G),$$
$$\Phi \Omega^{-1} = \Phi^{-1} \Omega.$$

Based on the equations $(\Omega^{-1}\Phi)^{\dagger} = \Phi^{\dagger}(\Omega^{\dagger})^{-1} = \Phi\Omega^{-1}$ and $(\Omega^{-1}\Phi)^{-1} = \Phi^{-1}\Omega$, we have

$$(\Omega^{-1}\Phi)^{\dagger} = (\Omega^{-1}\Phi)^{-1}. \quad (A.6)$$

Then,

$$\left(\mathbf{\Omega}^{-1}\mathbf{\Phi}\right)^{\dagger}\left(\mathbf{\Omega}^{-1}\mathbf{\Phi}\right) = \left(\mathbf{\Omega}^{-1}\mathbf{\Phi}\right)\left(\mathbf{\Omega}^{-1}\mathbf{\Phi}\right)^{\dagger} = I. \qquad \text{(A.7)}$$

Finally, we can obtain that $\mathbf{\Omega}^{-1}\mathbf{\Phi}$ is a unitary matrix.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] D. Kong, D. Qu, and T. Jiang, "Time domain channel estimation for OQAM-OFDM systems: algorithms and performance bounds," *IEEE Transactions on Signal Processing*, vol. 68, no. 2, pp. 322–330, 2014.

[2] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1170–1183, 2002.

[3] R. Nissel and M. Rupp, "Pruned DFT-spread FBMC: low PAPR, low latency, high spectral efficiency," *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4811–4825, 2018.

[4] R. Zakaria, D. Silva, and D. le Ruyet, "Lattice-reduction-aided equalization for MIMO-FBMC systems," *IEEE Communications Letters*, vol. 8, no. 1, pp. 101–104, 2019.

[5] H. Wang, "Low-Complexity MIMO-FBMC sparse channel parameter estimation for industrial big data communications," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3422–3430, 2021.

[6] D. Kong, J. Li, K. Luo, and T. Jiang, "Reducing pilot overhead: channel estimation with symbol repetition in MIMO-FBMC systems," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7634–7646, 2020.

[7] S. Hu, Z. Liu, Y. L. Guan, C. Jin, Y. Huang, and J. M. Wu, "Training sequence design for efficient channel estimation in MIMO-FBMC systems," *IEEE Access*, vol. 5, pp. 4747–4758, 2017.

[8] J. P. Javaudin, D. Lacroix, and A. Rouxel, "Pilot-aided channel estimation for OFDM/OQAM," in *The 57th IEEE Semiannual Vehicular Technology Conference, 2003. VTC 2003-Spring*, pp. 1581–1585, Jeju, Korea (South), April 2003.

[9] T. H. Stitz, T. Ihalainen, A. Viholainen, and M. Renfors, "Pilot-based synchronization and equalization in filter bank multicarrier communications," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, 2010.

[10] C. Lele, R. Legouable, and P. Siohan, "Channel estimation with scattered pilots in OFDM/OQAM," in *2008 IEEE 9th Workshop on Signal Processing Advances in Wireless Communications*, pp. 286–290, Recife, Brazil, July 2008.

[11] W. Cui, D. Qu, T. Jiang, and B. Farhang-Boroujeny, "Coded auxiliary pilots for channel estimation in FBMC-OQAM systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2936–2946, 2016.

[12] C. Lélé, J. P. Javaudin, R. Legouable, A. Skrzypczak, and P. Siohan, "Channel estimation methods for preamble-based OFDM/OQAM modulations," *European Transactions on Telecommunications*, vol. 19, no. 7, pp. 741–750, 2008.

[13] J. Du and S. Signell, "Novel preamble-based channel estimation for OFDM/OQAM systems," in *2009 IEEE International Conference on Communications*, Dresden, Germany, June 2009.

[14] C. Lélé, P. Siohan, and R. Legouable, "2 dB better than CP-OFDM with OFDM/OQAM for preamble-based channel estimation," in *2008 IEEE International Conference on Communications*, pp. 1302–1306, Beijing, China, May 2008.

[15] P. Singh, H. B. Mishra, A. K. Jagannatham, and K. Vasudevan, "Semi-blind, training, and data-aided channel estimation schemes for MIMO-FBMC-OQAM systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4668–4682, 2019.

[16] C. Lélé, P. Siohan, R. Legouable, and J.-P. Javaudin, "Preamble-based channel estimation techniques for OFDM/OQAM over the powerline," in *2007 IEEE International Symposium on Power Line Communications and Its Applications*, pp. 59–64, Pisa, Italy, March 2007.

[17] D. Kong, X. Zheng, Y. Zhang, and T. Jiang, "Frame repetition: a solution to imaginary interference cancellation in FBMC/O-QAM systems," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1259–1273, 2020.

[18] P. Liu, S. Jin, T. Jiang, Q. Zhang, and M. Matthaiou, "Pilot power allocation through user grouping in multi-cell massive MIMO systems," *IEEE Transactions on Communications*, vol. 65, no. 4, pp. 1561–1574, 2017.

[19] D. Kong, X.-G. Xia, P. Liu, and Q. Zhu, "MMSE channel estimation for two-port demodulation reference signals in new radio," *Science China Information Sciences*, vol. 64, no. 6, 2021.

[20] Y. Zhang, D. Wang, J. Wang, and X. You, "Channel estimation for massive MIMO-OFDM systems by tracking the joint angle-delay subspace," *IEEE Access*, vol. 4, pp. 10166–10179, 2016.

[21] D. Kong, Y. Xu, G. Song, J. Li, and T. Jiang, "A CP reduction scheme based on symbol repetition for narrow-band IoT systems," *IEEE Internet of Things Journal*, 2021.

WILEY | Hindawi

Research Article

High-Resolution ISAR Imaging Based on Improved Sparse Signal Recovery Algorithm

Junjie Feng [iD],[1] Yinan Sun,[2] and XiuXia Ji [iD][3]

[1]School of Physics and Electrical Engineering, Liupanshui Normal University, Liupanshui, Guizhou, China
[2]Wuhan University, School Electronic Information, Wuhan 430072, China
[3]Nanjing Vocational College of Information Technology, Nanjing 210023, China

Correspondence should be addressed to Junjie Feng; fhzy0929@163.com

In order to solve the problem of high-resolution ISAR imaging under the condition of finite pulses, an improved smoothed L0 norm (SL0) sparse signal reconstruction ISAR imaging algorithm is proposed. Firstly, the ISAR imaging is transformed into the optimization problem of minimum L0 norm. Secondly, a single-loop structure is used instead of two loop layers in SL0 algorithm which increases the searching density of variable parameter to ensure the recovery accuracy. Finally, the compared step is added to ensure the optimization solution along the steepest descent gradient direction. The experimental results show that the proposed algorithm has better imaging effect.

1. Introduction

Inverse synthetic-aperture radar (ISAR) has the characteristics of all-weather, all-time, long-range, and high-resolution which is widely used in military and civil fields, such as to detect, locate, track, and estimate the parameters of the target [1–4]. The range and azimuth resolution of ISAR imaging are related to radar system bandwidth and imaging accumulation angle, respectively. The range resolution is usually improved by increasing the bandwidth of the transmitted signal, and the azimuth resolution is improved by increasing the observation angle range. In a long observation time, the scattering characteristics of noncooperative moving targets change greatly, and there may be a large Doppler time-varying, which is not conducive to imaging and makes it difficult to improve the azimuth resolution. So implementing imaging in a short time duration is meaningful.

In recent years, compressive sensing (CS) has become very popular in signal processing [5–8]. It is a new signal processing theory proposed by Donoho which is developed from the theory of signal sparse decomposition and approximation. The theory shows that as long as the signal is compress-ible or sparse in a certain transform domain, the original signal can be reconstructed with a small number of sampled signals. It is a technique proposed to improve signal separation ability using a prior sparse property information of the signal.

In radar imaging, the targets are generally composed of strong scattering points, which are sparsely distributed in the imaging region. So radar signal processing based on compressive sensing theory has been widely studied [9–12], such as SAR imaging [13], ISAR imaging [14, 15], and MIMO radar imaging [16]. The CS theory is not sensitive to data loss, which can effectively improve the imaging problem of radar imaging system in the case of data loss and improve the signal processing ability. The Bayesian sparse signal reconstruction ISAR imaging algorithm based on hierarchical prior model correlation prior is proposed in [17], which can obtain better focusing effect in noisy environment. However, the algorithm is computational complexity. By using the sparsity of ISAR target in azimuth direction, the target image can be obtained by solving the L1 norm minimization in [18]. Combination of local sparsity constraint and nonlocal total variation is discussed in

[19]. A two-dimensional structure mode-coupled Bayesian (PC-SBL) ISAR imaging algorithm is proposed in [20], which can realize high-resolution ISAR imaging, but the algorithm has a large amount of computation, which is not conducive to real-time imaging.

Mohimani et al. proposed the smoothed L0 norm sparse signal recovery algorithm by using the continuous Gaussian function sequence with parameters as smoothing function to approximate the minimum L0 norm [21]. The steepest descent method and gradient projection principle by a double-layer loop is used to obtain the optimal solution. The outer layer establishes a sequence from large to small, and the inner layer uses the steepest descent method to obtain the approximate solution. The reconstruction speed of the algorithm is 2-3 times faster than basis pursuit algorithm when it has the same accuracy.

In order to achieve high-resolution ISAR imaging under the condition of finite pulses, an ISAR imaging algorithm with single-loop structure compared to SL0 sparse signal recovery algorithm is proposed. The imaging problem is transformed to a minimum L0 norm optimization problem to achieve high-resolution ISAR imaging. The steepest descent method should reduce the cost function at every step, but it is not really along the descending direction in the actual solving process. Therefore, in the above algorithm, the step of checking whether to descend is added in each iteration. If not, the midpoint of the previous point and the current point is taken for correction to ensure that the search direction is along the steepest descent direction. By using the improved algorithm, ISAR imaging is more intensive with limited pulse numbers. Real data ISAR images obtained using the proposed method is competitive to the several popular methods.

The paper is organized as follows. "ISAR Imaging Model" introduces the ISAR sparse imaging model. In "ISAR Imaging Algorithm," the proposed reconstruction algorithm is introduced in detail. Simulation and real data ISAR imaging results are presented in "Experimental Results." Finally, "Conclusion" provides the conclusion.

2. ISAR Imaging Model

Assuming that the target is located in far field, the radar transmits linear frequency modulization signal as

$$y(t) = \exp\left(j2\pi\left(f_0 t + \frac{1}{2}\mu_0 t^2\right)\right) t \in \left[-\frac{T}{2}, \frac{T}{2}\right], \quad (1)$$

where f_0 is the center frequency, μ_0 is the slope, and T is the pulse period. Then, the complex echo signal of scattering point $P(x, y)$ is

$$y(t) = \sigma e^{-j2\pi f_0 \tau(t) + j\pi\mu_0(t-\tau(t))^2}, \quad (2)$$

where $\tau(t)$ is time delay and σ is signal amplitude. After range compression, the signal can be expressed as

$$y(t) = \sigma \exp\left(-j2\pi f_0 \tau(t)\right) \sin c(\mu_0 T(t - \tau(t))). \quad (3)$$

After the envelope compensation, the target rotates around the coordinate axis at a constant angular velocity ω. Assuming the initial position of the scattering point is (x_0, y_0), the time delay is mainly determined by $x_0 \sin(\omega t) + y_0 \cos(\omega t)$. After the Taylor expansion, $x_0 \sin(\omega t) + y_0 \cos(\omega t)$ is expressed as

$$x_0 \sin(\omega t) + y_0 \cos(\omega t) \approx y_0 + x_0 \omega t. \quad (4)$$

Then, $\tau(t) \approx \tau_0 + (2(y_0 + x_0\omega t)/c)$, where c is light speed. The echo signal of range unite $\tau_0 + (2y_0/c)$ is

$$y(t) = \sigma e^{-\frac{j4\pi x_0 \omega t}{\lambda}}, \quad (5)$$

where $\lambda = c/f_0$ is the wavelength. After discretization, the above formula can be written as

$$y(n) = \sigma\alpha(n), \quad (6)$$

where $\alpha(n) = e^{j2\pi f_d n}$, $f_d = -2x_0\omega d/\lambda$, and d_t is the sampling interval. The received signal is written as a matrix; then

$$y = \Phi x + n, \quad (7)$$

where x is the vector composed of scattering coefficient σ and Φ is sparse matrix.

To solve x, the following sparse optimization strategy is used:

$$\min_x \|x\|_0 \\ \text{s.t.} \|y - \Phi x\|_2 < \tau, \quad (8)$$

where τ is a small constant which is related to the noise variance. The ISAR imaging problem is transformed into the reconstruction problem of formula (8).

3. ISAR Imaging Algorithm

To solve x, the smoothed L0 norm algorithm is proposed in [21]. The Gauss sequence $F_\sigma(x) = \sum_i \exp\left(-x_i^2/\sigma\right)$ is used to approach the L0 norm. The minimum L0 norm can be obtained by gradually reducing the value of control parameter σ. A two-layer method was proposed to solve the sparse signal recovery problem in the algorithm. In order to further improve the reconstruction effect of sparse signal, the one-order negative exponential function sequence $G_\sigma(x) = \sum_i \exp\left(-|x_i|/\sigma\right)$ is proposed as smoothed function to approach L0 norm in the paper. When $\sigma \longrightarrow \infty$, $G_\sigma(x)$ approximates L1 norm. When $\sigma \longrightarrow 0$, $G_\sigma(x)$ approximates L0 norm. Therefore, the sparse solution can be searched with high probability at the beginning of iteration.

In the double-loop structure of SL0 algorithm, the solution of $\sigma = \sigma_j$ is only as the initial value of $\sigma = \sigma_{j+1}$, so it is not necessary to solve the exact solution of $\sigma = \sigma_j$ in the inner loop. Therefore, the sparse signal reconstruction algorithm in this paper cancels the inner loop step and
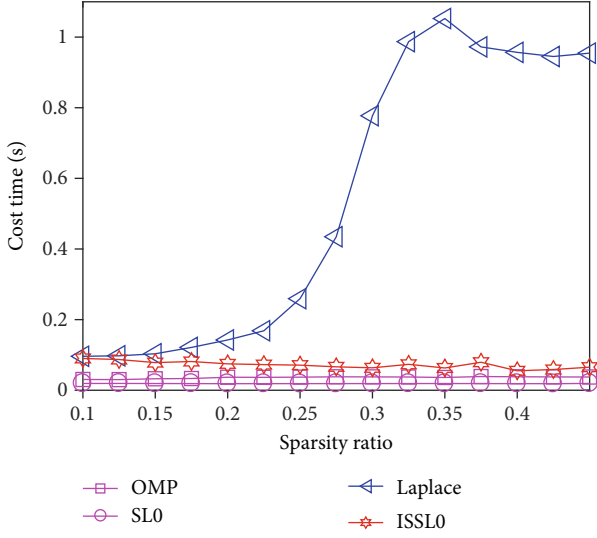
FIGURE 1: Computation costs of different algorithms.

FIGURE 2: Correct position estimation of different algorithms.

adopts the single loop instead of the double-loop structure in SL0 algorithm [52], the interval between control parameters σ_j and σ_{j+1} is reduced, only one gradient descent method is used to solve the minimum value of the cost function, and then the solution of x is given as the initial value of $\sigma = \sigma_{j+1}$. Due to the solution obtained by the steepest descent method is not necessarily in the feasible solution region, it is necessary to project the solution obtained by each iteration into the feasible solution region using the projection method. The proposed algorithm is called the improved single-loop smoothed L0 norm (ISSL0) algorithm. The ISSL0 algorithm can be expressed as follows:

Initialization is as follows:

(1) Let $\hat{x}_0$ be equal to the minimum L2 norm solution of $y = \Phi x$, obtained by $\hat{x}_0 = \Phi^H (\Phi \Phi^H)^{-1} y$

(2) Choose a suitable decreasing sequence for $\{\sigma\}$, $[\sigma_1, \sigma_2, \cdots, \sigma_J]$

for $j = 1, \cdots, L$

(i) Let $\sigma = \sigma_j$ and $\beta = L - l + 1/L$

(ii) Minimize the function $G_\sigma(x)$ on the feasible set $x = \{x : |\Phi x - y| < \varepsilon\}$

(iii) Let δ be the gradient of $G_\sigma(x)$

(iv) For x, let $x_1 \longleftarrow x - \mu\sigma\delta$

(v) If $|\Phi x_1 - y| > \varepsilon$, project x back into the feasible set:

$$x_1 \longleftarrow x_1 + \Phi^H (\Phi \Phi^H)^{-1} (y - \Phi x_1) \qquad (9)$$

(vi) Compare step

FIGURE 3: MSE of different algorithms.

While $G_\sigma(x_1) > G_\sigma(x)$

$$x_2 = (x + x_1)/2 \qquad (10)$$

update $G_\sigma(x_2)$
If $G_\sigma(x_2) < G_\sigma(x)$

$$x = x_2 \qquad (11)$$

(vii) $\hat{x}_j = x$.

(3) Final solution is $x = \hat{x}_J$.

FIGURE 4: ISAR images using 16 pulses. (a) OMP. (b) Laplace. (c) SL0. (d) ISSL0.

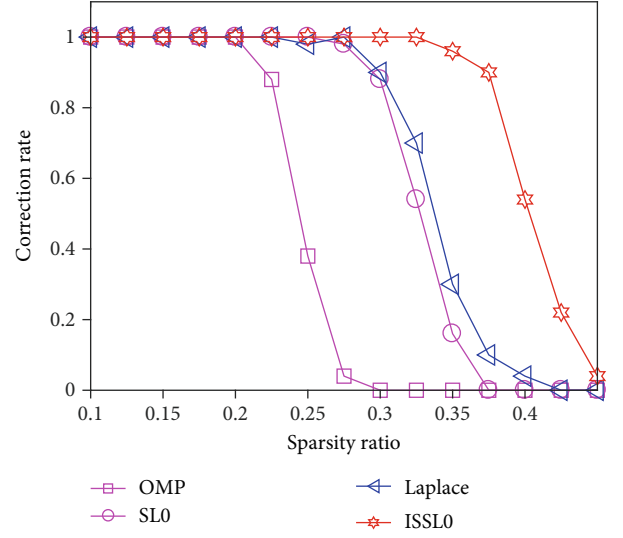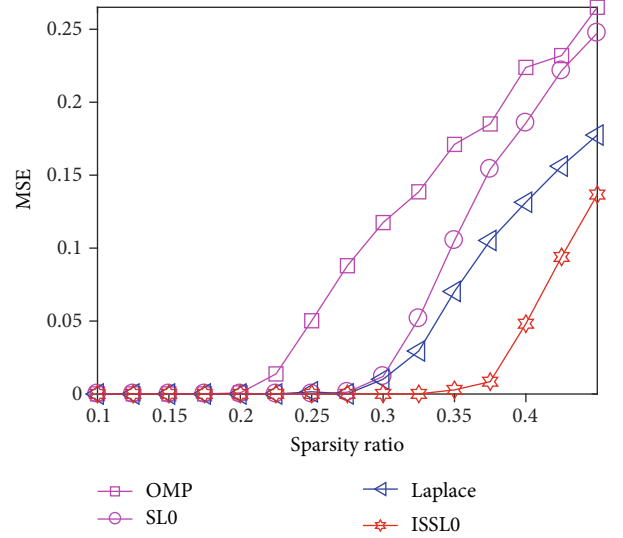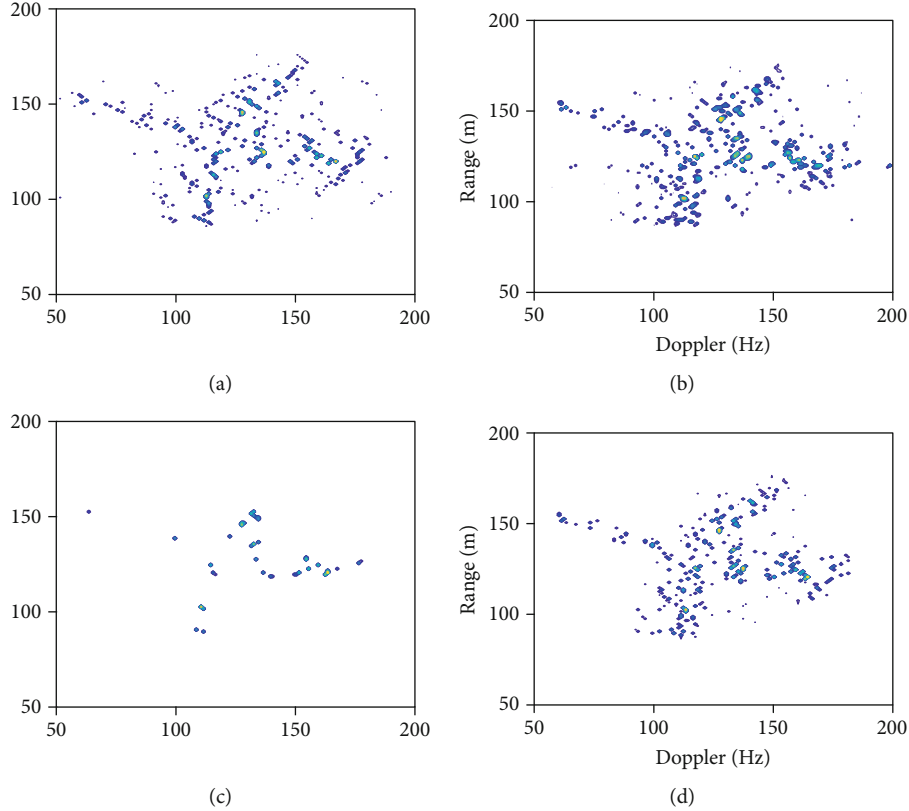The steepest descent method should reduce the cost function at every step, but it is not the descending direction in the actual solution process. Therefore, in the above algorithm, the step of checking whether to descend is added in the process of each iteration. If not, the midpoint of the previous point and the current point are taken for correction to ensure that the search direction is along the steepest descent direction. By using the single loop and compared step, the algorithm ensures the reconstruction accuracy and the computation amount cannot increase. The choice of step size factor in the steepest descent method is important. For a large step size, it may not converge, but for a very small step size, the computation efficiency is low. The step size should decrease when the searching point approaches the minimum solution. For search step factor, because the search point is far away from the minimum value point at the initial value, a larger step size is selected. When the search point is gradually close to the minimum value, the search step size should be gradually reduced. Regulatory factors are used to adjust the step; let $\mu = \beta(\max(|x|)/L)$.

4. Experimental Results

4.1. Simulation 1: One-Dimensional Sparse Signal Recovery. In order to verify the reconstruction performance of the algorithm, the paper tests the algorithm through the MATLAB processing platform. The signal model with noise is $y = \Phi x + n$, where Φ is constructed by selecting

its components from $N(0, 1)$ and x is the sparse signal, the nonzero elements of sparse signal consist of random signal ± 1. For the SL0 method, the numbers of outer loop and inner loop are 20 and 10, respectively. For the ISSL0 algorithm, the loop number is 200. For the ISSL0 algorithm, the loop number is 200. The experiment was implemented 100 times. The MSE is defined as $(1/N)\|x - x\wedge\|^2$. Changing the sparsity ratio of the signal, the curve of reconstruction time, correction rate, and MSE for different algorithms of sparse signal are shown in Figures 1–3. As can be seen from Figure 1, the ISSL0 algorithm takes more time than the OMP [22] algorithm and SL0 algorithm and shorter than the Laplace [23] algorithm. From Figures 2 and 3, we can see that the performances of ISSL0 algorithm are competitive with other algorithms.

4.2. Simulation 2: ISAR Imaging Using Real Data. A set of real data of the Yak-42 plane is used to demonstrate the performance of the proposed ISAR imaging algorithm. The related parameter descriptions of the radar data are listed as follows: the carrier frequency is 10 GHz with signal bandwidth of 400 MHz, and the range resolution is 0.375 m. The center carrier frequency is 5.52 GHz, and the pulse repetition frequency is 50 Hz. 256 echo pulses are selected as experimental data. 16, 32, and 64 pulses are implemented. For the SL0 method, the numbers of outer loop and inner loop are 20 and 10, respectively. For the ISSL0 algorithm, the loop number is 200. The simulation results are compared visually and
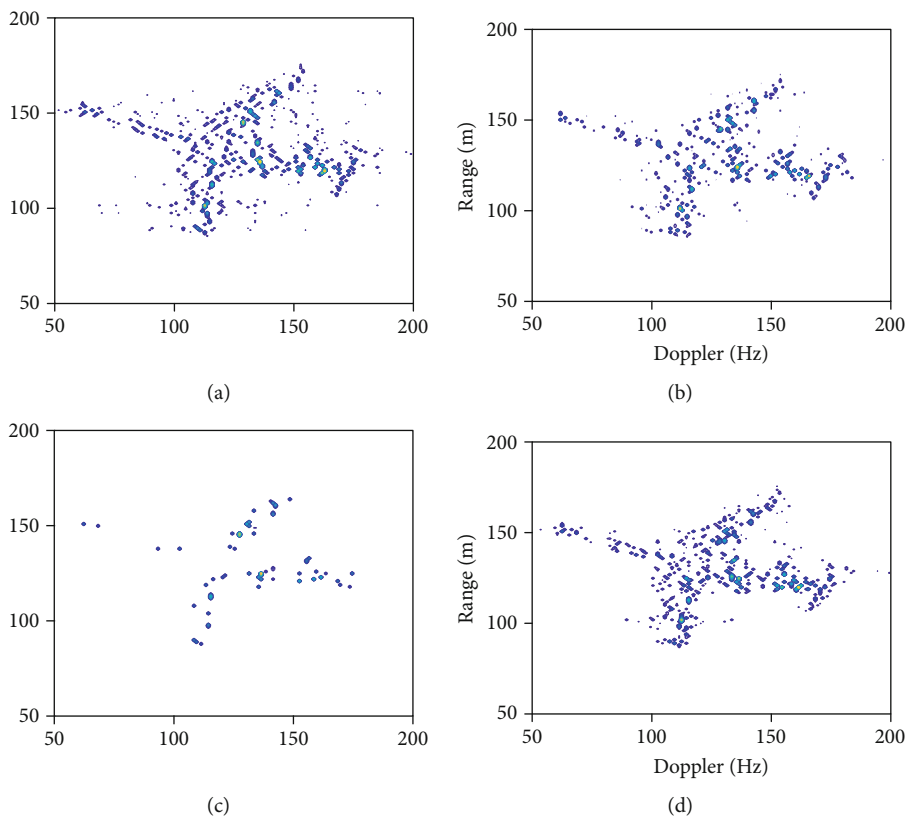
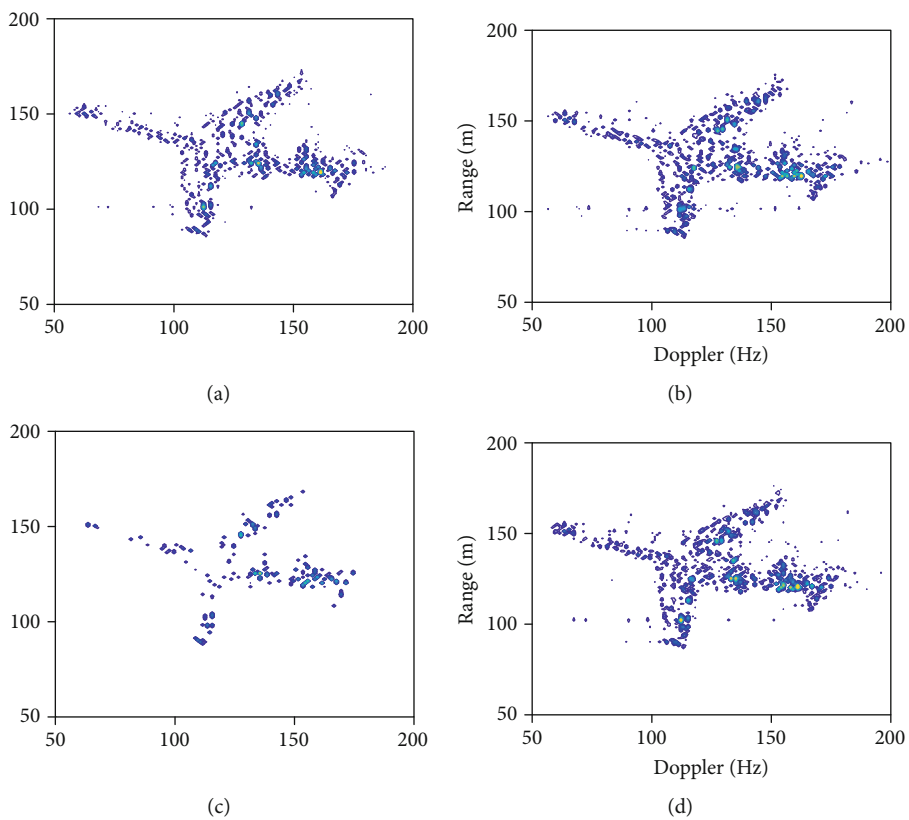Figure 5: ISAR images using 32 pulses. (a) OMP. (b) Laplace. (c) SL0. (d) ISSL0.

Figure 6: ISAR images using 64 pulses. (a) OMP. (b) Laplace. (c) SL0. (d) ISSL0.

quantitatively to those images obtained by OMP, Laplace, SL0, and ISSL0 algorithms in Figures 4–6. It is noticeable that more amounts of pulses generally lead to better imagery results. When the number of pulses is 64, the four sparse signal reconstruction algorithms are better. For the SL0 algorithm, when 16 and 32 pulses are used, some information of scattered points is missing. When the number of pulses is small, a large number of false scattering points appear in the OMP and Laplace algorithms. Regardless of the number of pulses, the ISSL0 algorithm keeps the aircraft in a good shape without false scattered points, and the geometric representation of the target is clear.

5. Conclusion

One improved sparse signal recovery ISAR imaging algorithm is proposed to ISAR imaging. A single-loop structure is proposed to solve the optimal solution. The revised step is added to ensure the searching direction is decrease compared to the traditional smoothed L0 norm recovery algorithm. The experiment results verify that the proposed algorithm can improve ISAR imaging quality.

Data Availability

All data sources are reliable and the corresponding author can be contacted.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

References

[1] D. H. Kim, D. K. Seo, and H. T. Kim, "Efficient classification of ISAR images," *IEEE Transaction on Antennas and Propagation*, vol. 53, no. 5, pp. 1611–1621, 2005.

[2] L. Zhang, M. Xing, C.-W. Qiu, J. Li, and Z. Bao, "Achieving higher resolution ISAR imaging with limited pulses via compressed sampling," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 3, pp. 567–571, 2009.

[3] H. Wang, Y. Quan, M. Xing, and S. Zhang, "ISAR imaging via sparse probing frequencies," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 451–455, 2011.

[4] Y. Chen, Q. Zhang, N. Yuan, Y. Luo, and H. Lou, "An adaptive ISAR-imaging-considered task scheduling algorithm for multi-function phased array radars," *IEEE Transactions on Image Processing*, vol. 63, no. 19, pp. 5096–5110, 2015.

[5] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.

[6] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[7] T. Wimalajeewa, H. Chen, and P. K. Varshney, "Performance limits of compressive sensing based signal classification," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2758–2770, 2012.

[8] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[9] L. Wan, K. Liu, Y.-C. Liang, and T. Zhu, "DOA and polarization estimation for non-circular signals in 3-D millimeter wave polarized massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 8, p. 1, 2021.

[10] L. Wan, L. Sun, K. Liu, X. Wang, Q. Lin, and T. Zhu, "Autonomous vehicle source enumeration exploiting non-cooperative UAV in software defined internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, pp. 1–13, 2020.

[11] L. Sun, L. T. Wan, and X. P. Wang, "Learning-based resource allocation strategy for industrial IoT in UAV-enabled MEC systems," *IEEE Transactions on Industrial Informatics*, vol. 9, p. 1, 2020.

[12] J. Lv, L. Huang, Y. Shi, and X. Fu, "Inverse synthetic aperture radar imaging via modified smoothed L_0 norm," *IEEE Antennas and Wireless Propagation Letters*, vol. 13, pp. 1235–1238, 2014.

[13] C. He, L. Liu, L. Xu, M. Liu, and M. Liao, "Learning based compressed sensing for SAR image super-resolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 4, pp. 1272–1281, 2012.

[14] J. Liu, X. Li, S. Xu, and Z. Zhuang, "ISAR imaging of non-uniform rotation targets with limited pulses via compressed sensing," *Progress In Electromagnetics Research B*, vol. 41, pp. 285–305, 2012.

[15] J. J. Feng, "ISAR imaging based on multiple measurement vector model sparse signal recovery algorithm," *Mathematical Problems in Engineering*, vol. 2020, Article ID 1743593, 8 pages, 2020.

[16] W. J. Zhang and A. Hoorfar, "A generalized approach for SAR and MIMO radar imaging of building interior targets with compressive sensing," *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 1052–1055, 2015.

[17] L. F. Zhao, L. Wang, G. Bi, and L. Yang, "An autofocus technique for high-resolution inverse synthetic aperture radar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6392–6403, 2014.

[18] X. Zhang, T. Bai, H. Meng, and J. Chen, "Compressive sensing-based ISAR imaging via the combination of the sparsity and nonlocal total variation," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 990–994, 2014.

[19] L. Zhang, Z.-J. Qiao, M.-D. Xing, J.-L. Sheng, R. Guo, and Z. Bao, "High-Resolution ISAR imaging by exploiting sparse apertures," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 997–1008, 2012.

[20] H. Duan, L. Zhang, J. Fang, L. Huang, and H. Li, "Pattern-coupled sparse Bayesian learning for inverse synthetic aperture radar imaging," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1995–1999, 2015.

[21] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on

smoothed l0 norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, 2009.

[22] J. F. Determe, J. Louveaux, L. Jacques, and F. Horlin, "Improving the correlation lower bound for simultaneous orthogonal matching pursuit," *IEEE Signal Processing Letters*, vol. 3, no. 11, pp. 1642–1646, 2016.

[23] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 53–63, 2010.

WILEY | Hindawi

Research Article

Generalized Sparse Polarization Array for DOA Estimation Using Compressive Measurements

Tao Chen [ID],[1] **Jian Yang [ID],**[1,2] **Weitong Wang [ID],**[1] **and Muran Guo [ID]**[1]

[1]*College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China*
[2]*Beijing Institute of Remote Sensing Equipment, Beijing 100854, China*

Correspondence should be addressed to Muran Guo; guomuran@hrbeu.edu.cn

The compressive array method, where a compression matrix is designed to reduce the dimension of the received signal vector, is an effective solution to obtain high estimation performance with low system complexity. While sparse arrays are often used to obtain higher degrees of freedom (DOFs), in this paper, an orthogonal dipole sparse array structure exploiting compressive measurements is proposed to estimate the direction of arrival (DOA) and polarization signal parameters jointly. Based on the proposed structure, we also propose an estimation algorithm using the compressed sensing (CS) method, where the DOAs are accurately estimated by the CS algorithm and the polarization parameters are obtained via the least-square method exploiting the previously estimated DOAs. Furthermore, the performance of the estimation of DOA and polarization parameters is explicitly discussed through the Cramér-Rao bound (CRB). The CRB expression for elevation angle and auxiliary polarization angle is derived to reveal the limit of estimation performance mathematically. The difference between the results given in this paper and the CRB results of other polarized reception structures is mainly due to the use of the compression matrix. Simulation results verify that, compared with the uncompressed structure, the proposed structure can achieve higher estimated performance with a given number of channels.

1. Introduction

In a traditional scalar sensor array, the time delay of the phased array is used to estimate the direction of arrival (DOA). In the practical application environment, however, the signal to be detected usually has certain polarization characteristics. The polarization sensitive array [1–3] can be used to obtain and utilize the spatial and polarization domain information of the signal source comprehensively, which lays a physical foundation for improving the overall performance of the array signal processing. Therefore, the concept of polarization is extended to wireless communication [4], radar systems [5], and many other fields of space science [6]. At present, the research on polarization parameter estimation mainly focuses on how to improve the estimation performance, such as estimation accuracy and degree of freedom (DOF). As the polarized state of a signal varies with polarization diversity, only if the polarization direction of a single antenna matches with the incoming wave, all energy of the incoming wave can be received; otherwise, the loss of

energy will occur [7]. Since a dual-polarized antenna [8, 9] can receive the signal energy along the horizontal and vertical branches, it can increase the receiving efficiency of signals.

To obtain a higher number of DOFs under the premise of a given number of sensors, sparse array structures [10–12] have been proposed under the coarray framework, which generate equivalent virtual array elements via extracting the correlation information of received signals. On the one hand, the nested array [11] can generate a difference coarray with all continuous lags, which is very useful for spatial smoothing-based estimation algorithms. However, since one subarray has the sensors placed with a half wavelength, the mutual coupling effects will compromise the estimation performance. In [13], a sparse nested array has been proposed, where the interelement spacing of the dense subarray is extended to suppress mutual coupling effects. The diversely polarized dipoles are used to further improve estimation performance. Moreover, the spatially spread orthogonal dipoles are exploited in the sparse nested array in [14], where a passive direction finding structure with high accuracy has been

proposed. Then, [15] further takes spatially spread square acoustic vector sensors into account, thus constructing a high-accuracy DOA estimation structure for an underdetermined case. On the other hand, the coprime array [12] can suppress the mutual coupling effect by using two sparse uniform linear arrays (ULA). In general, for a coprime array composed of an M-element subarray and an N-element subarray, up to $2M + N - 2$ uncorrelated sources can be distinguished. Therefore, the coprime array and its related improved structures [16] are also introduced into other radar structures [17]. In addition, [18] proposed an adaptive beamforming approach based on the coprime array, where the output performance is improved.

The DOA parameters, including the azimuth and elevation angle, are of great importance in many applications, especially the unmanned driving technology [19] which is one of the current hot research issues. Thus, different catalogs of DOA estimation methods have been proposed, for instance, the subspace methods [7], deep learning methods [20], and sparse reconstruction methods [21]. The compressive sensing (CS) theory is the kernel of sparse reconstruction methods. However, CS is first used in time domain to break through the limitation of the Nyquist Sampling Theorem [22], since a high sampling rate is usually required for wideband signals, thus leading to a high cost of analog-to-digital converters (ADCs). To be specific, if the number of nonzero elements in a vector is much less than that of the zero elements, then this vector is regarded as a sparse vector and can be recovered from a small sample set. Thus, the sampling rate is dramatically reduced. Then, as the polarization sensitive array develops, several DOA and polarization joint estimation algorithms have been proposed [23–25]. Moreover, in addition to the estimation algorithms, the CS has also been used in system design. The DOA estimation system exploiting compressive measurements can dramatically reduce the complexity and the computational burden. For example, the compressive measurement method has found many applications in the coprime arrays [26], one-bit quantization [27], modulated wideband converter [28], and MIMO radar [29].

Motivated by the above facts, in this paper, we mainly consider the orthogonal dipole antennas and propose a compressive measurement-based orthogonal dipole sparse array structure for the joint estimation of signal parameters. It is worth noting that there is no demand on the receive array structure, meaning that all kinds of dipole sparse array can be used. Based on the proposed structure, we first estimate the DOAs using a CS-based approach, and then the estimated DOAs are utilized to analyze the polarization parameters by a least-square estimation approach. In [24], the CS-based joint estimation of DOA and polarization parameters is proposed using a sparse array consisting of dual-polarized antenna elements. However, we would like to emphasize the contribution of this paper, as well as the difference with [24]:

(a) In this paper, a two-dimensional polarization signal model is established, in which the azimuth angle and elevation angle are both taken into account,

while [24] only considers the one-dimensional case, that is, only the azimuth angle is included in the signal model. In addition, it should be noted that a steering vector matrix $\psi(\theta_k, \phi_k, \gamma_k, \eta_k)$ is introduced into the signal model to describe the coherent structure in the polarizational and spatial domains

(b) The compression measurement method is applied to the proposed structure to compress the signal dimension by introducing a compression matrix Φ, therefore effectively reducing the number of channels required for subsequent digitization operations

(c) Considering that the Cramér-Rao bound (CRB) indicates the lower bound of the estimation error for an unbiased system, we first derive the CRB expression for the elevation angle and auxiliary polarization angle of the proposed structure. Then, theoretical performance analysis and simulation verification are made via the CRB expression in this paper

Note that we use the CS theory twice in this paper. One is in the system design part, in order to reduce the dimension of received signal vector. The other is in the DOA estimation algorithm, where the group sparsity is used to obtain an improved number of DOF. Using the proposed structure, the number of channels is effectively controlled, thus reducing the hardware cost. In addition, although the compression leads to a degradation on the estimation performance, the proposed structure still outperforms the conventional dipole sparse array with the same number of channels, which can be clearly observed from both the theoretical derived CRB expression and the experimentally obtained root mean square error (RMSE). Therefore, the proposed structure also provides a flexible alternative option for low complexity polarization sensitive array with a relatively high estimation performance. Numerical simulations are conducted to examine the performance of the proposed structure.

The following of this paper is organized as follows: In Section 2, we build the system model of the proposed structure. Then, a CS-based algorithm is proposed in Section 3 to jointly estimate the DOAs and polarization parameters. In Section 4, the CRB expression for the estimation of DOAs and polarization parameters of the proposed structure is derived. Numerical simulation results are shown in Section 5, where the corresponding analysis is given simultaneously. Finally, Section 6 concludes the whole paper.

2. System Model of the Proposed Structure

First, we would like to briefly review the receiving model of the polarization signals. Polarization sensitive receive array is composed of L orthogonal dipoles, each of which is aligned with the y-axis in the Cartesian coordinate system. For the sake of convenience, the set $\mathbb{S} = \{d_{y1}, d_{y2}, \cdots, d_{yL}\}$ is used to represent the positions of the sensors arranged in ascending order, and the antenna at the origin is assumed to be the reference, i.e., $d_{y1} = 0$. It is noted that the sensors can be sparsely placed. For instance, the sensors can be arranged as an
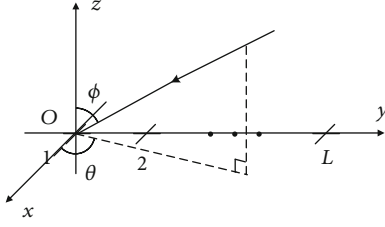
FIGURE 1: Schematic diagram of electromagnetic wave propagation.

extended coprime array according to set $\mathbb{S} = \{(Bad, 0 \leq a \leq 2A-1) \cup (Abd, 0 \leq b \leq B-1)\}$, where A and B are a pair of coprime integers. As shown in Figure 1, the signal from each dipole is processed separately; therefore, the array is divided into two subarrays according to the polarization receiving direction of the dipole:

(a) All dipoles pointing in the direction of the x-axis constitute *Subarray 1*

(b) All dipoles pointing in the direction of the y-axis constitute *Subarray 2*

Assume K narrow-band transverse electromagnetic (TEM) waves impinge upon the polarization sensitive array from the azimuth angle θ_k and elevation angle ϕ_k, where $\theta_k \in [0, \pi]$ and $\phi_k \in [-2/\pi, 2/\pi]$. It is assumed that each signal has an arbitrary elliptical electromagnetic polarization. The polarization of a TEM wave is often specified by two real parameters, namely, the auxiliary polarization angle γ_k ($\gamma_k \in [0, \pi/2]$) and the polarization phase difference η_k ($\eta_k \in [-\pi, \pi]$). The signal vector received by the polarization sensitive array, which has the dimension of $L_o = 2L$, is expressed as

$$\begin{aligned} \mathbf{x}(t) &= \sum_{k=1}^{K} [\psi(\theta_k, \phi_k, \gamma_k, \eta_k) \otimes \mathbf{u}(\theta_k, \phi_k)] \mathbf{s}_k(t) + \mathbf{n}(t) \\ &= \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \end{aligned} \tag{1}$$

where we use $\otimes$ to denote the Kronecker product. Denote $\mathbf{u}(\theta_k, \phi_k)$ as the L-dimensional spatial steering vector of the kth signal, expressed as

$$\mathbf{u}(\theta_k, \phi_k) = \left[e^{\frac{j2\pi d_{y1} \sin \theta_k \sin \phi_k}{\lambda}}, \cdots, e^{\frac{j2\pi d_{yL} \sin \theta_k \sin \phi_k}{\lambda}} \right]^T, \tag{2}$$

in which λ is the signal wavelength. For simplicity of notation, we denote $\mathbf{u}(\theta_k, \phi_k)$ as $\mathbf{u}_k$. A vector matrix $\psi(\theta_k, \phi_k, \gamma_k, \eta_k)$ which describes the polarization information of incoming signals is defined as follows:

$$\psi(\theta_k, \phi_k, \gamma_k, \eta_k) = \psi_k = \Xi_{\theta_k, \phi_k} \mathbf{h}_{\gamma_k, \eta_k}, \tag{3}$$

where Ξ and $\mathbf{h}$ are defined as

$$\Xi_{\theta_k, \phi_k} = \begin{bmatrix} -\sin \theta_k & \cos \phi_k \cos \theta_k \\ \cos \theta_k & \cos \phi_k \sin \theta_k \end{bmatrix},$$

$$\mathbf{h}_{\gamma_k, \eta_k} = \begin{bmatrix} \cos \gamma_k \\ \sin \gamma_k e^{j\eta_k} \end{bmatrix}, \tag{4}$$

respectively. Thus, $\mathbf{A} = [\psi_1 \otimes \mathbf{u}_1, \psi_2 \otimes \mathbf{u}_2, \cdots, \psi_K \otimes \mathbf{u}_K]$ is the manifold matrix of polarization received signal, and $\mathbf{s}(t) = [s_1(t), s_2(t), \cdots, s_K(t)]^T$ is the complex envelope. Noise vector $\mathbf{n}(t)$ is assumed to be the zero mean complex Gaussian processes, where each of its entries is statistically independent.

Then, let $\mathbf{x}_1(t)$ be the signal received on *Subarray 1* and $\mathbf{x}_2(t)$ the signal received on *Subarray 2*. Then, replace $\psi_k(1)$ and $\psi_k(2)$ with ψ_{xk} and ψ_{yk}, respectively. The received signal vector of subarrays are expressed as

$$\begin{aligned} \mathbf{x}_1(t) &= \sum_{k=1}^{K} [\psi_{xk} \otimes \mathbf{u}_k] s_k(t) + \mathbf{n}_1(t) = \mathbf{A}_1 \mathbf{s}(t) + \mathbf{n}_1(t), \\ \mathbf{x}_2(t) &= \sum_{k=1}^{K} [\psi_{yk} \otimes \mathbf{u}_k] s_k(t) + \mathbf{n}_2(t) = \mathbf{A}_2 \mathbf{s}(t) + \mathbf{n}_2(t). \end{aligned} \tag{5}$$

The proposed orthogonal dipole sparse array structure exploiting compressive measurements method is shown in Figure 2. The core principle of the compressive measurement method is to insert a combining network consisting of phase shifters and accumulators at the antenna outputs before subsequent digitization operations, which is equivalent to introducing a compression matrix $\Phi \in \mathbb{C}^{M \times L}$ ($M < L$) for linear operations mathematically. In this way, the received signal vector in the L-dimension is compressed to M-dimension and then output for subsequent signal processing [30].

It should be noted that the entries in the compression matrix are usually randomly selected from independent identically distributed parameters, and it is assumed that no additional noise is introduced during the compression process. In order to avoid information loss caused by data compression, the compression matrix can be optimized by various methods [27, 31, 32]. In this paper, Φ is selected to satisfy row-orthonormal, namely, $\Phi \Phi^H = \mathbf{I}_M$.

Then, the received signal vector of each subarray after compression is expressed as

$$\begin{aligned} \mathbf{y}_1(t) &= \Phi \left[\sum_{k=1}^{K} \mathbf{u}_k \psi_{xk} s_k(t) + \mathbf{n}_1(t) \right] = \bar{\mathbf{A}}_1 \mathbf{s}(t) + \bar{\mathbf{n}}_1(t), \\ \mathbf{y}_2(t) &= \Phi \left[\sum_{k=1}^{K} \mathbf{u}_k \psi_{yk} s_k(t) + \mathbf{n}_2(t) \right] = \bar{\mathbf{A}}_2 \mathbf{s}(t) + \bar{\mathbf{n}}_2(t). \end{aligned} \tag{6}$$
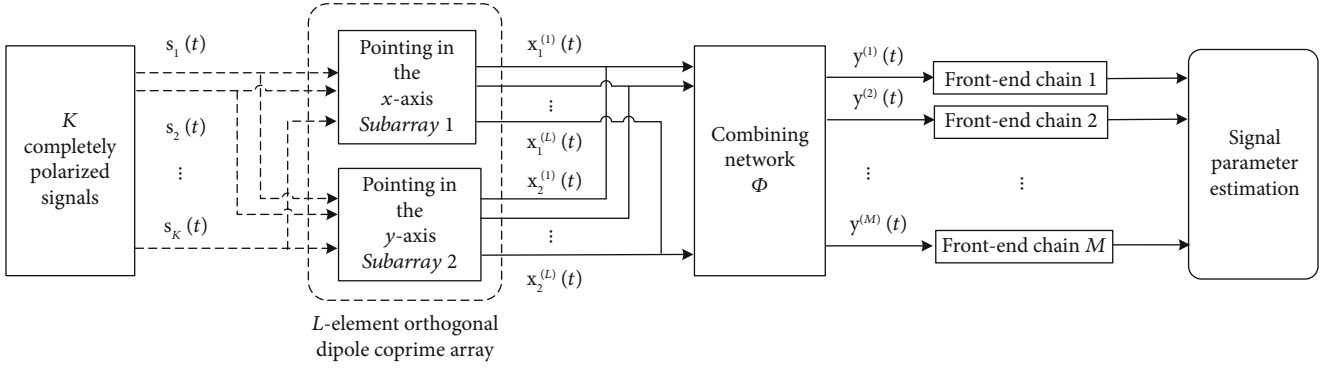
FIGURE 2: System model of the proposed compressive measurement-based orthogonal dipole sparse array structure.

Stacking the received vectors into a column vector yields

$$\mathbf{y}(t) = \tilde{\mathbf{A}}\mathbf{s}(t) + \tilde{\mathbf{n}}(t), \qquad (7)$$

where $\mathbf{y}(t) = [\mathbf{y}_1^T(t), \mathbf{y}_2^T(t)]^T$ and $\tilde{\mathbf{A}} = [\bar{\mathbf{A}}_1^T, \bar{\mathbf{A}}_2^T]^T$ is the steering vector after compression with the kth term being

$$\tilde{\mathbf{a}}_k = \psi_k \otimes \tilde{\mathbf{u}}_k = \left[\psi_{xk}, \psi_{yk}\right]^T \otimes [\Phi\mathbf{u}_k]. \qquad (8)$$

In addition, $\tilde{\mathbf{n}} = [\bar{\mathbf{n}}_1^T(t), \bar{\mathbf{n}}_2^T(t)]^T$ is the noise vector after compression.

Therefore, for the proposed structure, the length of the compressive received signal vector is $2M$, while the length of conventional sparse dipole array with the same configuration is $2L$. Assume that the number of snapshots is T. When the signal snapshots are used to compute the covariance matrix, the computational complexity of the proposed structure is $\mathcal{O}(4TM^2)$, while that of the conventional sparse dipole array is $\mathcal{O}(4TL^2)$.

3. Signal Parameter Estimation Approach

Since two-dimensional DOA estimation based on linear array cannot be realized, it is generally assumed that the signals and the linear array are in the yz plane, that is, the azimuth angle $\theta_k = \pi/2$. To avoid three-dimensional parameters, we search for ϕ_k, γ_k, and η_k; [24] proposed a different reformulation, where a CS-based approach is first used to estimate the DOAs, and then the polarization parameters are estimated utilizing the estimated DOAs and a least-square estimation approach. In this paper, the above method is improved and applied to the proposed structure.

We start with the array output covariance matrix $\mathbf{R}_{yy} = E[\mathbf{y}(t)\mathbf{y}^H(t)]$. The self-lag covariance matrix for the data vector $\mathbf{y}^{(i)}(t)$ and the cross-lag covariance matrix between $\mathbf{y}^{(i)}(t)$ and $\mathbf{y}^{(j)}(t)$ can be obtained as

$$\mathbf{R}_{yy}^{(i)} = \Phi\left(\mathbf{U}\psi(i)\mathbf{R}_{ss}\psi(i)^H\mathbf{U}^H + p_n\mathbf{I}_L\right)\Phi^H,$$

$$\mathbf{R}_{yy}^{(i,j)} = \Phi\mathbf{U}\psi(i)\mathbf{R}_{ss}\psi(j)^H\mathbf{U}^H\Phi^H, \qquad (9)$$

respectively ($1 \le i \ne j \le 2$), with $\mathbf{R}_{ss}$ representing the source covariance matrix and p_n indicating the noise power. In addition, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_K]$ is a spatial phase matrix with $L \times K$ dimension.

We denote the vectorized form of $\mathbf{R}_{yy}$ as $\mathbf{r}_{yy}$, which can be regarded as a received data vector at a virtual array with an extended coarray aperture. On the basis of the matrix algorithm, the vectorization covariance matrix of different subarrays are calculated as

$$\mathbf{r}_{yy}^{(i)} = \Phi_0\left[(\mathbf{U}^* \odot \mathbf{U})\mathrm{vec}\left(\psi(i)\mathbf{R}_{ss}\psi(i)^H\right) + p_n\tilde{\mathbf{i}}_{L^2}\right],$$

$$\mathbf{r}_{yy}^{(i,j)} = \Phi_0(\mathbf{U}^* \odot \mathbf{U})\mathrm{vec}\left(\psi(i)\mathbf{R}_{ss}\psi(j)^H\right), \qquad (10)$$

in which we denote $(\Phi^* \otimes \Phi) \in \mathbb{C}^{M^2 \times L^2}$ as Φ_0 for notational simplicity. A matrix $\mathbf{U}^* \odot \mathbf{U} = [\mathbf{u}_1^* \otimes \mathbf{u}_1, \cdots, \mathbf{u}_K^* \otimes \mathbf{u}_K]$ that leads to a series of virtual array elements is defined, and the Khatri-Rao product is denoted by $\odot$. Define a set of integers $\mathbb{D} = \{d_{y\alpha} - d_{y\beta} | d_{y\alpha}, d_{y\beta} \in \mathbb{S}\}$ to represent the locations of virtual sensors and arrange them in ascending order. Then, the corresponding array manifold can be represented as $\mathbf{U}_{\mathbb{D}} = [\mathbf{u}_{\mathbb{D}1}, \mathbf{u}_{\mathbb{D}2}, \cdots, \mathbf{u}_{\mathbb{D}K}]$.

The vectorized covariance matrices are stacked and simplified as

$$\begin{bmatrix} \mathbf{r}_{yy}^{(1)} \\ \mathbf{r}_{yy}^{(1,2)} \\ \mathbf{r}_{yy}^{(2,1)} \\ \mathbf{r}_{yy}^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi_0(\mathbf{U}^* \odot \mathbf{U})\mathbf{p}^{(1)} \\ \Phi_0(\mathbf{U}^* \odot \mathbf{U})\mathbf{p}^{(1,2)} \\ \Phi_0(\mathbf{U}^* \odot \mathbf{U})\mathbf{p}^{(2,1)} \\ \Phi_0(\mathbf{U}^* \odot \mathbf{U})\mathbf{p}^{(2)} \end{bmatrix} + \begin{bmatrix} p_n\Phi_0\tilde{\mathbf{i}}_{L^2} \\ 0 \\ 0 \\ p_n\Phi_0\tilde{\mathbf{i}}_{L^2} \end{bmatrix}, \qquad (11)$$

where $\tilde{\mathbf{i}}_{L^2} = \mathrm{vec}(\mathbf{I}_L)$ and the k-th item containing signal power is represented as

$$\mathbf{p}_k^{(1)} = p_k(\cos \gamma_k)^2, \qquad (12)$$

$$\mathbf{p}_k^{(2)} = p_k(\cos \phi_k \sin \gamma_k)^2, \qquad (13)$$

$$\mathbf{p}_k^{(1,2)} = -p_k \cos \phi_k \sin \gamma_k \cos \gamma_k e^{-j\eta_k}, \qquad (14)$$

$$\mathbf{p}_k^{(2,1)} = p_k \cos \phi_k \sin \gamma_k \cos \gamma_k e^{j\eta_k}. \qquad (15)$$

Discretizing the spatial domain $\Omega_{P,Q}$ $(P, Q \gg K)$ into a grid, let $\mathbf{d}_{\text{grid}}(\theta_p, \phi_q)$ $(1 \le p \le P, 1 \le q \le Q)$ represent the steering vector. Thus, the discretized array manifold corresponding to this grid can be obtained as

$$\mathbf{G}_{\text{grid},i} = \Phi_0 \left[\mathbf{d}_{\text{grid}}^*(\theta_1, \phi_1) \otimes \mathbf{d}_{\text{grid}}(\theta_1, \phi_1), \cdots, \mathbf{d}_{\text{grid}}^*(\theta_P, \phi_Q) \otimes \mathbf{d}_{\text{grid}}(\theta_P, \phi_Q) \right].$$
$$(16)$$

It can be known that there is a γ_k $(k \in [1, K])$ such that $\sin \gamma_k \longrightarrow 0$ or $\cos \gamma_k \longrightarrow 0$. In this case, $\mathbf{p}^{(1,2)}$ and $\mathbf{p}^{(2,1)}$ approach zero simultaneously due to the item $\sin \gamma_k \cos \gamma_k$. Thus, using $\mathbf{p}^{(1,2)}$ and $\mathbf{p}^{(2,1)}$ has no improvement on the estimation performance. Meanwhile, the computational complexity is increased. However, it is impossible for $(\cos \gamma_k)^2$ and $(\sin \gamma_k)^2$ to be equal to zero at the same time. Thus, both $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$ can be utilized for DOA estimation, and we have

$$\mathbf{r} = \tilde{\mathbf{G}}_{\text{grid}} \mathbf{p} + p_\mathbf{n} \tilde{\mathbf{I}} = \tilde{\mathbf{G}} \tilde{\mathbf{p}}, \qquad (17)$$

where $\mathbf{r} = \left[(\mathbf{r}_{\text{yy}}^{(1)})^T, (\mathbf{r}_{\text{yy}}^{(2)})^T \right]^T$, $\mathbf{p} = \left[(\mathbf{p}^{(1)})^T, (\mathbf{p}^{(2)})^T \right]^T$, $\tilde{\mathbf{I}} = \text{diag} \{ \Phi_0 \text{vec}(\mathbf{I}_L), \Phi_0 \text{vec}(\mathbf{I}_L) \}$, and $\tilde{\mathbf{G}}_{\text{grid}} = \text{diag} \{ \mathbf{G}_{\text{grid},1}, \mathbf{G}_{\text{grid},2} \}$. The nonzero entries of $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$ share the same support corresponding to the same grid. Thus, to utilize the group sparsity, hereby we define a sparse vector $\xi(\tilde{\mathbf{p}})$ as the ℓ_2-norm of each row in the matrix $[\mathbf{p}^{(1)}, \mathbf{p}^{(2)}]$. The group LASSO algorithm [21] is utilized to solve the group sparsity problem, and the minimization problem is as follows:

$$\widehat{\tilde{\mathbf{p}}} = \underset{\tilde{\mathbf{p}}}{\arg\min} \frac{1}{2} \| \mathbf{r} - \tilde{\mathbf{G}}_{\text{grid}} \tilde{\mathbf{p}} \|_2 + \mu_0 \| \xi(\tilde{\mathbf{p}}) \|_1. \qquad (18)$$

The nonzero items in $\widehat{\tilde{\mathbf{p}}}$ at its respective positions are the estimated DOAs.

The estimated elevation angle is denoted as $\widehat{\phi} = [\widehat{\phi}_1, \cdots, \widehat{\phi}_K]^T$, and the estimation of the array manifold is defined as $\hat{\mathbf{G}} = \mathbf{G}_{\text{grid}} \in \mathbb{C}^{M^2 \times K}$, which is used for the next polarization parameter estimation. Then, the vectors $\mathbf{p}^{(1)}$, $\mathbf{p}^{(1,2)}$, $\mathbf{p}^{(2,1)}$, and $\mathbf{p}^{(2)}$ which contain the polarization parameters can be obtained using the following least-square estimation:

$$\widehat{\mathbf{p}}^{(i,j)} = \left[\hat{\mathbf{G}}^H \hat{\mathbf{G}} \right]^{-1} \hat{\mathbf{G}}^H \mathbf{r}^{(i,j)}, \quad i, j = 1, 2, \qquad (19)$$

where the kth item is expressed in (12)–(15). Simplifying the above equations, we have the following equations:

$$\cos (2\gamma_k) = \frac{\widehat{\mathbf{p}}_k^{(1)} - \left(\widehat{\mathbf{p}}_k^{(2)} / \left(\cos \widehat{\phi}_k \right)^2 \right)}{p_k},$$
$$(20)$$
$$\cos (\eta_k) \sin (2\gamma_k) = \frac{\widehat{\mathbf{p}}_k^{(2,1)} - \widehat{\mathbf{p}}_k^{(1,2)}}{p_k \cos \widehat{\phi}_k}.$$

The estimated polarization phase difference $\widehat{\eta}_k$ can be obtained by expressing $\sin (2\gamma_k)$ in terms of $\cos (2\gamma_k)$, whereas the estimated auxiliary polarization angle can be calculated by the following equation:

$$\widehat{\gamma}_k = \frac{1}{2} \text{atan2} \left\{ \left| \widehat{p}_k^{(2)} / \widehat{p}_k^{(1)} \right|^{1/2}, \cos \widehat{\phi}_k \right\}. \qquad (21)$$

Thus, the joint estimation of DOA and polarization parameters is completed. The procedure of the proposed algorithm is summarized in Algorithm 1.

4. The Cramér-Rao Bound Analysis

Since the linear array is exploited, without loss of generality, the azimuth angle θ_k and polarization phase difference η_k are set as $\pi/2$, thus limiting the DOAs to the yz plane and the polarization state to the same great circle orbit of Poincare sphere. The CRB [33] provides a lower bound on the covariance matrix of any unbiased estimator. In fact, under mild regularity conditions, the maximum likelihood estimator achieves the CRB asymptotically, as the number of snapshots tends to infinity. Given T independent samples of a zero mean Gaussian process $\mathbf{y}(t)$ whose statistics depend on a parameter vector α

$$\alpha = \left[\Omega^T, Y^T, \mathbf{p}^T, p_n^{(1)}, p_n^{(2)} \right]^T, \qquad (22)$$

with $\Omega = [\phi_1, \phi_2, \cdots, \phi_K]^T$, $Y = [\gamma_1, \gamma_2, \cdots, \gamma_K]^T$, and $\mathbf{p} = [p_1, p_2, \cdots, p_K]^T$, the CRB is obtained by the inverse of the Fisher informative matrix (FIM) [34].

$$\text{CRB}(\alpha) = \frac{1}{T} \left(\mathbf{G}^H \Pi_\Delta^\perp \mathbf{G} \right)^{-1}, \qquad (23)$$

in which

$$\mathbf{G} = \mathbf{M} \left[\frac{\partial \mathbf{r}_{\text{yy}}}{\partial \phi_1}, \cdots, \frac{\partial \mathbf{r}_{\text{yy}}}{\partial \phi_K}, \frac{\partial \mathbf{r}_{\text{yy}}}{\partial \gamma_1}, \cdots, \frac{\partial \mathbf{r}_{\text{yy}}}{\partial \gamma_K} \right], \qquad (24)$$

$$\Delta = \mathbf{M} \left[\frac{\partial \mathbf{r}_{\text{yy}}}{\partial p_1}, \cdots, \frac{\partial \mathbf{r}_{\text{yy}}}{\partial p_K}, \frac{\partial \mathbf{r}_{\text{yy}}}{\partial p_n^{(1)}}, \frac{\partial \mathbf{r}_{\text{yy}}}{\partial p_n^{(2)}} \right], \qquad (25)$$

and $\mathbf{M} = \left(\mathbf{R}_{\text{yy}}^T \otimes \mathbf{R}_{\text{yy}} \right)^{-(1/2)}$.

procedure polarization signal parameter estimation

    Initialize $\mathbf{r}_{yy}^{(i)} \longleftarrow \text{vec}(\mathbf{R}_{yy}^{(i)})$, $\mathbf{r}_{yy}^{(i,j)} \longleftarrow \text{vec}(\mathbf{R}_{yy}^{(i,j)})$

    for (θ_p, ϕ_q), $1 \leq p \leq P$, $1 \leq q \leq Q$ **do**

        $\tilde{\mathbf{G}}_{\text{grid}} \longleftarrow \Phi_0[\mathbf{d}_{\text{grid}}^*(\theta_1, \phi_1) \otimes \mathbf{d}_{\text{grid}}(\theta_1, \phi_1), \cdots, \mathbf{d}_{\text{grid}}^*(\theta_P, \phi_Q) \otimes \mathbf{d}_{\text{grid}}(\theta_P, \phi_Q)]$

        $\widehat{\tilde{\mathbf{p}}} \longleftarrow \underset{\tilde{\mathbf{p}}}{\arg\min} 1/2\|\mathbf{r} - \tilde{\mathbf{G}}_{\text{grid}}\tilde{\mathbf{p}}\|_2 + \mu_0\|\xi(\tilde{\mathbf{p}})\|_1$

    end

    if $\widehat{\tilde{\mathbf{p}}}_i \neq 0$, $i \in [1, Q]$

        then $\widehat{\phi}_k \longleftarrow$ the position of $\widehat{\tilde{\mathbf{p}}}_i$, $k = 1, \cdots, K$

    end

    for $\widehat{\phi}_k \in [\widehat{\phi}_1, \cdots, \widehat{\phi}_K]$, $p_k \in [p_1, \cdots, p_K]$ **do**

        $\cos(2\gamma_k) \longleftarrow \widehat{\mathbf{p}}_k^{(1)} - (\widehat{\mathbf{p}}_k^{(2)}/(\cos\widehat{\phi}_k)^2)/p_k$

        $\cos(\eta_k)\sin(2\gamma_k) \longleftarrow \widehat{\mathbf{p}}_k^{(2,1)} - \widehat{\mathbf{p}}_k^{(1,2)}/p_k\cos\widehat{\phi}_k$

        $\widehat{\gamma}_k \longleftarrow 1/2\text{atan2}\{|\widehat{p}_k^{(2)}/\widehat{p}_k^{(1)}|^{1/2}, \cos\widehat{\phi}_k\}$

    end

    return $\widehat{\phi} \longleftarrow [\widehat{\phi}_1, \cdots, \widehat{\phi}_K]^T$, $\widehat{\eta} \longleftarrow [\widehat{\eta}_1, \cdots, \widehat{\eta}_K]^T$, $\widehat{\gamma} \longleftarrow [\widehat{\gamma}_1, \cdots, \widehat{\gamma}_K]^T$

end procedure

ALGORITHM 1: Procedure of the proposed algorithm.

Then, denote diag $\{\Phi, \Phi\}$ as $\Psi \in \mathbb{C}^{2M \times 2L}$, and vectorizing $\mathbf{R}_{yy}$ yields

$$\mathbf{r}_{yy} = \left(\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}}\right)\text{vec}(\mathbf{R}_{ss}) + \text{vec}(\mathbf{R}_{nn}). \quad (26)$$

Expand $(\tilde{\mathbf{A}}^* \otimes \tilde{\mathbf{A}})$ as $\sum_{k=1}^{K}[(\psi_k \otimes \tilde{\mathbf{u}}_k)^* \otimes (\psi_k \otimes \tilde{\mathbf{u}}_k)]$, so that the $\mathbf{r}_{yy}$ is finally simplified to the following form due to space limitation:

$$\mathbf{r}_{yy} = \sum_{k=1}^{K} \mathbf{V}_0\mathbf{H}_k\Gamma_k\mathbf{J}\mathbf{u}_{\mathbb{D}k}p_k + \mathbf{V}_0 p_n\text{vec}\left(\mathbf{I}_{L_o}\right). \quad (27)$$

Several definitions are given to illustrate the vectorized results in (27):

$$\mathbf{V}_0 = \Psi^* \otimes \Psi,$$

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{h}_k^*(1)\mathbf{I}_L & \\ & \mathbf{h}_k^*(2)\mathbf{I}_L \end{bmatrix} \otimes \begin{bmatrix} \mathbf{h}_k(1)\mathbf{I}_L & \\ & \mathbf{h}_k(2)\mathbf{I}_L \end{bmatrix},$$

$$\Gamma_k = \begin{bmatrix} -\mathbf{I}_L \\ \cos\phi_k\mathbf{I}_L \end{bmatrix} \otimes \begin{bmatrix} -\mathbf{I}_L \\ \cos\phi_k\mathbf{I}_L \end{bmatrix}. \quad (28)$$

Besides, the binary matrix $\mathbf{J}$ has the definition as follows:

$$<\mathbf{J}>_{:,d} = \text{vec}(\mathbf{I}(d)), \quad d \in \mathbb{D}, \quad (29)$$

where $\mathbf{I}(d)$ satisfies

$$<\mathbf{I}(d)>_{a,b} = \begin{cases} 1, & \text{if } a - b = d, \\ 0, & \text{otherwise}. \end{cases} \quad (30)$$

Thus, by utilizing the relationship $\mathbf{u}_k^* \otimes \mathbf{u}_k = \mathbf{J}\mathbf{u}_{\mathbb{D}k}$, $\mathbf{u}_k$ and $\mathbf{u}_{\mathbb{D}k}$ can be bridged.

Taking the derivatives of $\mathbf{r}_{yy}$ with respect to the DOA, polarization, signal power, and noise power, we have following results:

$$\frac{\partial \mathbf{r}_{yy}}{\partial \phi_k} = 2\mathbf{V}_0\mathbf{H}_k\Gamma'_k\mathbf{J}\mathbf{u}_k p_k + j\pi\mathbf{V}_0\mathbf{H}_k\Gamma_k\mathbf{J}\,\text{diag}(\mathbb{D})\cos\phi_k\mathbf{u}_k p_k,$$

$$(31)$$

$$\frac{\partial \mathbf{r}_{yy}}{\partial \gamma_k} = \mathbf{V}_0\left[\mathbf{H}'_k + \left(\mathbf{H}'_k\right)^*\right]\Gamma_k\mathbf{J}\mathbf{u}_k p_k, \quad (32)$$

$$\frac{\partial \mathbf{r}_{yy}}{\partial p_k} = \mathbf{V}_0\mathbf{H}_k\Gamma_k\mathbf{J}\mathbf{u}_k, \quad (33)$$

$$\frac{\partial \mathbf{r}_{yy}}{\partial p_n^{(1)}} = \mathbf{V}_0\text{vec}\left(\begin{bmatrix} \mathbf{I}_L & 0 \\ 0 & 0 \end{bmatrix}\right), \frac{\partial \mathbf{r}_{yy}}{\partial p_n^{(2)}} = \mathbf{V}_0\text{vec}\left(\begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I}_L \end{bmatrix}\right),$$
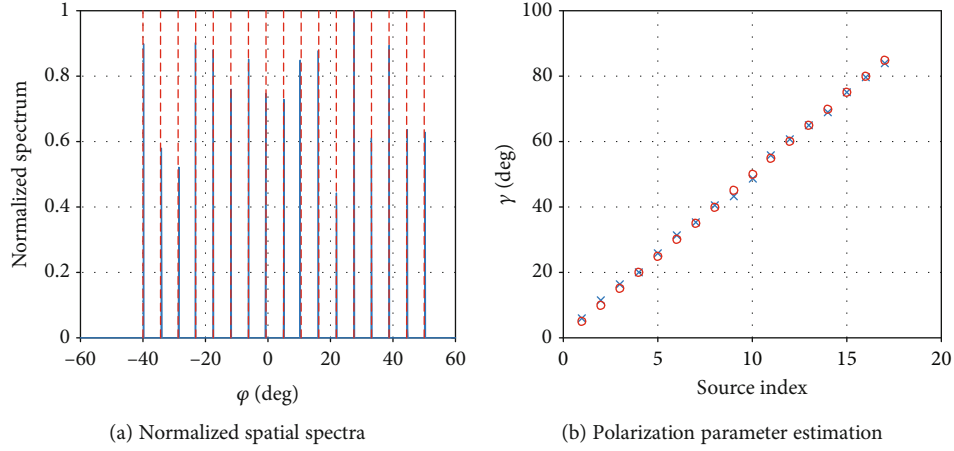
$$(34)$$

(a) Normalized spatial spectra

(b) Polarization parameter estimation

Figure 3: Signal parameter estimation of the proposed structure (SNR = 5 dB and 1,000 snapshots).

Proposed structure
Dipole coprime array 1
Dipole coprime array 2

(a) CRB (ϕ) versus no. of sources

Proposed structure
Dipole coprime array 1
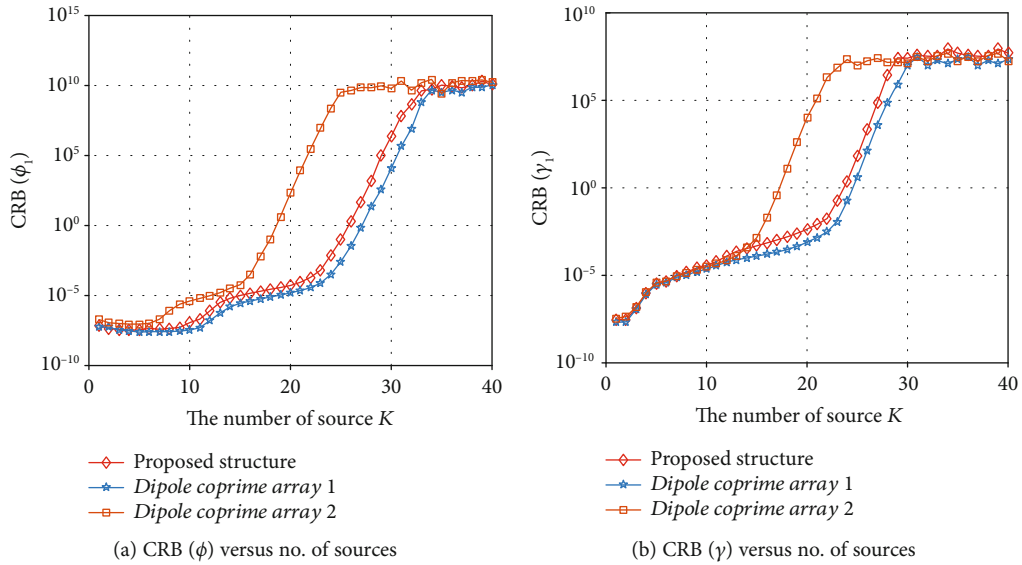Dipole coprime array 2

(b) CRB (γ) versus no. of sources

Figure 4: CRB versus the number of sources (SNR = 10 dB and 1,000 snapshots).

where

$$\Gamma'_k = \begin{bmatrix} \mathbf{0} \\ -\sin\phi_k \mathbf{I}_L \end{bmatrix} \otimes \begin{bmatrix} -\mathbf{I}_L \\ \cos\phi_k \mathbf{I}_L \end{bmatrix} + \begin{bmatrix} -\mathbf{I}_L \\ \cos\phi_k \mathbf{I}_L \end{bmatrix} \otimes \begin{bmatrix} \mathbf{0} \\ -\sin\phi_k \mathbf{I}_L \end{bmatrix},$$

$$\mathbf{H}'_k = - \begin{bmatrix} \sin\gamma_k \mathbf{I}_L & \\ & j\cos\gamma_k \mathbf{I}_L \end{bmatrix} \otimes \begin{bmatrix} \cos\gamma_k \mathbf{I}_L & \\ & j\sin\gamma_k \mathbf{I}_L \end{bmatrix}.$$

(35)

Substituting (31)–(34) into (23), (24), and (25) leads to the CRB for the proposed structure.

5. Simulation Results

Throughout our simulations, the 10-element coprime array with $A = 3$ and $B = 5$ is considered. Without additional instructions, we assume the channel number after compression is $M = 8$, and $\Phi \in \mathbb{C}^{8\times10}$ is generated from the standard complex Gaussian distribution. The incident signal is generated uniformly in the range of $\phi_k \in [-40°, 50°]$ with $\theta_k = \pi/2$, and the auxiliary polarization angle and polarization phase difference are evenly distributed in $\gamma_k \in [5°, 85°]$ and $\eta_k \in [-120°, 150°]$. Under the condition that SNR = 5 dB and 1,000 snapshots, the elevation angle is estimated using the CS-based approach. The angle range from $-60°$ to $60°$ is uniformly divided into grids with a $0.1°$ searching step.

The signal parameter estimation results of 17 sources using the proposed structure are shown in Figure 3, where the actual parameter values are marked in red. Compared with the results in [24], the proposed structure with channel number compression can still correctly analyze 17 signals

(a) RMSE (ϕ) versus SNR
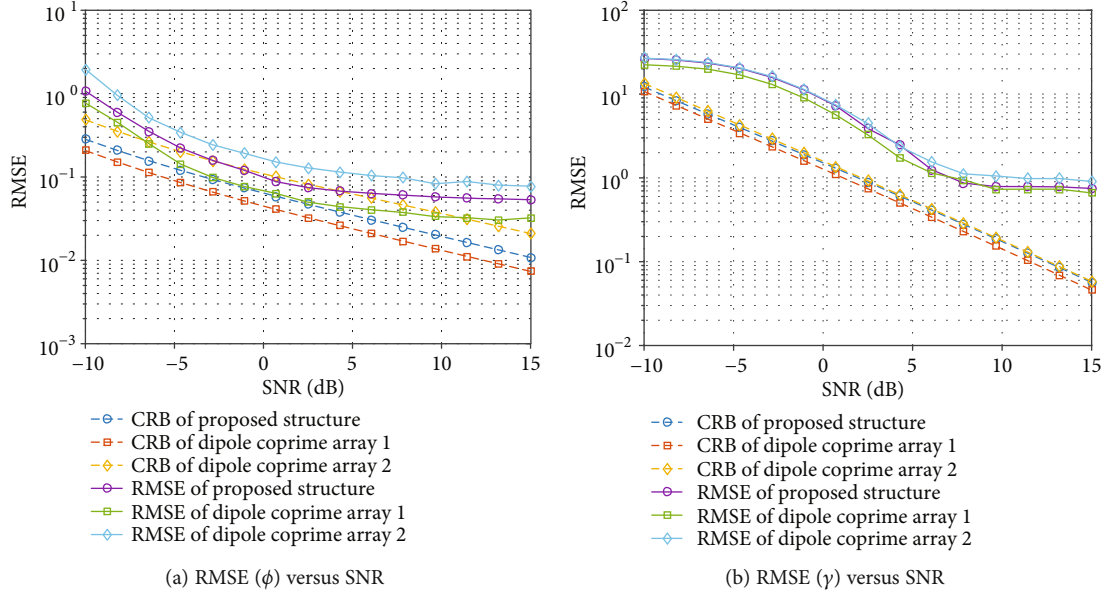
(b) RMSE (γ) versus SNR

FIGURE 5: RMSE versus SNR (1,000 snapshots).

with 10 physical elements. Then, the variation of CRB derived with the number of sources is simulated to verify the DOF that the proposed structure can obtain. To further analyze the estimation performance of the proposed structure, the following coprime array configurations are considered as the comparison structure: (a) the proposed structure with $L = 10, M = 8$; (b) the 10-element coprime array with the compression matrix Φ is equal to a unit matrix, that is, without compression, denoted as *dipole coprime array 1*; and (c) the 8-element coprime array with $A = 2$ and $B = 5$, namely, $L = M = 8$, denoted as *dipole coprime array 2*, in which the idea of the CACIS configuration idea proposed in [35] is used for reference.

Figure 4 describes the CRB curves of ϕ and γ versus the number of sources when taking 1,000 snapshots at 10 dB SNR. When the compression ratio is $L/M = 1.25$, the DOFs obtained by the proposed structure is not much less than that obtained by an uncompressed structure with the same number of physical elements, namely, *dipole coprime array 1*. However, compared with *dipole coprime array 2* with the same number of channels, the proposed structure can resolve more uncorrelated signal sources, showing the superiority of the proposed structure for the number of DOFs.

The RMSE versus SNR of the comparison structure is shown in Figure 5. The CRB (ϕ) and CRB (γ) of the three array configurations are presented by dash lines. The order of RMSE depicted by solid lines is consistent with that of CRB shown in Figure 4. It can be observed in Figure 4 that the downward trend of the RMSE curves can also fit the CRB curves well, while in Figure 4, the RMSE curves tend to be flat as the SNR increases since the search step is limited by the RIP criterion. Due to the largest number of channels, the estimation accuracy of *dipole coprime array 1* is the highest among the three configurations, while the computational complexity is also the highest. In the case of the same number of channels, the proposed structure has a lower RMSE than

dipole coprime array 2. In general, from the perspective of the number of inequalities, dimensionality reduction can inevitably lead to a decrease in estimation performance, such as DOF and estimation accuracy. In order to achieve the purpose of avoiding excessive system complexity, better estimation performance can be obtained by using the proposed structure for DOA estimation, which also verifies our previous analysis. On the other hand, in Figure 4, it can be observed that, in a low SNR region, the performance of the proposed structure is almost the same as that of the *dipole coprime array 2*, indicating that the compressive measurement-based structure has no significant improvement on the estimation of polarization parameters. However, in a large SNR region, the accuracy of the proposed structure approaches the *dipole coprime array 1*. We must note that this improvement is mainly caused by limitation of the step of the searching grid. To be specific, as the SNR increases, the estimation performance achieves the ceiling of current searching step, thus leading to the phenomenon that the proposed structure has almost the same performance as *dipole coprime array 1*. Theoretically speaking, by observing the CRB curves shown in Figures 3 and 4, the improvement on the estimation of polarization parameters is negligible.

6. Conclusion

In this paper, we proposed a compressive measurement-based orthogonal dipole sparse array structure, which can be used for high-performance signal parameter estimation with a small number of given elements. In the joint estimation algorithm of DOA and polarization parameters, the CS-based algorithm and the least-square estimation method were adopted. Then, based on compressive measurements, we derived the CRB expression for the elevation angle and auxiliary polarization angle. By comparing to the array configurations with the unit matrix as the compression matrix, we

considered the CRB curves versus the number of independent signal sources and SNR. Thus, we come to the conclusion that, under the condition that we control the system complexity by reducing the number of channels, better parameter estimation performance can be obtained by the proposed orthogonal dipole array structure, especially for the estimation of DOAs. Simulation results also verified the theoretical analysis.

Data Availability

If data is needed, please contact M. Guo (email: guomuran@hrbeu.edu.cn) for the code of numerical simulations.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

References

[1] V. H. Rumsey, G. A. Deschamps, M. L. Kales, J. I. Bohnert, and H. G. Booker, "Techniques for handling elliptically polarized waves with special reference to antennas: introduction," *Proceedings of the IRE*, vol. 39, no. 5, pp. 533-534, 1951.

[2] W. Lee and Yu Yeh, "Polarization diversity system for mobile radio," *IEEE Transactions on Communications*, vol. 20, no. 5, pp. 912–923, 1972.

[3] D. Giuli, "Polarization diversity in radars," *Proceedings of the IEEE*, vol. 74, no. 2, pp. 245–269, 1986.

[4] R. G. Vaughan, "Polarization diversity in mobile communications," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 3, pp. 177–186, 1990.

[5] K. Ouchi, "Recent trend and advance of synthetic aperture radar with selected topics," *Remote Sensing*, vol. 5, no. 2, pp. 716–807, 2013.

[6] E. H. Satorius, Z. Ye, and E. D. Archer, "Polarization combining scheme for radio direction finding with multipath," in *2001 MILCOM Proceedings Communications for Network-Centric Operations: Creating the Information Force (Cat. No.01CH37277)*, vol. 1, pp. 383–387, McLean, VA, USA, 2001.

[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[8] Y. Tian, X. Sun, and S. Zhao, "Sparse-reconstruction-based direction of arrival, polarisation and power estimation using a cross-dipole array," *IET Radar, Sonar and Navigation*, vol. 9, no. 6, pp. 727–731, 2015.

[9] G. Zheng, "Two-dimensional DOA estimation for polarization sensitive array consisted of spatially spread crossed-dipole," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5014–5023, 2018.

[10] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Transactions on Antennas and Propagation*, vol. 16, no. 2, pp. 172–175, 1968.

[11] P. Pal and P. P. Vaidyanathan, "Nested arrays: a novel approach to array processing with enhanced degrees of free-dom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, 2010.

[12] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.

[13] J. He, Z. Zhang, T. Shu, and W. Yu, "Sparse nested array with aperture extension for high accuracy angle estimation," *Signal Processing*, vol. 176, article 107700, 2020.

[14] J. He, L. Li, and T. Shu, "Sparse nested arrays with spatially spread orthogonal dipoles: high accuracy passive direction finding with less mutual coupling," *IEEE Transactions on Aerospace and Electronic Systems*, 2021.

[15] J. He, L. Li, and T. Shu, "Sparse nested arrays with spatially spread square acoustic vector sensors for high accuracy under-determined direction finding," *IEEE Transactions on Aerospace and Electronic Systems*, 2021.

[16] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. D. Zhang, "Direction of-arrival estimation for coprime array via virtual array interpolation," *IEEE Transactions on Signal Processing*, vol. 66, no. 22, pp. 5956–5971, 2018.

[17] Z. Zhang, C. Zhou, Y. Gu, J. Zhou, and Z. Shi, "An IDFT approach for coprime array direction-of-arrival estimation," *Digital Signal Processing*, vol. 94, pp. 45–55, 2019.

[18] C. Zhou, Y. Gu, S. He, and Z. Shi, "A robust and efficient algorithm for coprime array adaptive beamforming," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1099–1112, 2018.

[19] L. Wan, L. Sun, K. Liu, X. Wang, Q. Lin, and T. Zhu, "Autonomous vehicle source enumeration exploiting non-cooperative UAV in software defined internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.

[20] L. Wan, Y. Sun, L. Sun, Z. Ning, and J. J. P. C. Rodrigues, "Deep learning based autonomous vehicle super resolution DOA estimation for safety driving," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, 2020.

[21] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based DOA estimation using co-prime arrays," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3967–3971, Vancouver, BC, Canada, 2013.

[22] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[23] J. Yang, T. Chen, L. Shi, and C. Zhang, "Joint DOA and polarization estimation based on multi-polarization sensitive array," in *International Conference in Communications, Signal Processing, and Systems*, Springer, Harbin, China, 2017.

[24] B. K. Chalise, Y. D. Zhang, and B. Himed, "Compressed sensing based joint DOA and polarization angle estimation for sparse arrays with dual-polarized antennas," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 251–255, Anaheim, CA, U. S, 2018.

[25] L. Wan, K. Liu, Y.-C. Liang, and T. Zhu, "DOA and polarization estimation for non-circular signals in 3-D millimeter wave polarized massive MIMO systems," *IEEE Transactions on Wireless Communications*, 2021.

[26] C. Zhou, Y. Gu, Y. D. Zhang, Z. Shi, T. Jin, and X. Wu, "Compressive sensing-based coprime array direction-of-arrival estimation," *IET Communications*, vol. 11, no. 11, pp. 1719–1724, 2017.

[27] T. Chen, M. Guo, and X. Huang, "Direction finding using compressive one-bit measurements," *IEEE Access*, vol. 6, no. 1, pp. 41201–41211, 2018.

[28] T. Chen, X. Han, and Y. Yu, "A sub-Nyquist sampling digital receiver system based on array compression," *Progress In Electromagnetics Research Letters*, vol. 88, pp. 21–28, 2020.

[29] T. Chen, J. Yang, and M. Guo, "A MIMO radar-based DOA estimation structure using compressive measurements," *Sensors*, vol. 19, no. 21, 2019.

[30] M. Guo, Y. D. Zhang, and T. Chen, "DOA estimation using compressed sparse array," *IEEE Transactions on Signal Processing*, vol. 66, no. 15, pp. 4133–4146, 2018.

[31] M. Ibrahim, V. Ramireddy, A. Lavrenko et al., "Design and analysis of compressive antenna arrays for direction of arrival estimation," *Signal Processing*, vol. 138, no. 9, pp. 35–47, 2017.

[32] Y. Gu, Y. D. Zhang, and N. A. Goodman, "Optimized compressive sensing-based direction-of-arrival estimation in massive MIMO," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3181–3185, New Orleans, LA, 2017.

[33] P. Stoica, E. G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: a textbook derivation," *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 148–150, 2001.

[34] C. Liu and P. P. Vaidyanathan, "Cramer-Rao bounds for coprime and other sparse arrays, which find more sources than sensors," *Digital Signal Processing*, vol. 61, pp. 43–61, 2017.

[35] S. Qin, Y. D. Zhang, and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1377–1390, 2015.

WILEY | Hindawi

Research Article

Crosscorrelation and DOA Estimation for L-Shaped Array via Decoupled Atomic Norm Minimization

Yu Zhang [1] **Yinan Sun,** [2] **Gong Zhang** [1] **Xinhai Wang,** [3] **and Yu Tao** [4]

[1]*Key Lab of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[2]*Electronic Information School, Wuhan University, Wuhan 430072, China*
[3]*Nanjing Marine Radar Institute, Nanjing 211153, China*
[4]*School of Electronic and Information Engineering, Changshu Institute of Technology, Suzhou 215500, China*

Correspondence should be addressed to Gong Zhang; gzhang@nuaa.edu.cn

A novel two-phase method for two-dimensional (2D) direction-of-arrival (DOA) estimation with L-shaped array based on decoupled atomic norm minimization (DANM) is proposed in this paper. In the first phase, given the sample crosscorrelation matrix, the gridless DANM technique considering the noise and finite snapshots effects is employed to exploit the structure and sparse properties of the crosscorrelation matrix. The resulting DANM-based algorithm not only enables the crosscorrelation matrix reconstruction (CCMR) but also reconstructs the covariance matrix of the L-shaped array. Hence, sequentially, in the second phase, the conventional 2D DOA estimators for the L-shaped array can be adopted for the angle estimation. With appropriate 2D DOA estimators, the resulting proposed algorithms can not only achieve better performance but also detect more source number, compared with conventional crosscorrelation-based DOA estimators. Moreover, the proposed method, termed CCMR-DANM, not only has blind characteristic that it does not require the prior information of source numbers but also is more efficient than the existing CCMR-based counterparts. Numerical simulations demonstrate the effectiveness and outperformance of the proposed method.

1. Introduction

The problem of two-dimensional (2D) direction-of-arrival (DOA) estimation plays an important role in array signal processing and has attracted much interest in the area of wireless communications, radar and sonar [1–7]. For 2D DOA estimation, many array structures, such as rectangular arrays, circular arrays, and L-shaped arrays, have been developed. Among these arrays, since the L-shaped array can achieve better estimation performance than others, it has attracted a lot of attentions and many corresponding algorithms for 2D DOA estimation have been proposed in last several decades [8–15]. Moreover, these algorithms can be divided into three categories. The first is to separately estimate the angles corresponding to each uniform linear subarray

based on the covariance matrix of each subarray with conventional 1D DOA estimators, such as MUSIC [16] and ESPRIT [17]. However, an extra pairing operation is needed in these algorithms [18]. The second is to jointly estimate the two angles based on the covariance matrix of the L-shaped array. They can detect more source numbers than the first ones and do not need an extra pairing [8, 9]. The last is based on the crosscorrelation of the L-shaped array, which is naturally contaminated by less noise, compared with the covariance matrix. As a result, the corresponding algorithms can achieve a better performance in low signal-to-noise ratio (SNR) [10–15]. It is worth noting that in practical applications, all these three kinds of methods need the prior information of source numbers and are employed with sample matrices (no matter the sample covariance matrices or the

sample crosscorrelation matrices). Note that the sample matrix is calculated from finite collected snapshots of the observed signals which is contaminated with additive noise. Hence, the sample matrix cannot equip the ideal structure of the ideal matrix, which leads to performance degradation especially in low SNR and small number of snapshots [19].

To overcome these disadvantages, the basic idea is to first reconstruct the ideal matrix from the sample matrix before employing DOA estimation methods. Inspired by this thought, several crosscorrelation matrix reconstruction-(CCMR-) based methods are proposed. Specifically, cross-correlation atomic norm minimization (CC-ANM) utilising the 2D ANM technique [20] for CCMR is proposed in [21]. And [22] grafts the covariance fitting criterion [23] for CCMR and proposes a crosscorrelation gridless sparse iterative covariance-based estimation (CC-GLS) method. However, a high-dimensional two-level Toeplitz matrix is needed to be constructed in both methods, which leads to high computational complexity [18]. In contrast, we propose an efficient decoupled ANM- (DANM-) based method for CCMR in [24], while the method does not consider the finite snapshot effect, which results in performance degradation in practical applications.

In this paper, given the sample crosscorrelation matrix, the DANM technique [18] considering the noise and finite snapshots effects is employed to exploit the structure and sparse properties of the crosscorrelation matrix. Moreover, the resulting DANM-based algorithm not only enables the crosscorrelation matrix reconstruction but also reconstructs the covariance matrix of the L-shaped array. Hence, the conventional 2D DOA estimators for the L-shaped array can be sequentially adopted for the angle estimation. The proposed two-phase method is more computationally efficient than the aforementioned two CCMR-based methods and can achieve better estimation performance compared with traditional crosscorrelation-based methods. Numerical simulations demonstrate the effectiveness and outperformance of the proposed method.

The rest of this paper is organized as follows. Section 2 presents the signal model and problem formulation for 2D DOA estimation with L-shaped array. Section 3 proposes an efficient DANM-based two-phase method for crosscorrelation and DOA estimation. Section 4 discusses the computational complexity and extends the proposed method to the sparse L-shaped array cases. Section 5 presents simulation results, followed by conclusions in Section 6.

Throughout this paper, a, $\mathbf{a}$ and $\mathbf{A}$ denote a scalar, a vector, and a matrix, respectively. $\mathrm{Tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$. $\mathrm{T}(\mathbf{a})$ denotes the Hermitian Toeplitz matrix with the first column being $\mathbf{a}$. $\mathrm{diag}\,(\mathbf{A})$ is a column vector formed from the elements of the main diagonal of $\mathbf{A}$ and $\mathrm{diag}\,(\mathbf{a})$ generates a diagonal matrix with the diagonal elements constructed from $\mathbf{a}$. $\mathrm{vec}(\cdot)$ stacks all the columns of a matrix into a vector. $\mathbf{I}_a$ is an a-size identity matrix, and $\mathbf{I}_\Omega$ is a selection matrix with index set Ω. $\mathbf{0}_a$ and $\mathbf{1}_a$ are the $a \times 1$ zeros and one vectors, respectively. $\otimes$ is the Kronecker product. $\mathbb{E}\{\cdot\}$ denotes expectation, and $\mathrm{Var}(\cdot)$ denotes variance. We use $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ to denote the transpose, the conjugate, and the conjugate transpose operation, respectively.

2. Problem Formulation

Consider K far-field narrowband source signals $\{\mathbf{s}_i(t)\}_{i=1}^K$ impinging on an L-shaped array from distinct directions at element angles $\{\phi_i\}_{i=1}^K$ and azimuth angles $\{\theta_i\}_{i=1}^K$, as shown in Figure 1. The L-shaped array consists of two ULAs of M omnidirectional sensors which are uniformly spaced with a spacing of d along the x axis and y axis, respectively. The observed signals of the L-shaped array can be expressed as [24]

$$
\begin{aligned}
\mathbf{x}(t) &= \mathbf{A}_x \mathbf{s}(t) + \mathbf{n}_x(t), \\
\mathbf{y}(t) &= \mathbf{A}_y \mathbf{s}(t) + \mathbf{n}_y(t), t = 1, 2, \cdots, L,
\end{aligned}
\tag{1}
$$

where t indexes the snapshot; L denotes the number of collected snapshots; and $\boldsymbol{s}(t)$, $\mathbf{n}_x(t)$, and $\mathbf{n}_y(t)$ denote the vector of source signals and the vector of additive noise corresponding to the x subarray and y subarray at the snapshot t, respectively. $\mathbf{A}_x = [\mathbf{a}_x(\theta_1, \phi_1), \cdots, \mathbf{a}_x(\theta_K, \phi_K)]$ and $\mathbf{A}_y = [\mathbf{a}_y(\theta_1, \phi_1), \cdots, \mathbf{a}_y(\theta_K, \phi_K)]$ are the array manifold matrices of the x subarray and the y subarray, whose ith columns are the steering vectors of the ith source which satisfy

$$
\mathbf{a}_x(\theta_i, \phi_i) = \left[1, e^{-j2\pi 1/2 \cos{(\theta_i)} \sin{(\phi_i)}}, \cdots, e^{-j2\pi 1/2(M-1) \cos{(\theta_i)} \sin{(\phi_i)}} \right]^T,
$$

$$
\mathbf{a}_y(\theta_i, \phi_i) = \left[1, e^{-j2\pi 1/2 \sin{(\theta_i)} \sin{(\phi_i)}}, \cdots, e^{-j2\pi 1/2(M-1) \sin{(\theta_i)} \sin{(\phi_i)}} \right]^T.
\tag{2}
$$

Herein, the spaced distance d is assumed to be equal to half of the wavelength λ. Moreover, let $f_{i,1} = (1/2) \cos{(\theta_i)} \sin{(\phi_i)}$, $f_{i,2} = (1/2) \sin{(\theta_i)} \sin{(\phi_i)}$ and $\Omega = \{(f_{1,1}, f_{1,2}), \cdots, (f_{K,1}, f_{K,2})\}$ denote the frequencies that correspond to the direction on x and y axes and the set of corresponding frequencies, respectively. Note that $\{f_{i,1}, f_{i,2}\} \longleftrightarrow \{\theta_i, \phi_i\}$ is the one-to-one mapping. Once the estimation of $\{f_{i,1}, f_{i,2}\}$ is obtained, the corresponding $\{\theta_i, \phi_i\}$ can be retrieved as

$$
\begin{aligned}
\widehat{\phi}_i &= \arcsin\left(2\sqrt{\hat{f}_{i,1}^2 + \hat{f}_{i,2}^2} \right), \\
\widehat{\theta}_i &= \arccos\left(\frac{\hat{f}_{i,2}}{\hat{f}_{i,1}} \right).
\end{aligned}
\tag{3}
$$

Hence, in the following paper, we consider estimation of $\{f_{i,1}, f_{i,2}\}$ instead of $\{\theta_i, \phi_i\}$ for notational simplicity. Moreover, in this paper, the source signals $\mathbf{s}(t)$ are assumed uncorrelated with each other and the noise $\mathbf{n}_x(t)$ and $\mathbf{n}_y(t)$ are i.i.d. additive white Gaussian random processes satisfying $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ and are statistically independent of $\mathbf{s}(t)$. Therefore, under the above assumptions, we have the
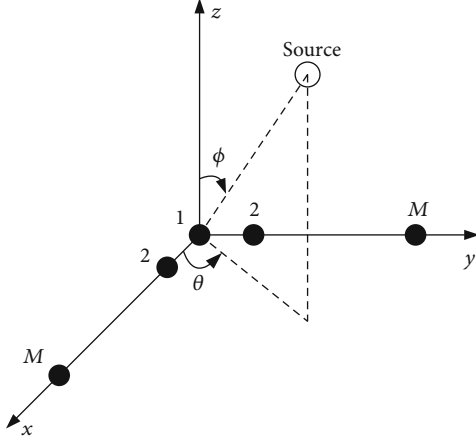
FIGURE 1: L-shaped array configuration.

crosscorrelation matrix between $\mathbf{x}(t)$ and $\mathbf{y}(t)$ can be expressed as

$$
\begin{aligned}
\mathbf{R}_{xy} = E\left[\mathbf{x}(t)\mathbf{y}^H(t)\right] &= \mathbf{A}_x\mathbf{R}_s\mathbf{A}_y + \mathbf{N}_1 = \mathbf{R}_c + \mathbf{N}_1 \\
&= \sum_{i=1}^{K} r_i\mathbf{a}_x\left(f_{i,1}\right)\mathbf{a}_y^H\left(f_{i,2}\right) + \mathbf{N}_1,
\end{aligned}
\tag{4}
$$

where $\mathbf{N}_1$ is the matrix with only the first element of main diagonal being σ^2 and zero otherwise. $\mathbf{R}_s = \text{diag}\{\mathbf{r}\}$ is the source correlation matrix with $\mathbf{r} = [r_1, \cdots, r_K]^T \geq 0$, i.e., $\mathbb{E}(|s_i(t)|^2) = r_i$. $\mathbf{R}_c = \mathbf{A}_x\mathbf{R}_s\mathbf{A}_y$ is the noise-free crosscorrelation matrix, while, in practical applications, $\mathbf{R}_{xy}$ can only be estimated from the finite L snapshots by $\widehat{\mathbf{R}}_{xy} = 1/L\sum_{t=1}^{L}\mathbf{x}(t)\mathbf{y}^H(t)$, which not only is contaminated by the noise but also contains errors caused by the finite snapshot effect. The goal of this paper is to recover the noise-free crosscorrelation matrix $\mathbf{R}_c$ and then the unknown 2D DOAs $\{\theta_i, \phi_i\}_i$ from the sample crosscorrelation matrix $\widehat{\mathbf{R}}_{xy}$.

3. Proposed Method

In this section, we propose an efficient DANM-based crosscorrelation and DOA estimation method for L-shaped array. To this end, the standard DANM technique is firstly presented. Then, the sparse representation of the noise-free crosscorrelation matrix is presented, which enables the DANM technique to exploit its sparse property. Further, simultaneously considering the noise and finite snapshot effects, and the structure property of the crosscorrelation matrix, we propose the original DANM-based formulation for CCMR which is intractable since the structure constraint. To make the formulation tractable, an effective relaxation is proposed. Moreover, an estimation error constraint leading to easy setting of the user-specific parameter is also proposed.

3.1. Prior Art: DANM Technique. We now review the standard DANM technique for harmonic retrieval. Define a matrix $\mathbf{Z}$ as

$$
\mathbf{Z} = \sum_{k=1}^{K} \alpha_k\mathbf{a}_x\left(f_{k,1}\right)\mathbf{a}_y^H\left(f_{k,2}\right),
\tag{5}
$$

where $\alpha_k \in \mathbb{C}$. Then, based on (5), a matrix-form atom set of infinite size is defined as [18, 15]

$$
\mathscr{A} = \left\{\mathbf{a}_x(f_1)\mathbf{a}_y^H(f_2) \mid f_1, f_2 \in \left[-\frac{1}{2}, \frac{1}{2}\right)\right\}.
\tag{6}
$$

Accordingly, the atomic norm of $\mathbf{Z}$ over the atom set $\mathscr{A}$ is defined as

$$
\begin{aligned}
\|\mathbf{Z}\|_{\mathscr{A}} = \inf \Bigg\{ &\sum_k |\alpha_k| \mid \sum_k \alpha_k\mathbf{a}_x\left(f_{k,1}\right)\mathbf{a}_y^H\left(f_{k,2}\right), \mathbf{a}_x\left(f_{k,1}\right)\mathbf{a}_y^H \\
&\cdot \left(f_{k,2}\right) \in \mathscr{A}; \alpha_k \in \mathbb{C} \Bigg\},
\end{aligned}
\tag{7}
$$

which seeks the sparsest (under l_1-norm measure) decomposition of $\mathbf{Z}$ over $\mathscr{A}$. Consider the matrix $\mathbf{Z}$ is contaminated by the noise matrix $\mathbf{N}$ and the matrix at hand is $\widehat{\mathbf{Z}} = \mathbf{Z} + \mathbf{N}$. Then, according to the DANM theory [18, 25], the atomic decomposition yields the true structure in (5), through the following DANM formulation:

$$
\tilde{\mathbf{Z}} = \arg \min_{\mathbf{Z}}\|\mathbf{Z}\|_{\mathscr{A}}\text{s.t.}\left\|\widehat{\mathbf{Z}} - \mathbf{Z}\right\|_F^2 \leq \eta,
\tag{8}
$$

where η is a user-specified parameter for error tolerance. Moreover, (8) is equivalent to the following semidefinite positive (SDP) formulation

$$
\begin{aligned}
\left\{\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2, \tilde{\mathbf{Z}}\right\} = \arg \min_{\mathbf{z}_1, \mathbf{z}_2, \mathbf{Z}} &\frac{1}{2M}\left(\text{Tr}(\text{T}(\mathbf{z}_1)) + \text{Tr}(\text{T}(\mathbf{z}_2))\right) \\
\text{s.t.} \quad &\left\|\widehat{\mathbf{Z}} - \mathbf{Z}\right\|_F^2 \leq \eta \\
&\begin{bmatrix} \text{T}(\mathbf{z}_1) & \mathbf{Z} \\ \mathbf{Z}^H & \text{T}(\mathbf{z}_2) \end{bmatrix} \succeq \mathbf{0}.
\end{aligned}
\tag{9}
$$

3.2. Standard DANM-Based Formulation for CCMR. Note that $\mathbf{R}_c = \sum_{i=1}^{K} r_i\mathbf{a}_x(f_{i,1})\mathbf{a}_y^H(f_{i,2})$. Apparently, $\mathbf{R}_c$ has a sparse linear atomic representation over the matrix-form atom set in (6). And we introduce a new matrix-form atomic norm

$$
\begin{aligned}
\|\mathbf{R}_c\|_{\mathscr{A}}^+ = \inf \Bigg\{ &\sum_i r_i \mid \sum_i r_i\mathbf{a}_x\left(f_{i,1}\right)\mathbf{a}_y^H\left(f_{i,2}\right), \mathbf{a}_x\left(f_{i,1}\right)\mathbf{a}_y^H \\
&\cdot \left(f_{i,2}\right) \in \mathscr{A}; r_i \geq 0, \forall i \Bigg\}.
\end{aligned}
\tag{10}
$$

Note that this norm differs from the original atomic norm of the standard DANM in (7), because of the extra constraint $\mathbf{r} \geq \mathbf{0}$.

Given $\mathbf{R}_c$, it is possible to retrieve the components $\{r_i, f_{i,1}, f_{i,2}\}$ of its sparest representation by calculating its atomic norm. Considering the obtained at hand is the sample cross-correlation containing the estimation error, it boils down to

$$\tilde{\mathbf{R}}_c = \arg \min_{\mathbf{R}_c} \|\mathbf{R}_c\|_{\mathscr{A}}^+$$
$$\text{s.t.} \quad \mathscr{F}\{\hat{\mathbf{R}}_{xy}, \mathbf{R}_c\} \le \beta. \tag{11}$$

where $\mathscr{F}\{\hat{\mathbf{R}}_{xy}, \mathbf{R}_c\}$ is the function quantifying the estimation error and β indicates the error tolerance threshold. It is similar to the DANM formulation introduced in (8), but defined on the norm $\|\cdot\|_{\mathscr{A}}^+$ instead of $\|\cdot\|_{\mathscr{A}}$. To show the intricacy of this difference, we rewrite (11) in the following equivalent form:

$$\{\tilde{\mathbf{R}}_c, \tilde{\mathbf{r}}\} = \min_{\mathbf{R}_c, \mathbf{r}} \|\mathbf{R}_c\|_{\mathscr{A}} = \|\mathbf{r}\|_1 \tag{12a}$$

$$\text{s.t.} \quad \mathscr{F}\{\hat{\mathbf{R}}_{xy}, \mathbf{R}_c\} \le \beta \tag{12b}$$

$$\mathbf{R}_c = \mathbf{A}_x \, \text{diag}\,(\mathbf{r}) \mathbf{A}_y^H \tag{12c}$$

$$\mathbf{r} \ge \mathbf{0}. \tag{12d}$$

Note that (12c) is implicit in the objective function (11) but becomes an explicit constraint because of the new non-negative constraint (12d). Without (12d), the SDP implementation of the DANM in (9) can be used to reformulate (12) into a convex problem. However, because of the extra constraint on $\mathbf{r}$ in (12d), this problem becomes intractable, because $\mathbf{r}$ is intertwined with the other variable $\mathbf{R}_c$ in the form of (12c).

3.3. Effective Relaxation. To solve (12) in a tractable manner, we seek to relax $\mathbf{r} \ge \mathbf{0}$ to an effective form with respect to $\mathbf{R}_c$. To this end, we note that

$$\mathbf{R}_c(:,1) = \sum_{i=1}^K r_i \mathbf{a}_x (f_{i,1}),$$
$$\mathbf{R}_c(1,:)^H = \sum_{i=1}^K r_i \mathbf{a}_y (f_{i,2}), \tag{13}$$

where $\mathbf{R}_c(:,1)$ and $\mathbf{R}_c(1,:)$ denote the first column and the first row of $\mathbf{R}_c$, respectively. Moreover, according to the property of Toeplitz matrices [26], if $\mathbf{r} \ge 0$, we have

$$T(\mathbf{R}_c(:,1)) \succeq \mathbf{0},$$
$$T(\mathbf{R}_c(1,:)^H) \succeq \mathbf{0}. \tag{14}$$

Adopting (14) to replace (12d) and reformulating (12a)–(12c) into the original decoupled SDP form in (9), we reach the following effective SDP relaxation for (11):

$$\{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \tilde{\mathbf{R}}_c\} = \arg \min_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{R}_c} \frac{1}{2M} (\text{Tr}(T(\mathbf{u}_1)) + \text{Tr}(T(\mathbf{u}_2)))$$

$$\text{s.t.} \quad \mathscr{F}\{\hat{\mathbf{R}}_{xy}, \mathbf{R}_c\} \le \beta$$

$$T(\mathbf{R}_c(:,1)) \succeq \mathbf{0}, T(\mathbf{R}_c(1,:)^H) \succeq \mathbf{0}$$

$$\begin{bmatrix} T(\mathbf{u}_1) & \mathbf{R}_c \\ \mathbf{R}_c^H & T(\mathbf{u}_2) \end{bmatrix} \succeq \mathbf{0}. \tag{15}$$

With an appropriate definition of $\mathscr{F}\{\hat{\mathbf{R}}_{xy}, \mathbf{R}_c\}$ and β, (18) can be solved successfully via off-the-shelf convex solvers, such as CVX [27].

3.4. Estimation Error Constraint. To define a nice $\mathscr{F}\{\hat{\mathbf{R}}_{xy}, \mathbf{R}_c\}$ which leads to easy setting of β, we denote the estimation error matrix as

$$\mathbf{E} = \hat{\mathbf{R}}_{xy} - \mathbf{R}_c. \tag{16}$$

Denote by $\varepsilon_{p,q}$ the (p, q)th element of the estimation error matrix $\mathbf{E}$; then, one has the following proposition:

Proposition 1. *For adequately large L, and $\forall p + q - 2 > 0$, $\varepsilon_{p,q}$ is approximately circular complex Gaussian distributed with zero mean, and the variance is*

$$Var(\varepsilon_{p,q}) = \frac{1}{L} \left\{ \left(\sum_{i=1}^K r_i + \sigma^2 \right)^2 - \sum_{i=1}^K r_i^2 \right\}. \tag{17}$$
$$p + q - 2 > 0, \quad 1 \le p, q \le M.$$

Proof. Note that $\hat{\mathbf{R}}_{xy}$ is estimated from the L collected snapshots. Denote $\hat{R}_{p,q}$ as the (p, q)th element of $\hat{\mathbf{R}}_{xy}$. $x_p(t)$, $y_q(t)$, $s_i(t)$, $n_{x(p)}(t)$, and $n_{y(q)}(t)$ are similarly defined. We have $\forall p + q - 2 > 0$,

$$\hat{R}_{p,q} = \frac{1}{L} \sum_{t=1}^L x_p(t) y_q^*(t) = \frac{1}{L} \sum_{t=1}^L \left(\sum_{i=1}^K s_i(t) e^{-j2\pi(p-1)f_{i,1}} + n_{x(p)}(t) \right)$$

$$\cdot \left(\sum_{j=1}^K s_j^*(t) e^{j2\pi(q-1)f_{j,2}} + n_{y(q)}^*(t) \right)$$

$$= \sum_{k=1}^K \left(\frac{1}{L} \sum_{t=1}^L |s_k(t)|^2 \right) e^{-j2\pi(p-1)f_{k,1}} e^{j2\pi(q-1)f_{k,2}}$$

$$+ \frac{1}{L} \sum_{t=1}^L \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K s_i(t) s_j^*(t) e^{-j2\pi(p-1)f_{i,1}} e^{j2\pi(q-1)f_{j,2}}$$

$$+ \frac{1}{L} \left[\sum_{t=1}^L \left(\sum_{i=1}^K s_i(t) n_{y(q)}^*(t) e^{-j2\pi(p-1)f_{i,1}} + \sum_{j=1}^K s_j^*(t) n_{x(p)}(t) e^{j2\pi(q-1)f_{j,2}} \right) \right]$$

$$+ \frac{1}{L} \sum_{t=1}^L n_{x(p)}(t) n_{y(q)}^*(t) = \hat{R}_{p,q}^0 + \hat{R}_{p,q}^1 + \hat{R}_{p,q}^2 + \hat{R}_{p,q}^3,$$

$$\tag{18}$$

where $\widehat{\mathbf{R}}_{p,q}^0$, $\widehat{\mathbf{R}}_{p,q}^1$, $\widehat{\mathbf{R}}_{p,q}^2$, and $\widehat{\mathbf{R}}_{p,q}^3$ denote the first, the second, the third, and the fourth summand. Note that $\mathbb{E}(\widehat{\mathbf{R}}_{p,q}^0) = \sum_{i=1}^K r_i$ $e^{-j2\pi(p-1)f_{i,1}}e^{j2\pi(q-1)f_{i,2}} = \mathbf{R}_c(p,q)$, where $\mathbf{R}_c(p,q)$ is the (p,q) th element of $\mathbf{R}_c$. Hence, the (p,q)th estimation error $\varepsilon_{p,q}$ can be appropriately estimated as $\varepsilon_{p,q} = \widehat{\mathbf{R}}_{p,q}^1 + \widehat{\mathbf{R}}_{p,q}^2 + \widehat{\mathbf{R}}_{p,q}^3$. Since L is sufficiently large, $\varepsilon_{p,q}$ is approximately circular complex Gaussian distributed according to the central limit theorem [28]. Moreover, as the incident signals and the additive noise are mutually independent, we can easily obtain that the expectation of $\varepsilon_{p,q}$ is

$$\mathbb{E}\left(\varepsilon_{p,q}\right) = \mathbb{E}\left(\widehat{\mathbf{R}}_{p,q}^1\right) + \mathbb{E}\left(\widehat{\mathbf{R}}_{p,q}^2\right) + \mathbb{E}\left(\widehat{\mathbf{R}}_{p,q}^3\right) = \mathbf{0}, \quad (19)$$

and the variance of $\varepsilon_{p,q}$ is

$$\mathrm{Var}\left(\varepsilon_{p,q}\right) = \mathbb{E}\left(\left|\mathbf{R}\wedge_{p,q}^1\right|^2\right) + \mathbb{E}\left(\left|\mathbf{R}\wedge_{p,q}^2\right|^2\right) + \mathbb{E}\left(\left|\mathbf{R}\wedge_{p,q}^3\right|^2\right), \quad (20)$$

where

$$\mathbb{E}\left(\left|\mathbf{R}\wedge_{p,q}^1\right|^2\right) = \frac{1}{L}\sum_{i=1}^K \sum_{\substack{j=1 \\ j\neq i}}^K r_i r_j = \frac{1}{L}\left(\left(\sum_{i=1}^K r_i\right)^2 - \sum_{i=1}^K r_i^2\right),$$

$$\mathbb{E}\left(\left|\mathbf{R}\wedge_{p,q}^2\right|^2\right) = \frac{1}{L}\left(\sum_{i=1}^K r_i\sigma^2 + \sum_{j=1}^K r_j\sigma^2\right) = \frac{2}{L}\sum_{i=1}^K r_i\sigma^2,$$

$$\mathbb{E}\left(\left|\mathbf{R}\wedge_{p,q}^3\right|^2\right) = \frac{1}{L}\sigma^4. \quad (21)$$

Substituting (21) into (20), we can directly have (17).

Denote vec($\mathbf{E}$) as the vectorized estimation error matrix, and note that the distribution property of $\varepsilon_{1,1}$ is not defined in Proposition 1; then, one has

$$E\left[\|\mathbf{J}\mathrm{vec}(\mathbf{E})\|_2^2\right] = \sum_{p=1}^M \sum_{q=1,p+q-2>0}^M E\left(\left|\varepsilon_{p,q}\right|^2\right) = (M^2-1)\mathrm{Var}(\varepsilon_{2,1}), \quad (22)$$

where $\mathrm{Var}(\varepsilon_{2,1})$ is obtained by (20) and $\mathbf{J} = [\mathbf{0}_{M^2}, \mathbf{I}_{M^2-1}]$ is the selection matrix by which $E(|\varepsilon_{1,1}|^2)$ is dropped from $\mathbf{E}$. Hence, according to Proposition 1 and (22), we can define the function $\mathscr{F}\{\widehat{\mathbf{R}}_{xy}, \mathbf{R}_c\}$ as

$$\mathscr{F}\{\widehat{\mathbf{R}}_{xy}, \mathbf{R}_c\} = \left\|\mathbf{J}\mathrm{vec}\left(\widehat{\mathbf{R}}_{xy} - \mathbf{R}_c\right)\right\|_2^2, \quad (23)$$

and then, β can be easily set as

$$\beta = \mu\left\{(M^2-1)\mathrm{Var}(\varepsilon_{2,1})\right\}, \quad (24)$$

where μ is a user-specific weighting factor permitting (24) to be held in a high probability. It is worth noting that the threshold β is influenced by various factors, such as the variances of signals and noise, limited snapshots, and array geometry. In contrast, μ just introduces a scale to the threshold. Thus, μ has much smaller dynamic than β and is easier to choose, which is verified in the Numerical Results. In the following, an approximation of $\mathrm{Var}(\varepsilon_{2,1})$ is given.

Firstly, $\sum_{i=1}^K r_i + \sigma^2$ can be well estimated by averaging the diagonal elements of the observed covariance matrix along the x and y axes. It is worth noting that since the first diagonal element is common among the covariance matrices, it should be calculated only once. Hence, $\sum_{i=1}^K r_i + \sigma^2$ can be estimated by

$$\frac{1}{(2M-1)}\left\{\left[\mathrm{diag}\left(\mathbf{R}\wedge_x\right) + \mathrm{diag}\left(\mathbf{R}\wedge_y\right)\right]^T \mathbf{1}_M - \widehat{R}_y(1,1)\right\}, \quad (25)$$

where $\widehat{\mathbf{R}}_x$ and $\widehat{\mathbf{R}}_y$ denote the sample covariance matrix along the x and y axes estimated from the L collected snapshots, respectively. In practical implementation, the diagonal elements of $\widehat{\mathbf{R}}_x$ can be calculated as $\mathrm{diag}(\widehat{\mathbf{R}}_x) = \mathrm{diag}(1/L\sum_{t=1}^L$ $\mathrm{diag}(\mathbf{x}(t))\mathrm{diag}^*(\mathbf{x}(t)))$, as well as $\mathrm{diag}(\widehat{\mathbf{R}}_y)$. $\widehat{R}_y(1,1)$ is the $(1,1)$th element of $\widehat{\mathbf{R}}_y$. Next, according to (4), the (p,q)th $(p+q-2>0)$ element of $\mathbf{R}_{xy}$ as $R_{p,q}$ can be expressed as $R_{p,q} = \sum_{i=1}^K r_i e^{j2\pi(p-1)f_{i,1}} e^{-j2\pi(q-1)f_{i,2}}$. Thus, the expectation of the squared modulus of $R_{p,q}$ is

$$\mathbb{E}\left\{|R_{p,q}|^2\right\} = \sum_{i=1}^K r_i^2 + \sum_{i=1}^K \sum_{\substack{i'=1 \\ i'\neq i}}^K r_i r_i' \times \cos\left(2\pi(p-1)(f_{i,1} - f_{i'1})\right.$$

$$\left. + 2\pi(q-1)(f_{i',2} - f_{i,2})\right). \quad (26)$$

The term $\cos(2\pi(p-1)(f_{i,1} - f_{i'1}) + 2\pi(q-1)(f_{i',2} - f_{i,2}))$ varies for different p, q, i, i'. Moreover, the phase mod $(2\pi(p-1)(f_{i,1} - f_{i'1}) + 2\pi(q-1)(f_{i',2} - f_{i,2}), 2\pi)$ is uniformly distributed in $[-\pi, \pi)$ with respect to uniform distributions of $f_{i,1}$ and $f_{i,2}$ within $[-0.5, 0.5)$. Thus, the cosine term has a mean of zero. Further, if we take the average of the module of $\widehat{R}_{p,q}, (p+q-2>0)$, the cosine term will be eliminated and only the first term is retained. Hence, we can approximate $\sum_{i=1}^K r_i^2$ with the average of the module of $\widehat{R}_{p,q}, (p+q-2>0)$ as

$$\sum_{i=1}^K r_i^2 \approx \frac{\left\|\mathbf{J}\mathrm{vec}(\widehat{\mathbf{R}}_{xy})\right\|_2^2}{(M^2-1)}. \quad (27)$$

Finally, substituting (25) and (27) into (17), the approximation of $\mathrm{Var}(\varepsilon_{2,1})$ is obtained, and then, β can be determined via (24). With the obtained β and the function $\mathscr{F}\{\widehat{\mathbf{R}}_{xy}, \mathbf{R}_c\}$ in (26), we have the proposed CCMR-

based DANM (CCMR-DANM) method for L-shaped array as follows:

$$\left\{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \tilde{\mathbf{R}}_c\right\} = \arg \min_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{R}_c} \frac{1}{2M} \left(\mathrm{Tr}(\mathrm{T}(\mathbf{u}_1)) + \mathrm{Tr}(\mathrm{T}(\mathbf{u}_2))\right)$$

$$\text{s.t.} \quad \left\|\mathbf{J}\mathrm{vec}\left(\widehat{\mathbf{R}}_{xy} - \mathbf{R}_c\right)\right\|_2^2 \leq \beta$$

$$\mathrm{T}(\mathbf{R}_c(:,1)) \succeq \mathbf{0}, \mathrm{T}\left(\mathbf{R}_c(1,:)^H\right) \succeq \mathbf{0}$$

$$\begin{bmatrix} \mathrm{T}(\mathbf{u}_1) & \mathbf{R}_c \\ \mathbf{R}_c{}^H & \mathrm{T}(\mathbf{u}_2) \end{bmatrix} \succeq \mathbf{0}.$$

$$(28)$$

Compared with the SDP formulation of standard DANM in (9), beside the error tolerance constraint, an extra constraint in (14) is included in the proposed SDP formulation. It is an effectiveness relaxation of the positive constraint $\mathbf{r} \geq 0$, which is introduced by the prior information of the crosscorrelation matrix. Moreover, the importance of this extra constraint is verified in simulations.

3.5. DOA Retrieval. By solving (28), we note that besides the estimation of $\mathbf{R}_c$, the estimation of $\mathrm{T}(\mathbf{u}_1)$ and $\mathrm{T}(\mathbf{u}_1)$ also are obtained, which can be expressed in the form of an augmented matrix as

$$\mathbf{R}_\mathrm{A} = \begin{bmatrix} \mathrm{T}(\mathbf{u}_1) & \mathbf{R}_c \\ \mathbf{R}_c{}^H & \mathrm{T}(\mathbf{u}_2) \end{bmatrix}. \qquad (29)$$

Next, we develop three kinds of angle estimation methods based on $\mathrm{T}(\mathbf{u}_1)$ and $\mathrm{T}(\mathbf{u}_2)$, $\mathbf{R}_c$, and $\mathbf{R}_\mathrm{A}$, respectively.

First, since $\mathrm{T}(\mathbf{u}_1)$ and $\mathrm{T}(\mathbf{u}_2)$ contain the unknown angle information, the Vandermonde decomposition-based methods such as MUSIC and ESPRIT [17, 29, 30] can be employed for angle estimation in each dimension. Of course, a pairing operation needs to be done to finally obtain the angle pairs [18].

Second, considering $\mathbf{R}_c$ as the noise-free crosscorrelation matrix, conventional 2D DOA estimation algorithms based on crosscorrelation [18–15] can be grafted with the proposed CCMR-DANM method for angle estimation. In this paper, we adopt the JSVD algorithm [10] as a representation for comparison in simulations.

Last, according to the decoupled atomic norm theory [18, 25], we have

$$\mathbf{R}_\mathrm{A} = \begin{bmatrix} \mathrm{T}(\mathbf{u}_1) & \mathbf{R}_c \\ \mathbf{R}_c{}^H & \mathrm{T}(\mathbf{u}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_y \end{bmatrix} \frac{1}{\sqrt{M}} \mathbf{R}_s \begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_y \end{bmatrix}^H = \mathbf{H}\mathbf{R}_s'\mathbf{H}^H,$$

$$(30)$$

where $\mathbf{H} = [\mathbf{A}_x^T, \mathbf{A}_y^T]^T$ and $\mathbf{R}_s' = (1/\sqrt{M})\mathbf{R}_s$. Hence, the conventional covariance-based 2D DOA estimation algorithms, such as [8, 9], can be incorporated into the proposed CCMR-DANM method.

Remark 2. Note that all the matrices we obtained through (28) are low rank and their rank values are equal to the source numbers. In the three kinds of proposed angle estimation methods, either eigenvalue decomposition (EVD) or singular value decomposition (SVD) is taken on these low-rank matrices. Hence, we can determine the source numbers by counting the number of eigenvalues or singular values, which is larger than a predefined threshold, e.g., $0.05\lambda_{\max}$, where $\lambda_{\max}$ is the maximum value of the eigenvalues. In this sense, the proposed methods can be done in a blind mode without knowing the source numbers *a priori*.

Remark 3. It is worth noting that the three kinds of proposed angle estimation methods have different behaviors in terms of estimation accuracy and computational complexity. The first Vandermonde decomposition-based method is computationally efficient but requires an extra pairing operation. The third $\mathbf{R}_\mathrm{A}$-based method is relatively computational expensive than the other two. The second crosscorrelation-based method is not only computationally efficient but also can achieve automatic pairing. Moreover, note that although the proposed method works on sample crosscorrelation matrix, the proposed method grafting with the $\mathbf{R}_\mathrm{A}$-based method can detect $2M - 1$ sources, which is larger than that of the conventional crosscorrelation-based DOA estimation methods.

In summary, the proposed CCMR-DANM method is widely applicable in practical implementations.

4. Discussions

Two related issues regarding the proposed solution in Section 3 are discussed in this section. First, the computational complexity of the proposed method is analyzed and compared with that of existing CCMR-based methods. Second, we extend the proposed method to the sparse L-shaped array cases.

4.1. Computational Complexity. Note that the CC-ANM, CC-GLS, and the proposed method are all two-phase CCMR-based methods. They have common operations in sample crosscorrelation matrix construction and DOA estimation. The only difference is the technique used for CCMR. Hence, we only compare the computational complexities of the CCMR operation in these three methods. According to Vandenberghe and Boyd [31], the computational complexity for solving the SDP formulation of (28) is $\mathcal{O}\{(M^2 + 2M)^2 (4M)^{2.5}\}$. Moreover, the computational complexities of CCMR in the proposed CCMR-DANM, CC-ANM, and CC-GLS are listed in Table 1.

Further, Figure 2 presents the computational complexities of CCMR in different algorithms versus the number of antennas at each subarray. Obviously, the proposed CCMR-DANM has the least computational cost compared with CC-ANM and CC-GLS when $M > 4$. Further, the computational gaps between the proposed and the other two algorithms become large with the growth of the number of

TABLE 1: The computational complexities of CCMR in different algorithms.

Algorithm	Computational complexity
CC-ANM	$\mathcal{O}\left\{\left(3M^2 - 2M\right)^2 M^{4.5}\right\}$
CC-GLS	$\mathcal{O}\left\{\left(2M^2 - 2M\right)^2 M^{4.5}\right\}$
Proposed	$\mathcal{O}\left\{\left(M^2 + 2M\right)^2 (4M)^{2.5}\right\}$

antennas. Hence, the proposed method offers us a more efficient choice for CCMR, especially with large M.

4.2. Extension to Sparse L-Shaped Array Cases. In the sparse L-shaped array case, we consider there are two sparse linear array (SLA) arranged along the x and y axes. Moreover, each SLA is a subarray of a virtual ULA. In other words, the manifolds of the SLAs termed $\mathbf{A}_x{}'$ and $\mathbf{A}_y{}'$ can be expressed as [22]

$$\begin{aligned} \mathbf{A}_{x'} &= \mathbf{I}_{\Omega_x} \mathbf{A}_x, \\ \mathbf{A}_{y'} &= \mathbf{I}_{\Omega_y} \mathbf{A}_y, \end{aligned} \tag{31}$$

where $\mathbf{I}_{\Omega_x}$ is the selection matrix for the SLA in the axis with Ω_x being the index set indicating the index of remaining antennas and $\mathbf{I}_{\Omega_y}$ is similarly defined. Hence, we have

$$\mathbf{R}_{x'y'} = \mathbf{I}_{\Omega_x}(\mathbf{R}_c + \mathbf{N}_1)\mathbf{I}_{\Omega_y}^H, \tag{32}$$

and sequentially, the error estimation matrix can be defined as

$$\mathbf{E}' = \widehat{\mathbf{R}}_{x'y'} - \mathbf{I}_{\Omega_x}\mathbf{R}_c\mathbf{I}_{\Omega_y}. \tag{33}$$

Moreover, in this paper, we consider that the two SLAs are two redundancy linear arrays [22], which means the virtual crosscorrelation as $\mathbf{R}_{xy}$ can be fully determined by $\mathbf{R}_{x'y'}$. Further, each element in $\mathbf{E}'$ has similar distribution as that in $\mathbf{E}$. In other words, we can similarly defined the function $\mathscr{F}\{\widehat{\mathbf{R}}_{x'y'}, \mathbf{I}_{\Omega_x}\mathbf{R}_c\mathbf{I}_{\Omega_y}\}$ and the tolerant bound β' as those in (23) and (24), respectively. Then, we have the proposed CCMR-DANM for sparse L-shaped array as

$$\left\{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \tilde{\mathbf{R}}_c\right\} = \min_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{R}_c} \frac{1}{2M}\left(\mathrm{Tr}(\mathbf{T}(\mathbf{u}_1)) + \mathrm{Tr}(\mathbf{T}(\mathbf{u}_2))\right)$$

$$\text{s.t.} \quad \left\|\mathbf{J}'\left(\mathrm{vec}\left(\widehat{\mathbf{R}}_{x'y'}\right) - \left(\mathbf{I}_{\Omega_y} \otimes \mathbf{I}_{\Omega_x}\right)\mathrm{vec}(\mathbf{R}_c)\right)\right\|_2^2 \le \beta'$$

$$\mathbf{T}(\mathbf{R}_c(:,1)) \succcurlyeq 0, \, \mathbf{T}\left(\mathbf{R}_c(1,:)^H\right) \succcurlyeq 0$$

$$\begin{bmatrix} \mathbf{T}(\mathbf{u}_1) & \mathbf{R}_c \\ \mathbf{R}_c{}^H & \mathbf{T}(\mathbf{u}_2) \end{bmatrix} \succcurlyeq 0, \tag{34}$$

where $\mathbf{J}'$ is similarly defined as $\mathbf{J}$.

5. Numerical Results

In this section, we present numerical examples to verify the effectiveness of the proposed method and evaluate the performance of the proposed method for L-shaped array DOA estimation. Unless specifically stated, in simulations, we consider three source signals, i.e., $K = 3$, with DOAs (15°, 10°), (25°, 20°), and (35°, 30°), impinge onto the L-shaped array with $M = 8$. The number of collected snapshots L and Monte Carlo trials N is set to 200 and 500, respectively. We use root mean square error (RMSE) to evaluate DOA estimation precision of the proposed method with comparison to the MUSIC algorithm for L-shaped array (replaced with MUSIC hereafter), the JSVD [10], the CESA [12], the CC-GLS [22], and the CRB [9]. We omit the CC-ANM for comparison since it has similar estimation performance as CC-GLS and it is more computationally expensive [22]. The RMSE of the azimuth and elevation angle estimations is defined as

$$\mathrm{RMSE} = \frac{1}{K}\sum_{i=1}^{K}\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[\left(\phi\wedge_{i,n} - \phi_i\right)^2 + (\theta\wedge_{i,n} - \theta_i)^2\right]}, \tag{35}$$

where $\widehat{\phi}_{i,n}$ and $\widehat{\theta}_{i,n}$ denote estimates of ϕ_i and θ_i from the nth trial, respectively.

First, to guarantee the proposed CCMR-DANM method works in a best status, it is important to choose an appropriate user-specific weighting factor μ. The RMSEs of the proposed method with different μ versus SNR are presented in Figure 3. The curves in Figure 3 indicate that with different μ, the resulting proposed methods have different estimation performance. Moreover, the proposed method equipped with $\mu = 1$ has the best performance. Hence, in following simulations, we set $\mu = 1$ to implement the proposed CCMR-DANM method.

Next, let us evaluate the estimation performance of the proposed CCMR-DANM method compared with the aforementioned methods. Moreover, the proposed method without the constraint (14), termed proposed w/o 2, is also simulated for comparison. Figure 4 shows the RMSEs of these methods versus SNR. As shown in Figure 4, the proposed method is effective and outperforms both the conventional L-shaped DOA estimation methods and the existing gridless CCMR-based methods. Moreover, there is a nonnegligible gap between the proposed method and the proposed w/o 2 method. It indicates that the performance of the proposed method degrades greatly when the constraint in (14) is dropped, which means the constraint in (14) is indispensable for the proposed method.

To further examine the performance, we present the RMSEs of aforementioned methods versus the number of snapshots with SNR = 0 dB in Figure 5. The result likewise indicates that the proposed method has better estimation accuracy than the conventional methods. Combining the estimation performance comparisons with the analysis in Remark 2, we can conclude that the proposed method not only has the better estimation performance compared with
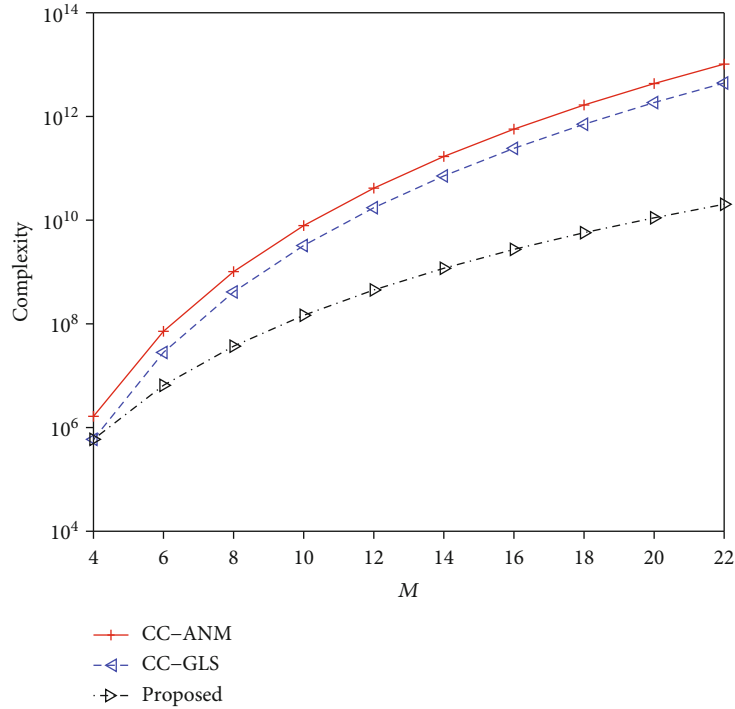
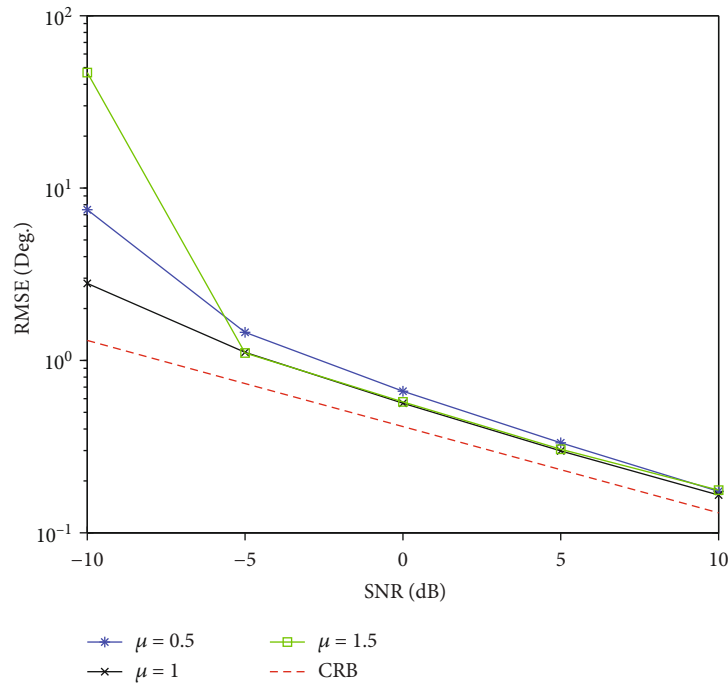FIGURE 2: The computational complexities of CCMR in different algorithms versus M.

FIGURE 3: RMSE of the proposed method with different μ versus SNR.

the conventional L-shaped DOA estimation methods but also can be implemented without knowing the source numbers. Moreover, combining the estimation performance comparisons with the analysis of computational complexity in Section 4.1, we can conclude that the proposed method not only is more efficient but also has better estimation performance, compared with the existing gridless CCMR-based methods, such as CC-GLS and CC-ANM.

Last, let us consider the sparse L-shaped array cases. Assume both of the SLAs are sampled from the 8 element
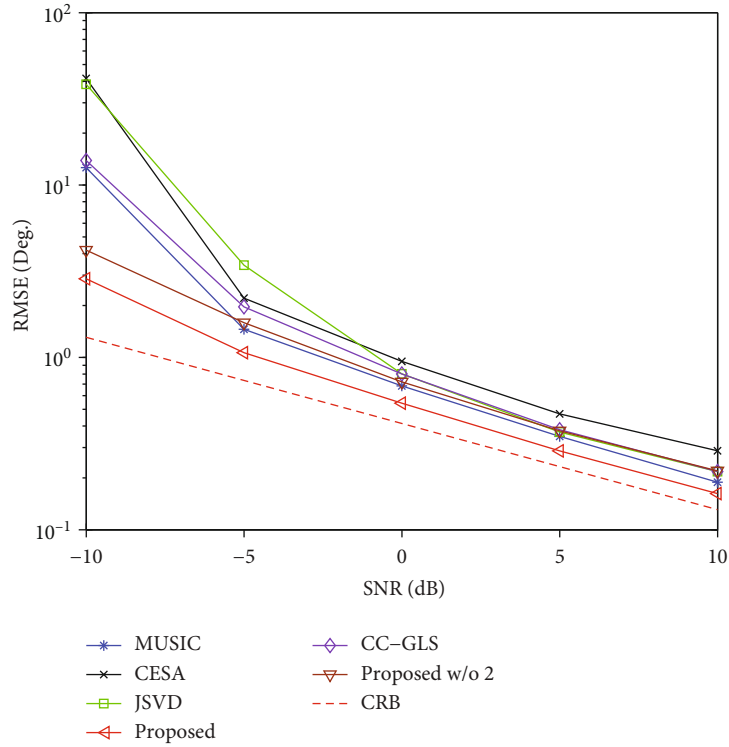
FIGURE 4: RMSE versus SNR with $L = 300$.
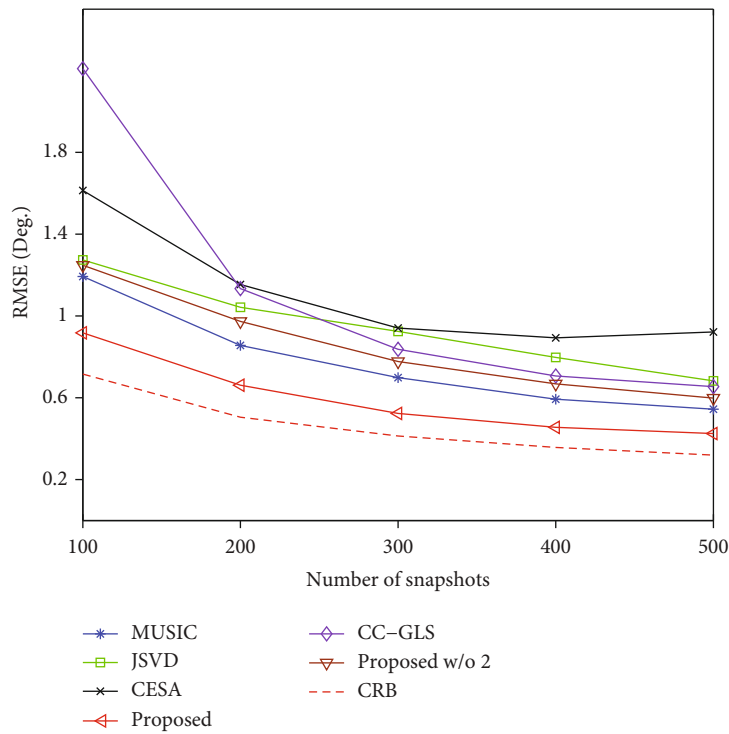
FIGURE 5: RMSE versus the number of snapshots with SNR = 0 dB.

ULA with the selection matrix $\mathbf{I}_\Omega$, where $\Omega = \{1, 2, 4, 7, 8\}$. And the same three source signals with DOAs (15°, 10°), (25°, 20°), and (35°, 30°) impinge onto the sparse L-shaped array. As shown in Figure 6, the 2D DOAs can be clearly observed via the proposed method. In other words, the proposed method is applicable to the sparse L-shaped array cases.
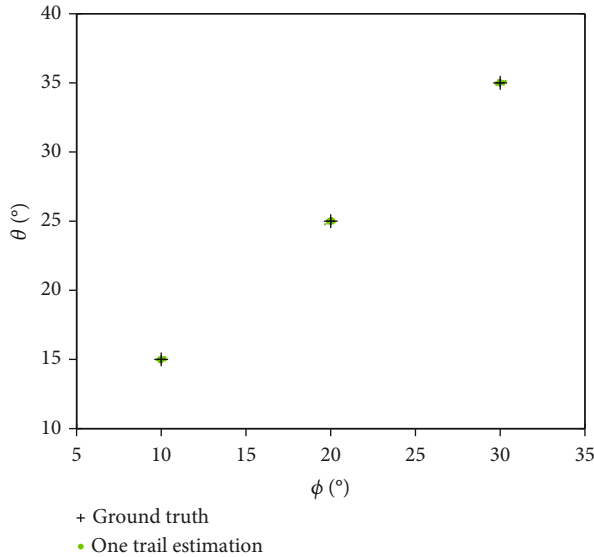
Figure 6: Angle estimation via the proposed CCMR-DANM in a sparse L-shaped array case with $L = 200$, SNR = 5 dB.

6. Conclusion

In this paper, a new CCMR-based two-phase method for 2D DOA estimation with L-shaped array via DANM is proposed. In the first phase, the DANM technique simultaneously considering the noise and finite snapshot effects and structure property is employed to reconstruct the noise-free crosscorrelation matrix from the sample counterpart. Then, in the second phase, conventional 2D DOA estimation methods for L-shaped array can be adopted for 2D-DOA estimation. Numerical simulations demonstrate the effectiveness and outperformance of the proposed method.

Data Availability

All data sources are reliable and the corresponding author can be contacted.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L. Wan, X. Kong, and F. Xia, "Joint range-doppler-angle estimation for intelligent tracking of moving aerial targets," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1625–1636, 2018.

[2] X. Wang, L. Wang, X. Li, and G. Bi, "Nuclear norm minimization framework for DOA estimation in mimo radar," *Signal Processing*, vol. 135, pp. 147–152, 2017.

[3] Y.-Y. Dong and X. Chang, "Computationally efficient 2D DOA estimation for L-shaped array with unknown mutual coupling," *Mathematical Problems in Engineering*, vol. 2018, Article ID 5454719, 9 pages, 2018.

[4] M.-M. Liu, C.-X. Dong, Y.-Y. Dong, and G.-Q. Zhao, "Super-resolution 2D DOA estimation for a rectangular array via reweighted decoupled atomic norm minimization," *Mathematical Problems in Engineering*, vol. 2019, Article ID 6797168, 13 pages, 2019.

[5] T. Liu, F. Wen, L. Zhang, and K. Wang, "Off-grid DOA estimation for colocated MIMO radar via reduced-complexity sparse Bayesian learning," *IEEE Access*, vol. 7, pp. 99907–99916, 2019.

[6] F. Wen and J. Shi, "Fast direction finding for bistatic EMVS-MIMO radar without pairing," *Signal Processing*, vol. 173, article 107512, 2020.

[7] J. Shi, F. Wen, and T. Liu, "Nested MIMO radar: coarrays, tensor modeling and angle estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 1, pp. 573–585, 2021.

[8] J. Li and R. T. Compton, "Two-dimensional angle and polarization estimation using the ESPRIT algorithm," *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 5, pp. 550–555, 1992.

[9] J. Li and D. Jiang, "Joint elevation and azimuth angles estimation for L-shaped array," *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 453–456, 2017.

[10] J.-F. Gu and P. Wei, "Joint SVD of two cross-correlation matrices to achieve automatic pairing in 2-D angle estimation problems," *IEEE Antennas and Wireless Propagation Letters*, vol. 6, pp. 553–556, 2007.

[11] G. Wang, J. Xin, N. Zheng, and A. Sano, "Computationally efficient subspace-based method for two-dimensional direction estimation with L-shaped array," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3197–3212, 2011.

[12] N. Xi and L. Liping, "A computationally efficient subspace algorithm for 2-D DOA estimation with L-shaped array," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 971–974, 2014.

[13] N. Tayem, K. Majeed, and A. A. Hussain, "Two-dimensional DOA estimation using crosscorrelation matrix with L-shaped array," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 1077–1080, 2016.

[14] Y.-Y. Dong, C.-x. Dong, J. Xu, and G.-q. Zhao, "Computationally efficient 2-D DOA estimation for L-shaped array with automatic pairing," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 1669–1672, 2016.

[15] Y.-Y. Dong, C.-x. Dong, W. Liu, H. Chen, and G.-q. Zhao, "2-D DOA estimation for L-shaped array with array aperture and snapshots extension techniques," *IEEE Signal Processing Letters*, vol. 24, no. 4, pp. 495–499, 2017.

[16] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[17] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[18] Z. Zhang, Y. Wang, and Z. Tian, "Efficient two-dimensional line spectrum estimation based on decoupled atomic norm minimization," *Signal Processing*, vol. 163, pp. 95–106, 2019.

[19] Y. Zhang, G. Zhang, and X. Wang, "Array covariance matrix-based atomic norm minimization for off-grid coherent direction-of-arrival estimation," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3196–3200, New Orleans, LA, USA, 2017.

[20] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1030–1042, 2015.

[21] X. Wu, W.-P. Zhu, and J. Yan, "Gridless two-dimensional DOA estimation with L-shaped array based on the cross-covariance matrix," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3256–3260, Calgary, AB, Canada, 2018.

[22] X. Wu, W.-P. Zhu, and J. Yan, "An efficient gridless 2-D DOA estimation method for sparse and uniform L-shaped arrays," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, Cannes, France, 2019.

[23] P. S. Hongbin Li and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1314–1323, 1999.

[24] Y. Zhang, G. Zhang, and X. Wang, "Cross-correlation matrix-based atomic norm minimization for L-shaped array DOA estimation," in *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–4, Xiamen, China, 2017.

[25] Z. Tian, Z. Zhang, and Y. Wang, "Low-complexity optimization for two-dimensional direction-of-arrival estimation via decoupled atomic norm minimization," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3071–3075, New Orleans, LA, USA, 2017.

[26] Y. Zhang, G. Zhang, Y. Kong, and F. Wen, "Gridless sparsity-based DOA estimation for sparse linear array," *The Journal of Engineering*, vol. 2019, no. 20, pp. 6629–6632, 2019.

[27] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014, http://cvxr.com/cvx.

[28] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, John Wiley & Sons, 2007.

[29] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-RAO bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, 1989.

[30] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound: further results and comparisons," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2140–2150, 1990.

[31] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, 1996.

Research Article

Lora RTT Ranging Characterization and Indoor Positioning System

Qiang Liu,[1] XiuJun Bai,[2] Xingli Gan ,[3] and Shan Yang[2]

[1]*CHN Energy, Huanghua Port Affairs Co. Ltd, Cangzhou, Hebei, China*
[2]*China Unicom Smart City Research Institute, Baoding, Hebei, China*
[3]*Zhejiang University of Science and Technology, Hangzhou, Zhejiang, China*

Correspondence should be addressed to Xingli Gan; ganxingli@163.com

In recent years, indoor positioning systems (IPS) are increasingly very important for a smart factory, and the Lora positioning system based on round-trip time (RTT) has been developed. This paper introduces the ranging characterization, RTT measurement, and position estimation method. In particular, a particle filter localization method-aided Lora pseudorange fitting correction is designed to solve the problem of indoor positioning; the cumulative distribution function (CDF) criteria are used to measure the quality of the estimated location in comparison to the ground truth location; when the positioning error on the x-axis threshold is 0.2 m and 0.6 m, the CDF with pseudorange correction is 61% and 99%, which are higher than the 32% and 85% without pseudorange correction. When the positioning error on the y-axis threshold is 0.2 m and 0.6 m, the CDF with pseudorange correction is 71% and 99.9%, which are higher than the 52% and 94.8% without pseudorange correction.

1. Introduction

Indoor positioning systems are increasingly very important for a smart factory, such as finding the location of workers, goods, or vehicles [1–4]. However, the Global Positioning System (GPS) is unable to provide the indoor positioning service, which is commonly used for outdoor positioning. The research on indoor positioning technologies has been conducted for more than three decades, such as Radio Frequency Identification (RFID) [5, 6], WiFi [7, 8], Ultra Wide Band (UWB) [9–11], Pseudolite [12–14], Bluetooth [15], and Inertial Navigation System (INS), but they may not be suitable for Internet of Things (IoT) applications in terms of cost, application mode, and terminal power consumption.

With the continuous progress of sensor and Internet of Things [16, 17] technologies, increasing attention has been paid to IPS using Lora WAN. Semtech has developed a Lora positioning system based on round-trip time [18], which is called the SX1280 transceiver. The SX1280 transceiver family

provides ultra-long-range communication in the 2.4 GHz band with a time-of-flight functionality; its radio is fully compliant with all worldwide 2.4 GHz radio regulations including EN 300440, FCC CFR 47 Part 15 [19], and the Japanese ARIB STD-T66 [20]. Very small wearable products to track and localize assets in logistic chains and people for safety can easily be designed thanks to the high level of integration and the ultralow current consumption which allows the use of miniaturized batteries.

But the most difficult challenge for Lora indoor positioning is the ranging error caused by multipath in the indoor environment. Liang et al. carry out the study focused on the indoor propagation of the Lora signal, and the main contribution of this work is to measure the round-trip time and packet delivery ratio by changing send power, payload length, and air rate in a multilevel building from the 1st floor to the 12th floor [21]. Huynh and Brennan verify the UWB transmission characteristics of an indoor RTT signal by generating synthetic received signals using ray tracing plus

Rayleigh distributed random multipath clusters as well as random amplitude and delay factors [22]. Staniec and Kowal describe outcomes of measurement campaigns during which the Lora performance was tested against a heavy multipath propagation and a controlled [23], Lora configurational space is divided into three distinct sensitivity regions: in the white region, it is immune to both interference and multipath propagation; in the light-grey region, it is only immune to the multipath phenomenon but sensitive to interference; and in the dark grey region, Lora is vulnerable to both phenomena.

On the other hand, location based on a filtering algorithm is the only effective solution currently known, and the Bayesian filtering algorithm occupies an important role. In the early days, the Kalman filter positioning algorithm is mainly used to solve the problem achieving efficient state estimation for linear Gaussian systems [24]. Moreover, to deal with unwanted errors and nonlinear distortions, particle filter (PF) is applied as a nonparametric filter to location, which is recursive implementations of Monte Carlo-based statistical processing [25–27] and performs well in localization efficiency, stability, and accuracy.

In the following sections, a Lora indoor positioning system is introduced, which overcomes the problem of long distance, low power consumption, and low cost. The main contributions and research content of this paper are as follows: Firstly, a Lora-aided particle filter localization method is designed to solve the problem of indoor positioning. Secondly, numerous experiments were carried out with real Lora RTT measurement data to evaluate the performance of the proposed approach; we used the CDF criteria to measure the quality of the estimated location in comparison to the ground truth location. The results show that the indoor positioning accuracy is improved obviously with the help of the piecewise fitting correction method. At the same time, the Lora indoor positioning system can achieve a positioning accuracy of 1 m under the condition of LOS.

2. Background and Related Work

2.1. Lora RTT Measurement. In this paper, we focus on one of the latest techniques called the RTT scheme-based ranging and localization [28–30], which can give accurate measurements by the time stamp from the initiator (Lora_A) to the responder (Lora_B) with nanosecond resolution. Figure 1 shows the RTT measurement illustration; the pseudorange measurement can be built as

$$\rho_{AB} = R_{AB} + t_{r\_A} - t_{s\_B}, \qquad (1)$$

$$\rho_{BA} = R_{BA} + t_{r\_B} - t_{s\_A}, \qquad (2)$$

where ρ_{AB} or ρ_{BA} is the pseudorange measurement between Lora_A and Lora_B, R_{AB} or R_{BA} is the geometric range between Lora_A and Lora_B, t_{r_A} is the clock offset of Lora_A at the receiving time, t_{s_A} is the clock offset of Lora_A at the transmitting time, t_{r_B} is the clock offset of Lora_B at the receiving time, and t_{s_B} is the clock offset of Lora_B at the transmitting time.
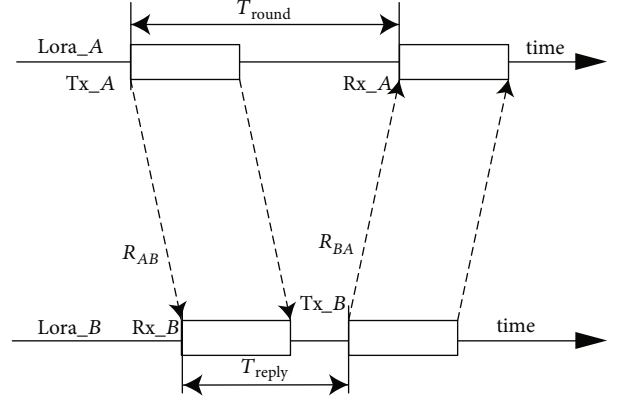

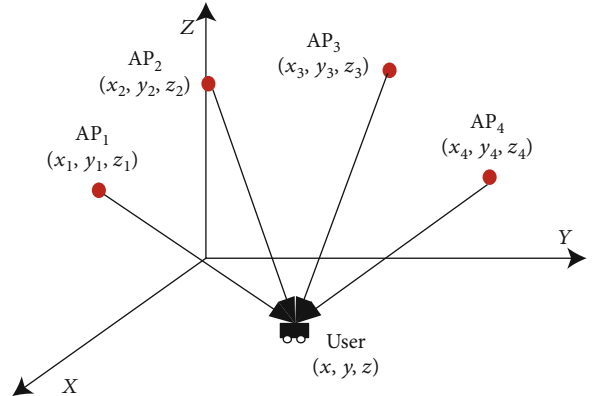
FIGURE 1: RTT measurement illustration.

FIGURE 2: Position estimation by four distance measurements.

If the clock characteristics of Lora_A and Lora_B are stable, $t_{r\_A} = t_{s\_A}$ and $t_{r\_B} = t_{s\_B}$ at the adjacent time of receiving and transmitting. The time-of-flight (ToF) measurements can be written as

$$\begin{aligned} T_{\text{round}} &= \rho_{AB} + T_{\text{reply}} + \rho_{BA} = R_{AB} + R_{BA} + t_{r\_B} \\ &\quad - t_{s\_A} + t_{r\_A} - t_{s\_B} + T_{\text{reply}} = R_{AB} + R_{BA} + T_{\text{reply}}, \end{aligned} \qquad (3)$$

where T_{round} is the time-of-flight (ToF) measurements of Lora_A from Tx_A to Rx_A, T_{reply} is the time difference of Lora_B between Rx_B and Tx_B.

The geometric distance can be calculated by

$$R_{AB} = \frac{\left(T_{\text{round}} - T_{\text{reply}} \right)}{2}. \qquad (4)$$

The pseudorange measurement can be built as

$$R_{AB} = \frac{\left(T_{\text{round}} - T_{\text{reply}} \right)}{2} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2} + \varepsilon, \qquad (5)$$

TABLE 1: Relationship between ranging error, clock error, and flight time.

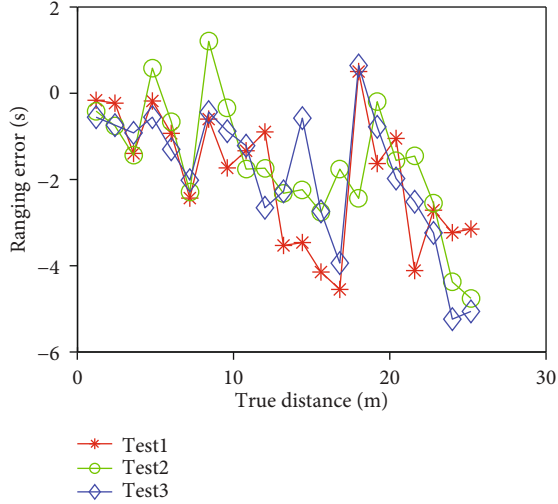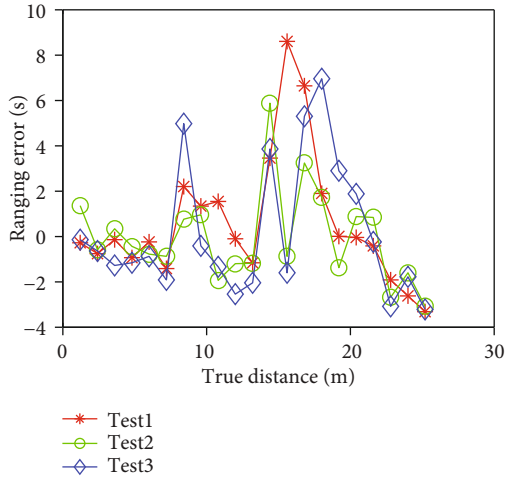| Clock error $(T_{reply} + R_{AB})$	0.1 ppm	0.5 ppm	5 ppm	25 ppm
$1.5\,\mu s$	1.5×10^{-4} ns	7.5×10^{-4} ns	7.5×10^{-3} ns	3.75×10^{-2} ns
$10\,\mu s$	1.5×10^{-3} ns	7.5×10^{-3} ns	7.5×10^{-2} ns	3.75×10^{-1} ns
$100\,\mu s$	1.5×10^{-2} ns	7.5×10^{-2} ns	7.5×10^{-1} ns	3.75 ns
$1000\,\mu s$	1.5×10^{-1} ns	7.5×10^{-1} ns	7.5 ns	37.5 ns
$10000\,\mu s$	1.5 ns	7.5 ns	75 ns	375 ns

FIGURE 3: Ranging error under LOS.

FIGURE 4: Ranging error under NLOS.

where x_A, y_A, and z_A are the transmitting antenna coordinates of Lora_A; x_B, y_B, and z_B are the transmitting antenna coordinates of Lora_B; and ε is measurement error.

2.2. Position Estimation. Four distance measurements of Lora transceivers whose positions are known are used to determine the three-dimensional coordinates of an unknown position [31], as shown in Figure 2. The position of user

(x_u, y_u, z_u) can be calculated by

$$
\begin{cases}
R_{1u} = \sqrt{(x_1 - x_u)^2 + (y_1 - y_u)^2 + (z_1 - z_u)^2} + \varepsilon_{1u}, \\
R_{2u} = \sqrt{(x_2 - x_u)^2 + (y_2 - y_u)^2 + (z_2 - z_u)^2} + \varepsilon_{2u}, \\
R_{3u} = \sqrt{(x_3 - x_u)^2 + (y_3 - y_u)^2 + (z_3 - z_u)^2} + \varepsilon_{3u}, \\
R_{4u} = \sqrt{(x_4 - x_u)^2 + (y_4 - y_u)^2 + (z_4 - z_u)^2} + \varepsilon_{4u},
\end{cases}
\tag{6}
$$

where R_{iu} is the pseudorange measurement of AP_i, i is the AP index, and u is the user index, (x_i, y_i, z_i) are the locations of AP_i, and ε_{iu} is the measurement error of AP_i.

The geometric distance can be calculated by

$$
R_{iu} = \frac{(x_u - x_i)x_u + (y_u - y_i)y_u + (z_u - z_i)z_u}{\sqrt{(x_i - x_u)^2 + (y_i - y_u)^2 + (z_i - z_u)^2}}
$$
$$
+ \varepsilon_{iu} = \begin{bmatrix} e_x^i, e_y^i, e_z^i \end{bmatrix} \begin{bmatrix} x_u \\ y_u \\ z_u \end{bmatrix} + \varepsilon_{iu},
\tag{7}
$$

where $[e_x^i, e_y^i, e_z^i]$ is called a geometry matrix.

The observation equations of Lora RTT can be expressed as the following matrix form:

$$
\begin{bmatrix} R_{1u} \\ R_{2u} \\ \vdots \\ R_{4u} \end{bmatrix} = \begin{bmatrix} e_x^1 & e_y^1 & e_z^1 \\ e_x^2 & e_y^2 & e_z^2 \\ e_x^3 & e_y^3 & e_z^3 \\ e_x^4 & e_y^4 & e_z^4 \end{bmatrix} \begin{bmatrix} x_u \\ y_u \\ z_u \end{bmatrix} + \begin{bmatrix} \varepsilon_{1u} \\ \varepsilon_{2u} \\ \vdots \\ \varepsilon_{4u} \end{bmatrix}.
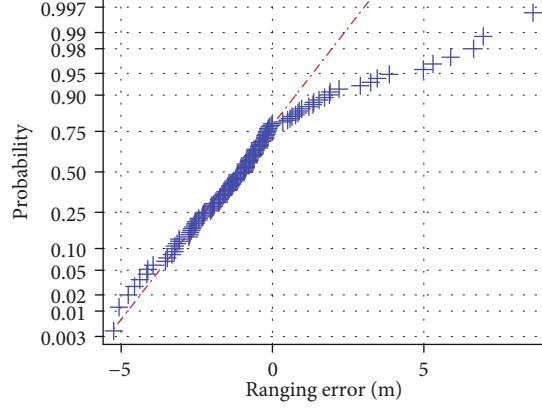\tag{8}
$$

The matrix on the right-hand side of Equation (8) is defined as A and ε, and the two-column vectors on the left-hand side are defined as b. Equation (8) can be written as

$$
b = AX + \varepsilon.
\tag{9}
$$

If an initial value X_0 is used for the solution-updating process, the Newton-Raphson method is described
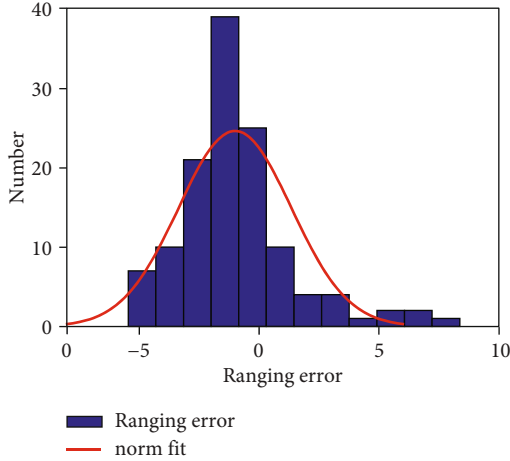
(a)

(b)

FIGURE 5: Probability characterization: (a) histogram with a distribution fit and (b) normal probability plot.
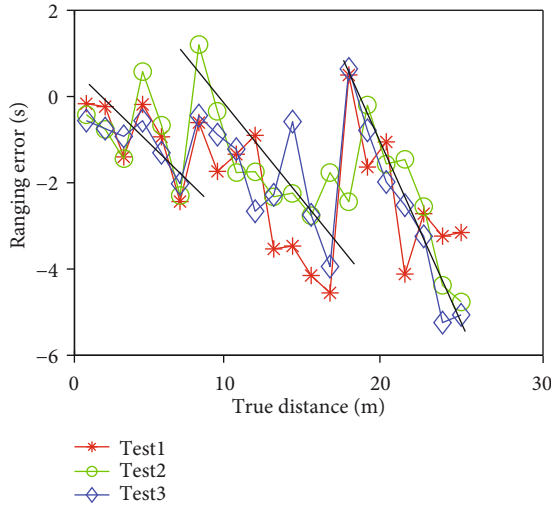

FIGURE 6: Piecewise linear fitting correction.

as $X_0 = (x_{u,0}, y_{u,0}, z_{u,0})$. The least-squares updated solution can be represented as

$$\Delta X = \left(A^T A\right)^{-1} A^T b + \varepsilon. \tag{10}$$

Then, the position X_1 can be updated iteratively according to

$$X_1 = X_0 + \Delta X. \tag{11}$$

It should be noted that the measurement error is considered Gaussian white noise, when using the least-squares method.

3. Ranging Characterization of Lora

3.1. Clock Error. It is supposed that the clock offsets of devices Lora_A and Lora_B are ε_A and ε_B; therefore, the ranging error will increase with the increase of flight time. The equa-
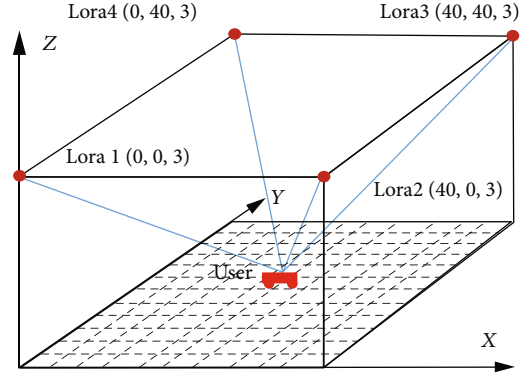

FIGURE 7: Simulation conditions of geometry distribution.

tion of ranging error is as follows:

$$\varepsilon_t = \frac{(\varepsilon_A - \varepsilon_B)}{2} \times \left(T_{\text{reply}} + R_{AB}\right), \tag{12}$$

where ε_t is ranging error caused by the clock offsets.

If the clock error of Lora_A and Lora_B is the same, Equation (12) can be written as

$$\varepsilon_t = \varepsilon_A \times \left(T_{\text{reply}} + R_{AB}\right). \tag{13}$$

It is supposed that the distance between the two Lora nodes is between 10 meters and 5000 meters; then, R_{AB} is from 3 ns to 1500 ns. The clock error of packaged crystal oscillator (PCO) is generally more than 25 ppm, the temperature compensated crystal oscillator (TCXO) is from 0.5 ppm to 5 ppm, and the oven-controlled crystal oscillator (OCXO) may be less than 0.1 ppm. According to the above parameters, we estimate the influence of clock error on Lora ranging accuracy, as shown in Table 1. It can be found that TCXO and OCXO are best used as the clock of the Lora positioning system.
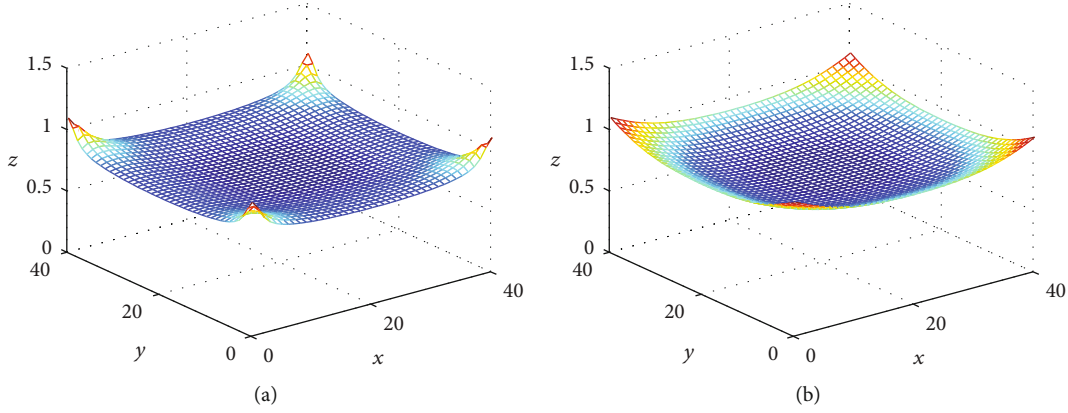
FIGURE 8: Indoor positioning system composed of four Lora transceivers: (a) HDOP and (b) GDOP in the projection area.
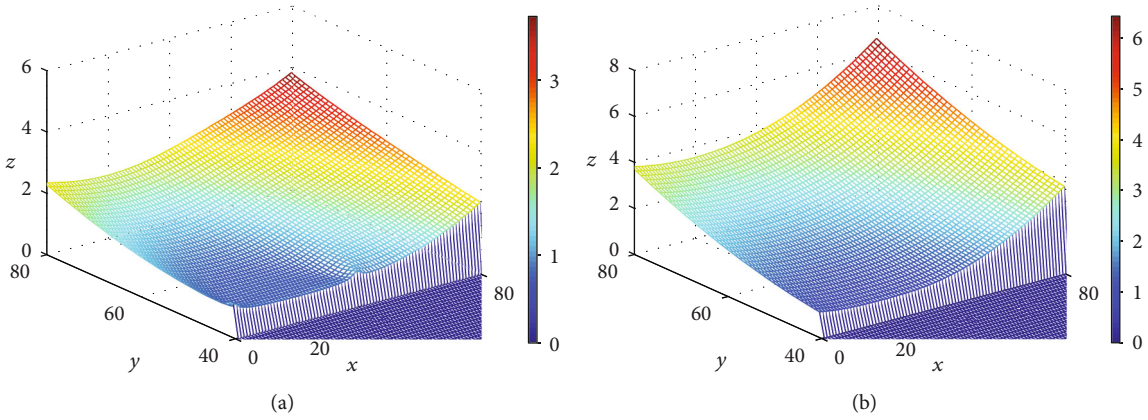

FIGURE 9: Indoor positioning system composed of four Lora transceivers: (a) HDOP and (b) GDOP outside the projection area.

3.2. Ranging Accuracy and Characterization. The reflection of the indoor environment to the Lora positioning signal is very serious, for example, irregular room structure, walking people, tables and chairs, and glass. Because of signal interference and multipath effect, Lora's ranging error may be non-Gaussian distribution. We collect the ranging data of Lora under different conditions such as line-of-sight (LOS), non-line-of-sight (NLOS), and human occlusion, which is compared with the real distance to analyze the ranging characterization.

Three sets of ranging data are collected under the LOS condition; the ranging error of Lora is shown in Figure 3. The average error is -0.61 m, the maximum error is 1.2 m, and the minimum error is -5.24 m. It can be found that the ranging error of LOS has a linear trend, which can be corrected by the polynomial fitting method.

Similarly, three sets of ranging data are collected under the NLOS condition, such as using people's bodies to block Lora's antenna; the ranging error of Lora is shown in Figure 4. The average error is -0.01 m, the maximum error is 8.6 m, and the minimum error is -3.32 m. It can be found that the ranging error of NLOS is much worse than that of LOS; therefore, it is better to use the Lora indoor positioning system under the condition of LOS.

In order to analyze the distribution characteristics of Lora ranging error, the six groups of data collected above are combined, as shown in Figure 5. Figure 5(a) shows the histogram with a distribution fit, and Figure 5(b) is the normal probability plot. It can be concluded that the probability distribution of Lora ranging error is non-Gaussian white noise. Therefore, it is necessary to use nonlinear filter to solve the location problem.

The piecewise fitting correction method of LOS is used to correct the Lora ranging value, as shown in Figure 6. The formula of the piecewise fitting correction method can be expressed as

$$\widehat{R} = R + f(R), \tag{14}$$

$$f(R) = \begin{cases} a_{10} + a_{11} \times R, & 0 \le R \le r_1, \\ a_{20} + a_{21} \times R, & r_1 < R \le r_2, \\ \quad\quad\quad\quad\quad\quad\quad \vdots \\ a_{m0} + a_{m1} \times R, & r_{m-1} < R \le r_m, \end{cases} \tag{15}$$

where R is the pseudorange measurement of Lora, $\widehat{R}$ is the corrected pseudorange, $f(R)$ is the correction function, a_{10}

TABLE 2: Particle filter algorithm.

| Process	Content
Initialization	Let $X_1^i \sim p(X_0)$, $i = 1, \cdots, N$, and $w_0^i = 1/N$
Iteration	
(1) Measurement update	For $i = 1, \cdots, N$, $w_t^i = (1/c_t)w_{t-1}^i p(Y_t/X_t^i)$, where the normalization weight is given by $c_t = \sum_{t=1}^N w_{t-1}^i p(Y_t/X_t^i)$
(2) Estimation	The filtering density is approximated by $p(X_t/Y_t) \approx \sum_{i=1}^N w_t^i \delta(X_t - X_t^i)$, $X_t \approx \sum_{i=1}^N w_t^i X_t^i$
(3) Resampling	Optionally at each time, take N samples with replacement from the set $\{X_t^i, w_t^i\}_{i=1}^N$, where the probability to take sample i is w_t^i and let $w_t^i = 1/N$
(4) Time update	Generate predictions according to the proposal distribution: $X_{t+1}^i \sim q(X_{t+1}/X_t^i, Y_{t+1})$, and compensate for the importance weight $w_{t+1}^i = w_t^i (p(X_{t+1}^i/X_t^i)/q(X_{t+1}^i/X_t^i, Y_{t+1}))$
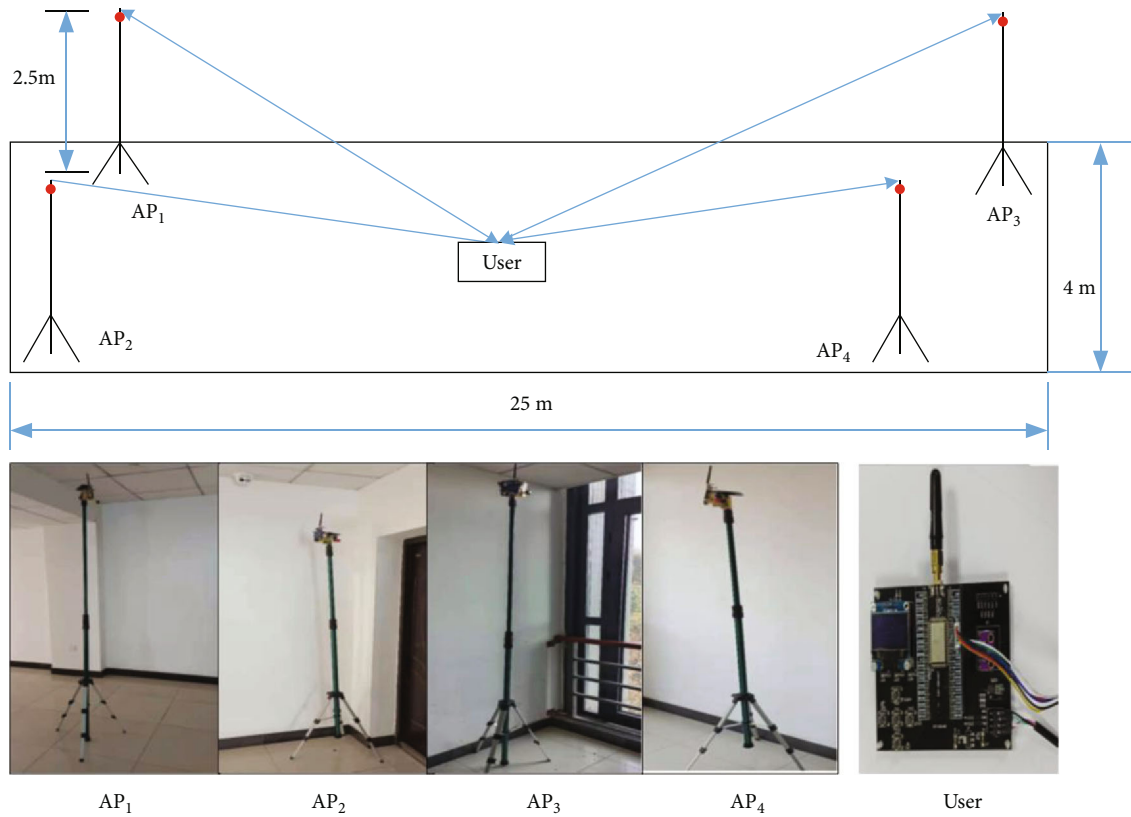
FIGURE 10: Experimental environment of Lora indoor positioning system.

and a_{11} are the linear correction factors, and r_{m-1} and r_m are the stages; the above data of LOS are divided into three sections; that is, $m = 3$.

3.3. Geometry Factor. The dilution of geometry precision (GDOP) [32, 33] can be expressed as

$$\text{cov} (\Delta X) = \sigma_\varepsilon^2 \cdot (A^T A)^{-1}. \tag{16}$$

If $(A^T A)^{-1}$ is defined as H, the diagonal elements of H are

as follows:

$$H = \begin{bmatrix} x\text{DOP}^2 & & \\ & y\text{DOP}^2 & \\ & & z\text{DOP}^2 \end{bmatrix}. \tag{17}$$

$HDOP$ and $GDOP$ can be defined as

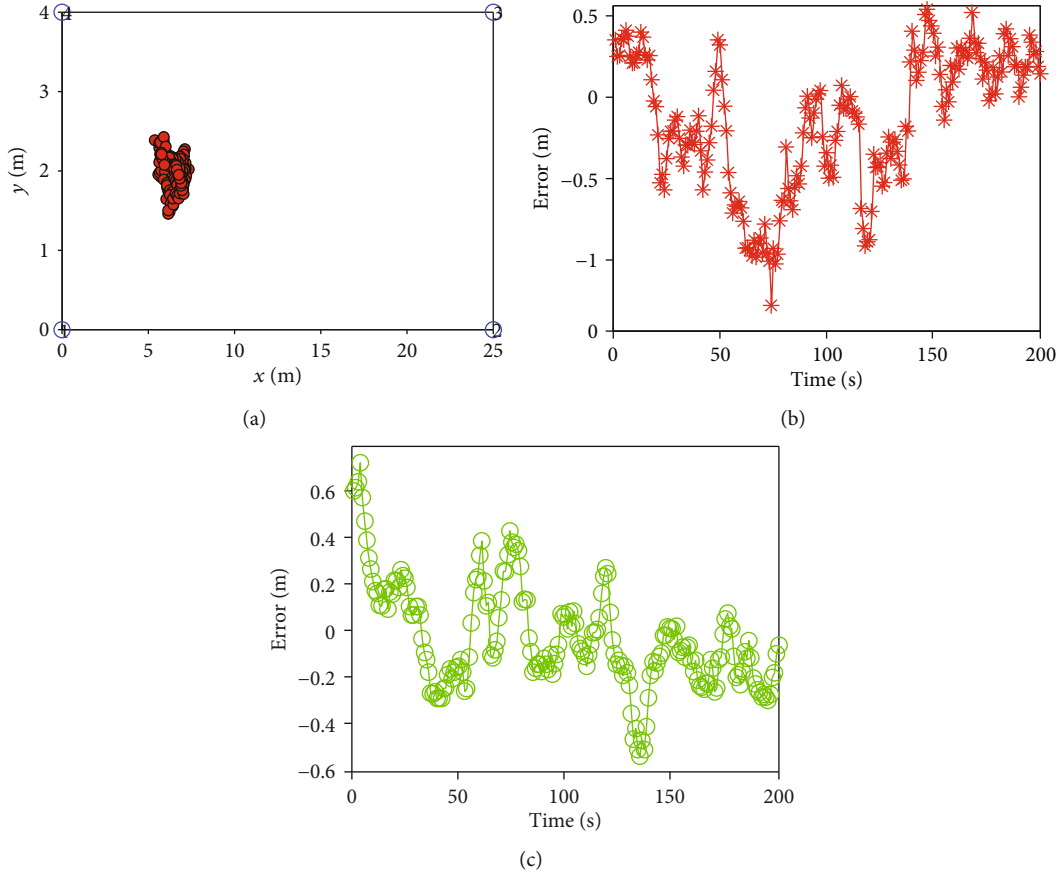$$HDOP = \sqrt{x\text{DOP}^2 + y\text{DOP}^2}, \tag{18}$$

(a)

(b)

(c)

FIGURE 11: Static test without pseudorange correction: (a) positioning results, (b) x-axis positioning error, and (c) y-axis positioning error.

$$GDOP = \sqrt{xDOP^2 + yDOP^2 + zDOP^2}, \quad (19)$$

where $xDOP^2$ means the dilution of precision (DOP) for the x-coordinate and $yDOP^2$ means the DOP for the y-coordinate.

In order to analyze the influence of the geometry factor on Lora positioning performance, a simulation of geometry distribution is designed, which is composed of four Lora transceivers. Suppose the length and width of the room are 40 meters and the height is 3 meters, as shown in Figure 7. The $HDOP$ of the indoor positioning system is given in Figures 8(a) and 9(a); the GDOP of the indoor positioning system is given in Figures 8(b) and 9(b). The results show that four Lora transceivers can obtain the suitable geometric distribution in their projection area, whose $HDOP$ and G DOP are less than 1.2. However, outside the projection area, the geometric distribution will deteriorate; $HDOP$ and G DOP will be greater than 2, which will make the positioning error more than twice of the ranging error.

4. Methodology Based on Particle Filter

4.1. Recursive Bayesian Estimation. Applied nonlinear filtering is based on discrete-time nonlinear state-space models relating a hidden state X_t to the observations Y_t, denote the observations at time t by $Y_t = \{y_0, \cdots, y_t\}$, the Bayesian solu-

tion to compute the posterior distribution is given by

$$\begin{cases} p\left(\dfrac{X_t}{Y_{t-1}}\right) = \int p\left(\dfrac{X_t}{X_{t-1}}\right) p\left(\dfrac{X_{t-1}}{Y_{t-1}}\right) dX_{t-1}, \\ p\left(\dfrac{X_t}{Y_t}\right) = \dfrac{p(y_t/X_t)p(X_t/Y_{t-1})}{p(y_t/Y_{t-1})}, \end{cases} \quad (20)$$

where t is the time stamp, x_t is the state variable, $p(X_{t-1}/Y_{t-1})$ is the posterior probability distribution of the last moment, $p(X_t/X_{t-1})$ is the state transition probability, $p(X_t/Y_{t-1})$ is the prior probability distribution, $p(y_t/X_t)$ is the likelihood function, and $p(y_t/Y_{t-1})$ is the normalization function.

4.2. Particle Filter. Supposed that N particles $\{X_t^i, w_t^i\}_{i=1}^N$ from the posterior probability $p(X_t/Y_t)$ of the state can be extracted, where X_t^i is the state of the particle, w_t^i is the weight of the particle; then,

$$p\left(\dfrac{X_t}{Y_t}\right) \approx \sum_{i=1}^N w_t^i \delta\left(X_t - X_t^i\right), \quad (21)$$

where δ is the Dirac delta function and N is the number of particles.
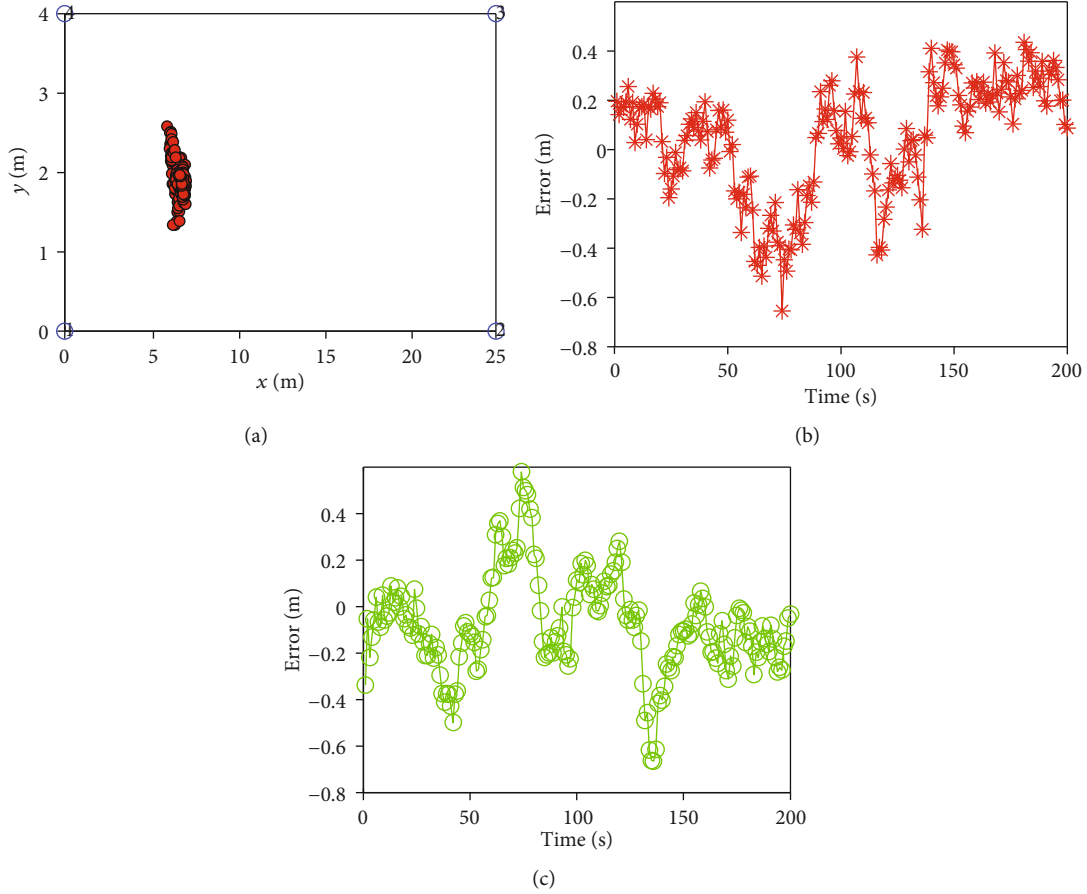
(a)

(b)

(c)

FIGURE 12: Static test with pseudorange correction: (a) positioning results, (b) x-axis positioning error, and (c) y-axis positioning error.

The Sequential Importance Sampling (SIS) method is used to calculate the weight of particles, which is written as

$$w_t^i = w_{t-1}^i \frac{p\left(Y_t/X_t^i\right)p\left(X_t^i/X_{t-1}^i\right)}{q\left(X_t^i/X_{t-1}^i, Y_t\right)}. \tag{22}$$

The prior probability distribution is used as the importance density function:

$$q\left(\frac{X_t^i}{X_{t-1}^i}, Y_t\right) = p\left(\frac{X_t^i}{X_{t-1}^i}\right). \tag{23}$$

Then, the formula for calculating the weight of particles is

$$w_t^i = w_{t-1}^i p\left(\frac{Y_t}{X_t^i}\right), \tag{24}$$

where t is time stamp, i is the number of particles, and w is the weight of particles.

4.3. Particle Filter Implementation. The particle filter algorithm is summarized in Table 2. Firstly, the state parameters and weights of particles are initialized; secondly, the iterative process of the particle filter algorithm is divided into four

steps, which includes measurement update, estimation, resampling, and time update.

5. Experimental Results and Analysis

5.1. Experimental Setup. The performance of the Lora positioning system is evaluated in a room as shown in Figure 10; the size of the room is about 25 meters long, 4 meters wide, and 2.5 meters high; and the antenna coordinates of Lora are surveyed precisely with a total station.

5.2. Experimental Results. Figure 11 shows the static test without pseudorange correction for the Lora positioning system, Figure 11(a) is the positioning results, Figure 11(b) is the x-axis positioning error, and Figure 11(c) is the y-axis positioning error. The average positioning error is 0.11 m in the x-axis and 0.07 m in the y-axis, the maximum positioning error is 1.25 m in the x-axis and 0.59 m in the y-axis, and the standard deviation of the x-axis and y-axis errors is 0.42 m and 0.18 m, respectively.

Figure 12 shows the static test without pseudorange correction (the piecewise fitting correction method) for the Lora positioning system, Figure 12(a) is the positioning results, Figure 12(b) is the x-axis positioning error, and Figure 12(c) is the y-axis positioning error. The average positioning error is 0.01 m in the x-axis and 0.07 m in the y-axis,
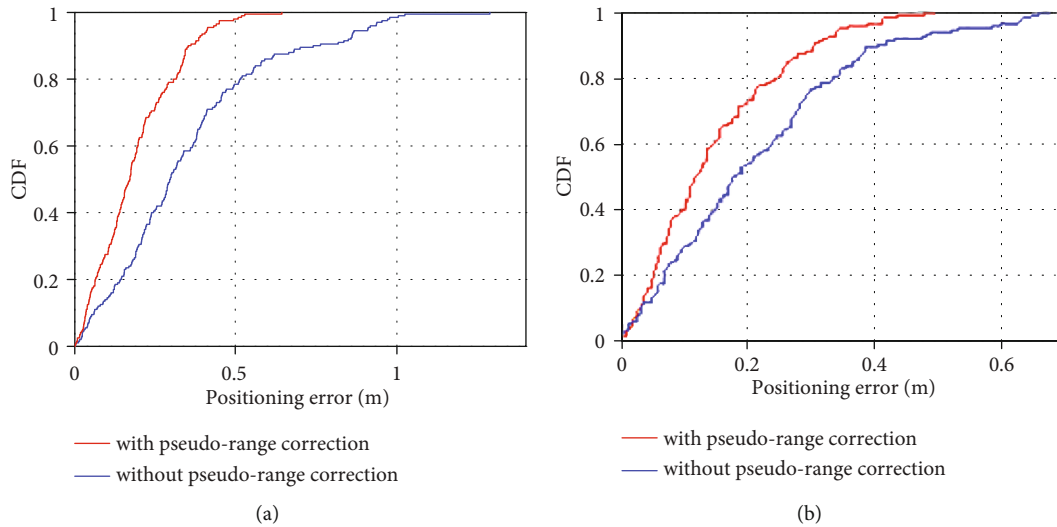
FIGURE 13: Comparison of CDF positioning errors: (a) x-axis and (b) y-axis.

the maximum positioning error is 0.72 m in the x-axis and 0.70 m in the y-axis, and the standard deviation of the x-axis and y-axis errors is 0.22 m and 0.12 m, respectively.

The positioning error in terms of the cumulative distribution function on the databases with and without pseudorange correction is shown in Figure 13. When the positioning error on the x-axis threshold is 0.2 m and 0.6 m, the CDF with pseudorange correction is 61% and 99%, which are higher than the 32% and 85% without pseudorange correction. When the positioning error on the y-axis threshold is 0.2 m and 0.6 m, the CDF with pseudorange correction is 71% and 99.9%, which are higher than the 52% and 94.8% without pseudorange correction.

6. Conclusions

The long-distance transmission of Lora wireless technology makes it possible to be widely used in the smart factory; this paper proposes Lora RTT measurement for indoor positioning, which has two key aspects of innovations: Firstly, a Lora-aided particle filter localization method is designed to solve the problem for indoor positioning. Secondly, numerous experiments were carried out with Lora RTT measurement data to evaluate the performance of the proposed approach; we used the CDF criteria to measure the quality of the estimated location in comparison to the truth location. The results show that the indoor positioning accuracy is improved obviously with the help of the piecewise fitting correction method. At the same time, the Lora indoor positioning system can achieve a positioning accuracy of 1 m under the condition of LOS. In the future, we will focus on Lora indoor positioning and pseudorange correction under the condition of NLOS.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L. Wan, L. Sun, K. Liu, X. Wang, Q. Lin, and T. Zhu, "Autonomous vehicle source enumeration exploiting non-cooperative UAV in software defined internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.

[2] L. Sun, L. Wan, and X. Wang, "Learning-based resource allocation strategy for industrial IoT in UAV-enabled MEC systems," *IEEE Transactions on Industrial Informatics*, 2020.

[3] L. Wan, Y. Sun, L. Sun, Z. Ning, and J. J. P. C. Rodrigues, "Deep learning based autonomous vehicle super resolution DOA estimation for safety driving," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, 2020.

[4] L. Sun, L. Wan, K. Liu, and X. Wang, "Cooperative-evolution-based WPT resource allocation for large-scale cognitive industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5401–5411, 2020.

[5] F. Adrion, A. Kapun, F. Eckert et al., "Monitoring trough visits of growing-finishing pigs with UHF-RFID," *Computers and Electronics in Agriculture*, vol. 144, pp. 144–153, 2018.

[6] F. Adrion, A. Kapun, E.-M. Holland, M. Staiger, P. Löb, and E. Gallmann, "Novel approach to determine the influence of pig and cattle ears on the performance of passive UHF-RFID ear tags," *Computers and Electronics in Agriculture*, vol. 140, pp. 168–179, 2017.

[7] B. Wang, X. Gan, X. Liu et al., "A novel weighted KNN algorithm based on RSS similarity and position distance for Wi-

Fi fingerprint positioning," *IEEE Access*, vol. 8, pp. 30591–30602, 2020.

[8] B. Wang, X. Liu, B. Yu, R. Jia, and X. Gan, "An improved WiFi positioning method based on fingerprint clustering and signal weighted Euclidean distance," *Sensors*, vol. 19, no. 10, pp. 2300–2319, 2019.

[9] Z. Hao, B. Li, and X. Dang, "A method for improving uwb indoor positioning," *Mathematical Problems in Engineering*, vol. 2018, 17 pages, 2018.

[10] J. Zhang, P. V. Orlik, Z. Sahinoglu, A. F. Molisch, and P. Kinney, "UWB systems for wireless sensor networks," *Proceedings of the IEEE*, vol. 97, no. 2, article 2008786, pp. 313–331, 2009.

[11] M. R. Mahfouz, C. Zhang, B. C. Merkl, M. J. Kuhn, and A. E. Fathy, "Investigation of high-accuracy indoor 3-D positioning using UWB technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 6, pp. 1316–1330, 2008.

[12] X. Gan, B. Yu, X. Wang et al., "A new array pseudolites technology for high precision indoor positioning," *IEEE Access*, vol. 7, pp. 153269–153277, 2019.

[13] X. Gan, B. Yu, L. Huang et al., "Doppler differential positioning technology using the BDS/GPS indoor array pseudolite system," *Sensors*, vol. 19, no. 20, pp. 4580–4580, 2019.

[14] L. Huang, X. Gan, B. Yu et al., "An innovative fingerprint location algorithm for indoor positioning based on array pseudolite," *Sensors*, vol. 19, no. 20, pp. 4420–4420, 2019.

[15] L. Ruan, L. Zhang, T. Zhou, and Y. Long, "An improved Bluetooth indoor positioning method using dynamic fingerprint window," *Sensors*, vol. 20, no. 24, pp. 7269–7269, 2020.

[16] S. N. Swamy and S. R. Kota, "An empirical study on system level aspects of internet of things (IoT)," *IEEE ACCESS*, vol. 8, pp. 188082–188134, 2020.

[17] P. J. Basford, F. M. Bulot, M. Apetroaie-Cristea, S. J. Cox, and S. J. Ossont, "LoRaWAN for smart city IoT deployments: a long term evaluation," *Sensors*, vol. 20, no. 3, pp. 648–648, 2020.

[18] I. Martin-Escalona and E. Zola, "Passive round-trip-time positioning in dense IEEE 802.11 networks," *Electronics*, vol. 9, no. 8, 2020.

[19] D. R. Novotny, J. R. Guerrieri, and D. G. Kuester, "Potential interference issues between FCC part 15 compliant UHF ISM emitters and equipment passing standard immunity testing requirements," *IEEE Electromagnetic Compatibility Magazine*, vol. 1, no. 3, pp. 92–96, 2012.

[20] ARIB, "Homepage," http://arib.or.jp/english/html/overview/doc/5-STD-T66v2_1-E.pdf.

[21] R. Liang, L. Zhao, and P. Wang, "Performance evaluations of Lora wireless communication in building environments," *Sensors*, vol. 20, no. 14, pp. 3828–3828, 2020.

[22] T. Huynh and C. Brennan, "Efficient UWB indoor localisation using a ray-tracing propagation tool," in *Proceedings of the Ninth IT & T Conference*, Dublin, Ireland, October, 2009Technological University Dublin.

[23] K. Staniec and M. Kowal, "Lora performance under variable interference and heavy-multipath conditions," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 6931083, 9 pages, 2018.

[24] Y. Wang, W. Zhang, F. Li, Y. Shi, F. Nie, and Q. Huang, "UAPF: a UWB aided particle filter localization for scenarios with few features," *Sensors*, vol. 20, no. 23, pp. 6814–6814, 2020.

[25] A. Dhital, P. Closas, and C. Fernández-Prades, "Bayesian filtering for indoor localization and tracking in wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, Article ID 227, p. 13, 2012.

[26] F. Gustafsson, "Particle filter theory and practice with positioning applications," *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 53–82, 2010.

[27] M. Khalaf-Allah, "Particle filtering for three-dimensional TDoA-based positioning using four anchor nodes," *Sensors*, vol. 20, no. 16, pp. 4516–4516, 2020.

[28] S. Dwivedi, A. De Angelis, D. Zachariah, and P. Handel, "Joint ranging and clock parameter estimation by wireless round trip time measurements," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2379–2390, 2015.

[29] C. MA, B. Wu, S. Poslad, and D. R. Selviah, "Wi-Fi RTT ranging performance characterization and positioning system design," *IEEE Transactions on Mobile Computing*, 2020.

[30] N. Podevijn, D. Plets, J. Trogh et al., "TDoA-based outdoor positioning with tracking algorithm in a public LoRa network," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 1864209, 9 pages, 2018.

[31] K. Zhao, T. Zhao, Z. Zheng, and C. Yu, "Optimization of time synchronization and algorithms with TDOA based indoor positioning technique for Internet of things," *Sensors*, vol. 20, no. 22, pp. 6513–6513, 2020.

[32] R. B. Langley, "Dilution of precision," *GPS world*, vol. 10, no. 5, pp. 52–59, 1999.

[33] J. D. Bard and F. M. Ham, "Time difference of arrival dilution of precision and applications," *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 521–523, 1999.

WILEY | Hindawi

Research Article
SAR Image Target Recognition Based on Monogenic Signal and Sparse Representation

XiuXia Ji [1] and **Yinan Sun** [2]

[1]*Nanjing Vocational College of Information Technology, Nanjing 210023, China*
[2]*Wuhan University, School Electronic Information, Wuhan 430072, China*

Correspondence should be addressed to XiuXia Ji; jixx@njcit.cn

It is necessary to recognize the target in the situation of military battlefield monitoring and civilian real-time monitoring. Sparse representation-based SAR image target recognition method uses training samples or feature information to construct an overcomplete dictionary, which will inevitably affect the recognition speed. In this paper, a method based on monogenic signal and sparse representation is presented for SAR image target recognition. In this method, the extended maximum average correlation height filter is used to train the samples and generate the templates. The monogenic features of the templates are extracted to construct subdictionaries, and the subdictionaries are combined to construct a cascade dictionary. Sparse representation coefficients of the testing samples over the cascade dictionary are calculated by the orthogonal matching tracking algorithm, and recognition is realized according to the energy of the sparse coefficients and voting recognition. The experimental results suggest that the new approach has good results in terms of recognition accuracy and recognition time.

1. Introduction

As a new kind of reconnaissance remote sensing device, SAR is widely used in aerial and space reconnaissance, monitoring, and intelligent tracking of moving aerial targets [1–3]. UAVs are widely used in military surveillance, smart home monitoring, and target tracking. UAV-borne SAR has become an important development direction of UAV remote sensing earth observation technology. In the situation of military battlefield monitoring and smart cities monitoring, it is necessary to classify and identify the target. SAR image target recognition mainly refers to radar detection of targets, processing of echo information, and determination of target attributes, categories, or types. The spectral property for SAR images is determined by the back-scattered signal that is simply the back reflected part of the microwaves scattered from the land cover. Since the back-scattered signal is very weak, it is very difficult to distinguish different types of targets in SAR images. At the same time, the inherent speckle noise will play a vital role in information extraction from SAR images. Being affected by the inherent speckle noise, SAR images are inferior in readability. In addition, the image features change tremendously as slight fluctuations of imaging parameters or the variation of surroundings which will affect the accuracy and speed of SAR image target recognition.

SAR image target recognition mainly consists of three stages: image processing, feature extraction, and classifier design. The purpose of image processing is to remove speckle noise, segment SAR image, and make it easy to extract features and recognize the image target. Feature extraction directly affects the accuracy of SAR image target recognition. Features used for image classification usually include principal component analysis (PCA), generalized 2-dimensional principal component analysis (G2DPCA), independent component analysis (ICA), and wavelet. For two-dimensional images, monogenic signal perfectly reproduces the monogenic amplitude of the signal energy, monogenic phase of

the signal structure information, and monogenic orientation of the signal geometry information, which has been widely used in the field of image processing [4–6].

Similarly, there are many classifiers for SAR image recognition, such as support vector machine (SVM) [7], k-nearest neighbor (KNN) [8], and neural network (NN) [9]. In the above recognition algorithms, in order to ensure the recognition performance, KNN classifier theoretically requires an infinite number of training samples, which is obviously difficult to meet in practical application. SVM classifier transforms the linear inseparable problem into a linear divisible problem by using spatial projection. The huge amount of computation caused by training seriously affects the recognition speed of SAR image targets. NN classifier uses the training samples to learn the parameters and weights of training networks. When the categories and numbers of training samples are large, the corresponding computation is also very large, which will lead to the lack of convergence in the training process. Therefore, the research of SAR image target recognition algorithm urgently needs to inject new elements.

In recent years, sparse representation of image signals has been widely concerned in the field of pattern recognition. Wright J first proposed a sparse representation-based classifier (SRC); it constructed an overcomplete dictionary with multiple kinds of training samples with label information and classified by using sparse representation coefficients of testing samples on the dictionary [10]. Now, sparse representation has been widely applied to face recognition [11–13], the direction of arrival estimation, tensor modeling and angle estimation [14–16], and SAR image target recognition [17, 18]. Extended maximum average correlation height (EMACH) filter is actually a kind of template filter; it is widely applied to the recognition of specific military targets because of its high matching ability and strong antinoise ability. EMACH combined with exponential wavelet fractal feature and G2DPCA feature is applied to target detection of SAR image, and EMACH combined with G2DPCA feature is used to complete SAR image target recognition.

SAR image target recognition algorithm based on sparse representation is designed from two aspects. The one is using dictionary learning to complete recognition. There are two main methods; one is to train dictionaries directly to make them discriminant. That is to say, the dictionary should be designed according to the training sample and has certain adaptability. If the features of training samples are directly extracted to construct overcomplete dictionary, the dictionary dimension is high and the redundancy is large, which will affect the speed of testing samples recognition. The second is to study and optimize the dictionary in order to improve its recognition ability. For example, Literature [19] used discriminant KSVD dictionary learning method to complete SAR image target recognition. The other one is using sparse coefficient to complete recognition. Unlike training multicategory dictionaries, this approach only requires training a whole dictionary without paying attention to the category of each dictionary. The recognition algorithm based on this method usually needs to consider adding the recognition error to the cost function when learning the dictionary, so that the trained dictionary has good recognition ability.

In this paper, SAR image target recognition is also studied from the point of dictionary design and sparse coefficient solving. EMACH filter is used to train samples and generate template samples; the monogenic features of all templates are extracted; three subdictionaries are generated according to the monogenic amplitude, the monogenic phase, and the monogenic orientation; the sparse coefficients of the testing samples' monogenic feature in each level dictionary are solved; and the category of samples to be tested is determined by the image reconstruction error.

The main contributions of this paper are summarized as follows:

(1) The proposed SAR image target recognition method uses sparse representation to identify SAR image target, which eliminates the need for additional suppression of speckle noise, reduces the steps of SAR image processing and saves time

(2) The proposed SAR image target recognition method uses three subdictionaries to construct an overcomplete cascade dictionary, and it obviously improves the recognition speed in the precondition of ensuring the recognition accuracy

The remainder of this paper is organized as follows. Section 2 gives a brief survey of the related works about SAR image target recognition method based on sparse representation. Section 3 describes the motivation and design of the proposed SAR image recognition method which is the major contribution of this paper. Experimental results and discussion are shown in Section 4. Finally, the concluding remarks are given in Section 5.

2. SAR Image Recognition Based on Sparse Representation

When sparse representation is originally used in face recognition, an overcomplete dictionary is mainly constructed by the training samples; the testing sample is expressed as a linear combination of atoms in the overcomplete dictionary. The face image can be accurately recovered according to the sparse representation coefficients, and the recognition of the target is realized according to the distance metric between the reconstructed image and the testing sample image.

There are similarities and differences between SAR image target recognition and face recognition. They are same in constructing the overcomplete dictionary with the training samples. If the testing sample can be represented linearly by the atoms of the overcomplete dictionary, and the coefficient of the corresponding target category in all the sparse representation coefficients is large, then the representation of the testing sample in the dictionary is sparse, and the recognition discrimination is completed according to the energy characteristics of the sparse representation coefficient. SAR image target recognition method based on sparse representation omits the reconstruction process in face recognition method. The framework of the specific recognition method is shown in Figure 1.
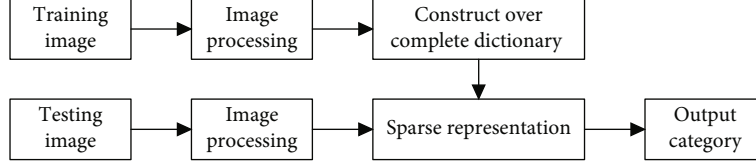
FIGURE 1: The framework of SAR image target recognition based on sparse representation.

In SAR image target recognition, it is assumed that there are k class samples; column vector set $D_i = [v_{i,1}, v_{i,2}, \cdots, v_{i,n_i}]$ $\in R^{m \times n_i}$ of matrix is composed of n_i training samples of class i target. Any kind of testing sample $y \in R^m$ can be expressed as a linear combination of the training samples.

$$y = \alpha_{i,1}v_{i,1} + \alpha_{i,2}v_{i,2} + \cdots + \alpha_{i,n_i}v_{i,n_i}, \quad (1)$$

where $\alpha_{i,j} \in R(j = 1, 2, \cdots, n_i)$ is the coefficient of linear representation of the testing sample in the dictionary.

If the class i of the testing sample y is unknown, all k class n training sample sets are formed into matrix D, that is,

$$D = [D_1, D_2, \cdots D_k] = [v_{1,1}, v_{1,2}, \cdots v_{k,n_k}]. \quad (2)$$

The linear representation of test sample y under all training samples is as follows.

$$y = \alpha_{1,1}v_{1,1} + \alpha_{1,2}v_{1,2} + \cdots + \alpha_{k,n_k}v_{k,n_k}, \quad (3)$$

$$y = D\hat{x}, \quad (4)$$

where $\hat{x} = [0, \cdots, 0, \cdots \alpha_{i,1}, \alpha_{i,2}, \cdots \alpha_{i,n_i}, \cdots, 0, \cdots, 0]^T \in R^n$ is a coefficient vector.

In an ideal case, only the same type of training sample factor as the testing sample may be nonzero in the $\hat{x}$; the corresponding coefficient for the other class samples shall be 0. Since the sparse coefficient $\hat{x}$ contains the information of the target category, the recognition of the target can be realized by the solution of the formula $y = D\hat{x}$.

In the SAR image target recognition method based on sparse representation, the structure of overcomplete dictionary is very critical. The dictionary must be of low dimension, and the atoms of the dictionary should correspond to the properties of the SAR image. At the same time, sparse representation coefficients of different classes of targets over the dictionary must be distinguishable.

Considering the above points, if the overcomplete dictionary is composed of SAR image pixels, that is, the overcomplete dictionary is constructed directly by stretching the central region of the image into a column vector. Because of the high dimension of the dictionary, it will directly affect the speed of solving the subsequent sparse representation coefficients. The overcomplete dictionary can be constructed by using the representation vector which can describe the characteristics of SAR image target. Thus, it can reduce the

dimension of dictionary atoms and improve the speed of sparse solution. In [20], G2DPCA features are used to construct the overcomplete dictionary, which reduces the dimension of extracted features, improves the recognition performance, and has good robustness to the change of target azimuth.

3. The Proposed SAR Image Target Recognition Method

The above SAR image target recognition method based on sparse representation takes the feature information extracted by all categories of the training samples or the training samples as the atoms of the overcomplete dictionary. When the number of the training samples is too large, the dictionaries generated by these two methods are too large, which is bound to affect the speed of recognition. In this paper, a cascade dictionary is adopted, that is, subdictionaries are generated according to the monogenic feature of the training samples and the category of the sample to be determined by the reconstruction error. For simplicity, the system framework is divided into the following four stages: image processing, EMACH train, feature extraction, and target recognition. Explanation of each step is described below.

3.1. Image Processing. First of all, we extract the region containing the target in the image and generate a new target image based on the biggest scattering point of the target section as the sample image. Since sparse representation can effectively remove noise [21], there is no need for additional noise removal of SAR images in this paper. Take T7 tank image (138×138) with angle as a sample. Select the target region of 45×45 contained in the center of the image as the sample image (as shown in Figure 2).

It can be seen from the images before and after the processing that the sample image after processing aims at description on the target, the detail information of the target has been enhanced, and the influence of the surrounding clutters on the target is reduced.

3.2. EMACH Train. EMACH obtains a two-dimensional function by training the sample image, and then, the correlation response of the image is obtained after doing the correlation operation with the image to be detected of the same size, and the target is judged according to the intensity of the response.

First input images of size 45×45 extend the pixel to a one-dimensional vector x_i with the length of 45×45 line by line from left to right and from top to bottom. Define h as

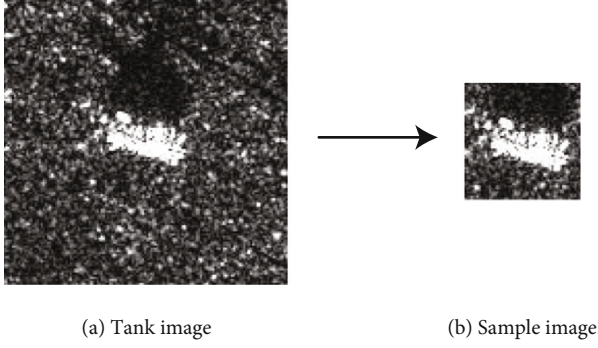(a) Tank image   (b) Sample image

FIGURE 2: Image processing.

the filter of EMACH. FFT2 () represents Fourier operation and assign $\beta \in (0, 1)$, $M = \text{FFT}(m)$, $X_i = \text{FFT}(x_i)$, and $i = 1, 2, \cdots, N$.

$$C_x^\beta = \frac{1}{N} \sum_{i=1}^{N} [x_i - (1 - \beta)m]^+ [x_i - (1 - \beta)m], \quad (5)$$

$$S_x^\beta = \frac{1}{N} \sum_{i=1}^{N} [X_i - (1 - \beta)M]^+ [X_i - (1 - \beta)M], \quad (6)$$

where m is the mean value of x_i. The symbol of "+" represents matrix transposition.

$$J(h) = \frac{h^+ C_x^\beta h}{h^+ (I + S_x^\beta) h}. \quad (7)$$

When the value of formula (7) is maximum, h is the EMACH filter.

In this paper, the sample image in MSTAR database is processed. The templates are obtained using EMACH, and all the images within the range of 12° azimuth are selected to train a template, and 30 templates are trained for each type of image, for a total of 150 templates. Figure 3 shows all EMACH template images.

3.3. *Feature Extraction*. The traditional Gabor filter uses an adjustable filter to filter amplitude and phase in different directions and scale. Riesz transform is introduced in the analysis of monogenic signal. The expression of the Riesz transform kernel in the spatial domain is defined as

$$\eta_z = (\eta_x, \eta_y) = \left(\frac{x}{2\pi \|z\|^3}, \frac{y}{2\pi \|z\|^3} \right). \quad (8)$$

Suppose $w = (u, v)$, u, v represent two coordinates in the frequency domain, the frequency domain response of Log-Gabor filter is

$$g_{LG}(w) = \exp \left\{ \frac{-[\log (w/\gamma_0)]^2}{2[\log (\sigma/\gamma_0)]^2} \right\}, \quad (9)$$

where γ_0 is the central frequency and σ is the scale of Log-Gabor filter. Log-Gabor filters of different scales can be obtained by modify σ.

The band communication number generated by the filter of 2D image f can be expressed as

$$h = f * F^{-1}(g_{LG}(w)), \quad (10)$$

where $*$ is the convolution operator and F^{-1} represents the inverse Fourier transform.

$$h_x = h * \eta_x, \quad (11)$$

$$h_y = h * \eta_y, \quad (12)$$

where h is the real part of monogenic transformation and h_x and h_y are the two imaginary parts.

For a given image f, the monogenic amplitude A, the monogenic phase ϕ, and the monogenic orientation θ can be calculated by the following formula:

$$A = \sqrt{h^2 + h_x^2 + h_y^2}, \quad (13)$$

$$\phi = \arctan \left(\frac{\sqrt{h_x^2 + h_y^2}}{h} \right), \quad (14)$$

$$\theta = \arctan \left(\frac{h_y}{h_x} \right). \quad (15)$$

Obtain the monogenic feature of the ind template image with s scale of log-Gabor filter; it can be described as

$$\left\{ A^{i,1}, \phi^{i,1}, \theta^{i,1}, A^{i,2}, \phi^{i,2}, \theta^{i,2}, \cdots, A^{i,S}, \phi^{i,S}, \theta^{i,S} \right\}. \quad (16)$$

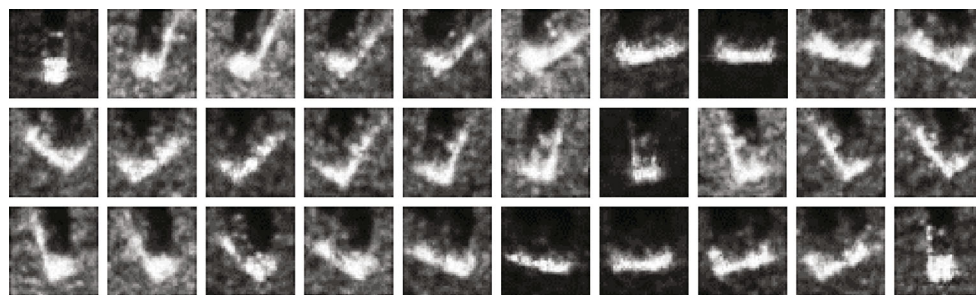Pull the monogenic feature into one-dimensional vector and normalize it:

$$D^i = \left\{ \chi_A^{i,1}, \chi_A^{i,2}, \cdots, \chi_A^{i,S}, \chi_\phi^{i,1}, \chi_\phi^{i,2}, \cdots, \chi_\phi^{i,S}, \chi_\theta^{i,1}, \chi_\theta^{i,2}, \cdots, \chi_\theta^{i,S} \right\}. \quad (17)$$

Assuming the total number of the training samples is n, and the subdictionary D_k can be expressed as:
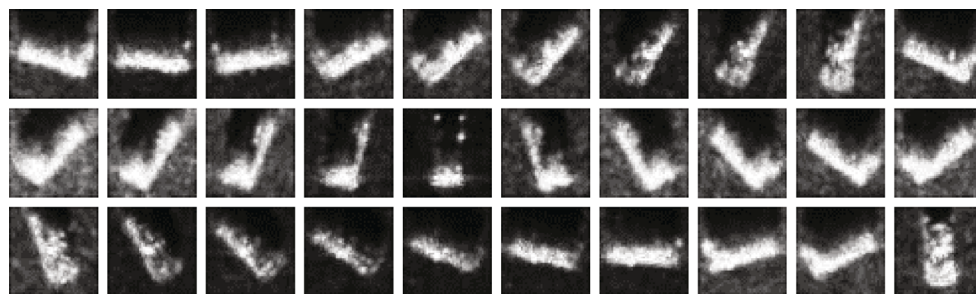
$$D_1 = \left\{ \chi_A^{1,1}, \chi_A^{1,2}, \cdots, \chi_A^{1,S}, \cdots, \chi_A^{n,1}, \chi_A^{n,2}, \cdots, \chi_A^{n,S} \right\}, \quad (18)$$

$$D_2 = \left\{ \chi_\phi^{1,1}, \chi_\phi^{1,2}, \cdots, \chi_\phi^{1,S}, \cdots, \chi_\phi^{n,1}, \chi_\phi^{n,2}, \cdots, \chi_\phi^{n,S} \right\}, \quad (19)$$
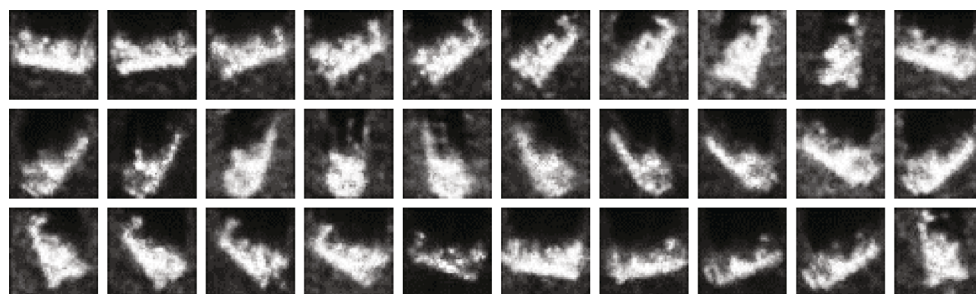
$$D_3 = \left\{ \chi_\theta^{1,1}, \chi_\theta^{1,2}, \cdots, \chi_\theta^{1,S}, \cdots, \chi_\theta^{n,1}, \chi_\theta^{n,2}, \cdots, \chi_\theta^{n,S} \right\}. \quad (20)$$
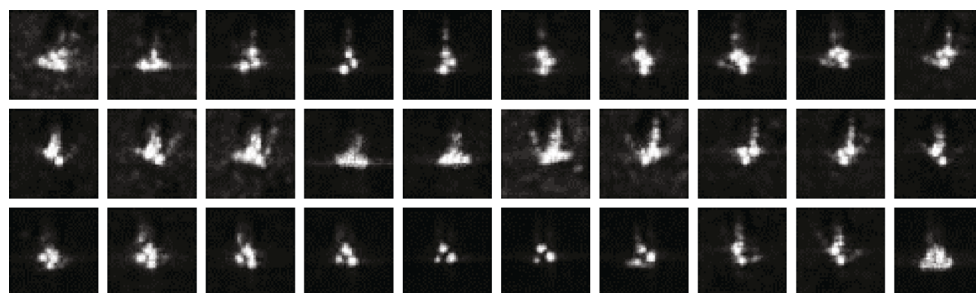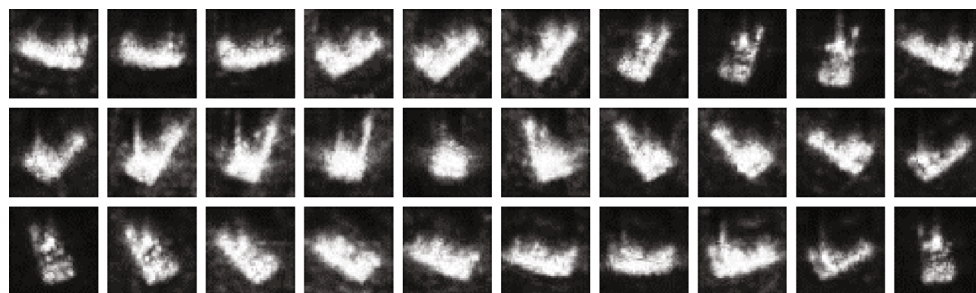
(a) Template images of BRDM2

(b) Template images of 2S1

(c) Template images of T72

(d) Template images of SLICY

(e) Template images of ZSU234

Figure 3: Template images.

Each subdictionary can be treated as a binary classifier, cascading all subdictionaries into a cascading dictionary.

3.4. Target Recognition. In early research, ensembles were shown empirically and theoretically to possess better accuracy than any single component classifier [22]. Since each classifier can only classify the samples of corresponding category, it can be considered a weak classifier. However, a plurality of weak classifiers are cascaded to form a strong classifier with strong recognition capability, and the cascade classifier designed by the hierarchical structure can make most of the samples of the previous categories be recognized directly in the previous classifiers, and only a small part of samples of the latter categories or the missed samples will pass all the subclassifiers, and the recognition time can be obviously reduced.

As stated earlier, the objective of this paper is to develop a cascade classifier that will improve SAR image target recognition accuracy and reduce recognition time. For this purpose, we trained three subdictionaries and then combined them into a cascade dictionary. Because the final classifier uses the cascade structure, the recognition performance of each subclassifier is very important. In this paper, the subclassifiers are designed based on the monogenic features, each monogenic feature generates a subdictionary, the sparse coefficient of the testing sample under each subdictionary is obtained, the testing sample is reconstructed by the coefficient, and the sample category is determined by the reconstruction error. The selection of reconstruction error has a great influence on the recognition performance of subclassifiers. In general, a larger reconstruction error threshold results in a shorter recognition time, but a higher recognition error rate is also generated. The reconstruction error threshold selected in this paper is large. In this way, although there are some samples missed recognition, the last voting mechanism guarantees the correct recognition of this part of samples.

The system framework of the proposed SAR image target recognition method based on monogenic signal and sparse representation is shown in Figure 4.

The main steps of the SAR image target recognition method based on monogenic signal and cascade dictionary are described below.

Step 1. Process the training samples.

Step 2. Train the samples with the EMACH filter and generate the template samples.

Step 3. Extract the monogenic features of the template samples, and generate subdictionaries $D_i\,(i = 1, 2, 3)$ according to the monogenic amplitude, the monogenic phase, and the monogenic orientation features.

Step 4. For any testing sample $y \in R^m$, the monogenic feature is extracted as $X = \{A, \varphi, \theta\}$. Repeat the following steps until the condition is met.

(i) For signal $X_i\,(i = 1, 2, 3)$, use the improved orthogonal matching pursuit algorithm to solve underdetermined linear equations $X_i = D_i\alpha$, and find out the most sparse coefficient α

$$\min_{\alpha} \|\alpha\|_1 \text{s.t.} \|y - D_i\alpha\|_2 < \varepsilon. \qquad (21)$$

(ii) Since all the atoms in the dictionary have labels, the sum of coefficients of the k class can be calculated:

$$r_k(y) = \text{sum}\left(\left|\alpha_{k,i}\right|\right), i = 1, 2, \cdots, n_k. \qquad (22)$$

(iii) Determine the category of testing sample according to the sum of coefficients

$$\text{identity}(y) = \arg\max_k\left(r_k(y)\right). \qquad (23)$$

(iv) If the category is the same as the previous category, directly output category k and skip to Step 5; otherwise, $i = i + 1$; if $i > 3$, skip to Step 5; otherwise, Step 4 is repeated

Step 5. If the testing sample falls into different categories under the three subdictionaries, the category is determined by the voting recognition according to the reconstruction error.

$$\text{identity}(y) = \arg\min_k\left(|X_i - D_i\alpha_k|\right). \qquad (24)$$

4. Experimental Results and Analysis

In this section, we describe and discuss the experimental results obtained on the study sites introduced in Section 3. We carried out several experiments with the aim to supply a complete analysis of the performance of the proposed SAR image target recognition method. We investigate different aspects: (i) we perform an experiment to evaluate the recognition accuracy of the proposed recognition method and (ii) we perform an in-depth comparative analysis of the performance of the proposed method with respect to the other recognition methods.

4.1. Data Set Used for Experiments. The experimental data used herein is from the SAR ground still data of the SAR ground target high resolution provided by the Working Group of the United States DARPA/AFRL MSTAR. As a result of the SAR imaging, even if the target is same, the azimuth difference can cause the difference of the characteristic information, and the difference of the azimuth angle can
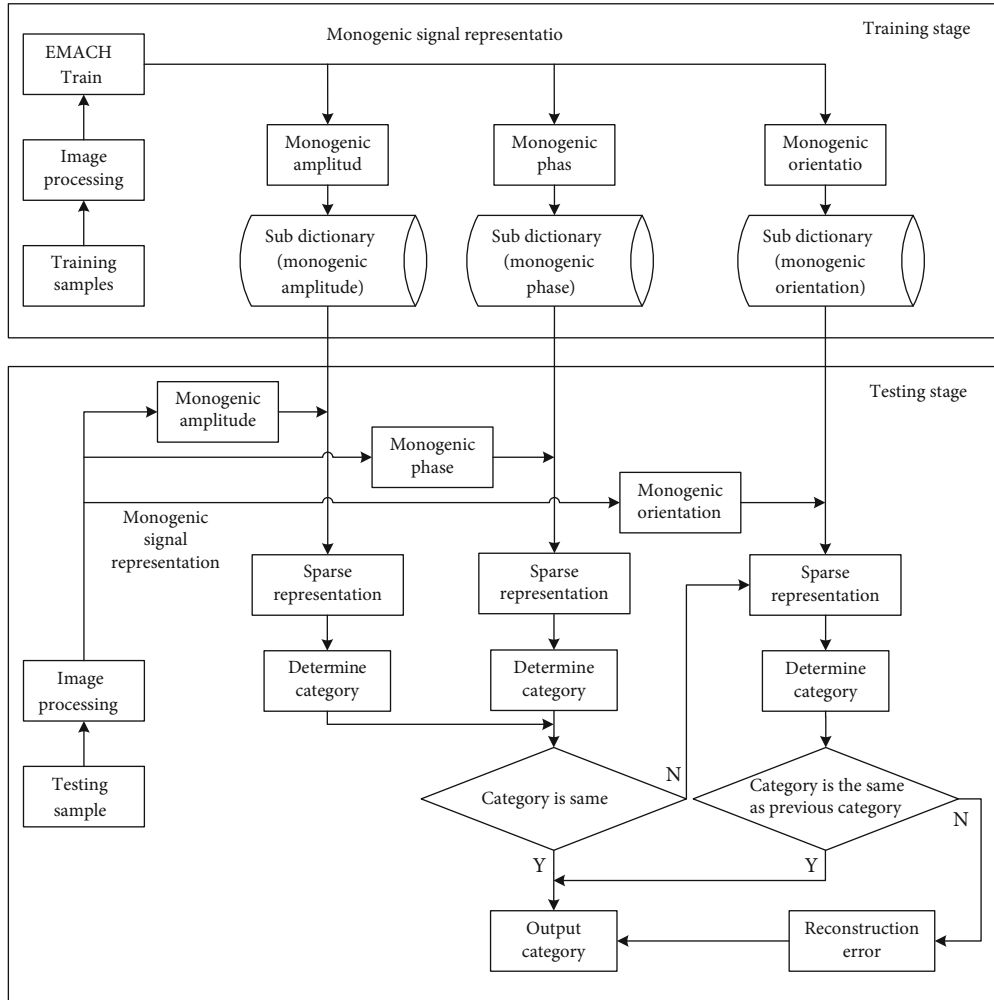
FIGURE 4: The system framework of cascade structure classifier.

cause the difference between the postimaging targets, because the training samples are required to contain all the imaging data at different angles. The SAR image data in the MSTAR database is comprehensive; the azimuth coverage of each target is from 0° to 360°. The experiment selects a subset of MSTAR database, including five types of SAR images: BRDM2, 2S1, T72, SLICY, and ZSU234. The imaging resolution of the image is 0.3 m × 0.3 m, and the azimuth range is from 0° to 360°. In the experiments, the imaging data at 17° are selected as the training samples, and the imaging data when the angle is 15° are used as the testing samples. Since some of the SAR images are noise polluted seriously from certain angle, the amount of images sample for every single target can be used is different. The amount of training samples and testing samples used in this experiment is shown in Table 1.

In order to compare efficiency of proposed method, we establish two characteristics as a comparison basis:

(i) Recognition accuracy: proportion of the testing samples correctly recognized by the algorithm.

(ii) Recognition time: the time it takes an algorithm to complete a given task.

The experiment was conducted with Matlab-2012b 64bit, installed on windows 10 professional 64bit with an Intel Core i7 Processor (8MCache, up to 3.90 GHz), and 16 GB of RAM.

4.2. Recognition Performance Analysis. The monogenic features of each kind of sample images are extracted and drawn into column vectors to form three subdictionaries $D_i (i = 1, 2, 3)$, concatenate the subdictionaries, and form a cascade classifier shown in Figure 4. For the testing samples, the region with 45×45 size of the target in the center of the image is also selected to realize the segmentation, and the recognition accuracy of each class of testing samples with the proposed method in this paper is shown in Table 2.

The correct sample of primary recognition refers to the samples that do not need to participate in the final voting mechanism recognition, and the correct sample of voting recognition refers to the samples that need to participate in the final voting mechanism recognition, that is, the samples that are missed in cascade recognition. Suppose the total number of samples is N, correct sample number of primary

Target type	Training sample	Testing sample
BRDM2	298	274
2S1	299	274
T72	232	196
SLICY	299	288
ZSU234	299	274

recognition is n_1, and correct sample number of voting recognition is n_2.

$$\text{Primary recognition accuracy} = \frac{n_1}{N} \times 100\%, \qquad (25)$$

$$\text{Voting recognition accuracy} = \frac{n_2}{N} \times 100\%, \qquad (26)$$

$$\text{Recognition accuracy} = \frac{n_1 + n_2}{N} \times 100\%. \qquad (27)$$

As can be seen from Table 2, in the design of the classifier, because the reconstruction error threshold is large, about 10% of the samples are missed in the first recognition. However, because of the existence of voting mechanism, this part of the missed samples has been correctly recognized at the time of the final vote.

Additionally, in order to further assess the effectiveness of the proposed recognition method, this paper compared it with the other four methods including the traditional classifiers, such as the SVM recognition method (method 1), the KNN recognition method (method 2), the method with the dictionary directly generated by image pixels (method 3), and the method with the dictionary generated by G2DPCA features in [20] (method 4).

In method 1, SVM is a small sample learning classification method, which has strong generalization ability. Nonlinear processing can be easily realized by introducing kernel function mapping. Suppose monogenic feature is $x = \{x_1, x_2, \cdots x_M\}$, where M is the number of categories of the samples. For the tagged training sets (x_i, y_i), $i = 1, 2, \cdots, M$, $y_i \in \{1, -1\}$, SVM is to solve the following optimization problems.

$$\min_{w,b,\eta} \frac{1}{2} w^T w + C \sum_{i=1}^{K} \eta_i \text{ subject to } y_i \left(w^T f(x_i) + b \right) \geq 1 - \eta_i \eta_i \geq 0, \qquad (28)$$

where C is the penalty factor of the error term.

Usually, RBF kernel function is used to map the training vector to high dimensional space.

$$K\left(x_k, x_g\right) = \varphi(x_k)^T \varphi(x_g) = \exp\left(-\gamma \|x_k - x_g\|^2\right) > 0, \qquad (29)$$

where γ is a nuclear factor. In this paper, the kernel function selected is $k(u, v) = \exp(-\gamma \|u - v\|)$, $\gamma = 5$.

In method 2, KNN is a simple and effective technique for objects classification according to the closest training examples in the feature space. KNN rates the neighbor of a test sequence among the training sample and uses the class labels of the nearest neighbor to predict the test vector class.

For the tagged training sets (x_i, y_i), $y_i \in \{1, -1\}$, the Euclidean distance is often used as the distance metric to measure the similarity between two vectors.

$$d^2\left(x_i, x_j\right) = \|x_i - x_j\|^2 = \sum_{k=1}^{d} \left(x_{ik} - x_{jk}\right)^2. \qquad (30)$$

Parameter k represents the number of neighbors in a set of training observations which are nearest to the given observation in validation or testing data set. Variation of this parameter will affect the accuracy of each binary classifier inside an expert. In this paper, parameter $k = 5$.

In method 3, the dictionary is constructed directly with the image pixels. In method 4, the dictionary is constructed directly with the G2DPCA feature.

Experimental results, for each data set, are represented separately for all methods in Figure 5; the graphical representations of average recognition accuracy and the recognition time are shown in Figures 6 and 7.

As can be seen from Figures 5 to 7, on the premise of the same feature extraction, the recognition accuracy of the SVM classifier is the lowest, and recognition time is the longest. The other two classifiers include the cascade dictionary classifier designed in this paper and the classifier designed to construct the dictionary with the monogenic features of the samples. Their recognition rates are obviously higher than that designed by the dictionary directly generated with the image pixels. The recognition accuracy of the proposed method, method 3, and method 4 are all higher than that of SVM and KNN methods.

A recognition method can not be considered superior to other methods if it requires a great deal of time to yield relatively small improvements. It can be seen from Figures 6 and 7 that by comparing the results of method 4 with that of the proposed method, the recognition accuracy is slightly lower than that of the method 4. However, the average recognition time of method 4 is 19.14 seconds, while that of the proposed method is only 10.75 seconds, the recognition speed is obviously improved. The main reason is that about 86% of the samples are correctly identified at the first recognition and do not need to pass the rest of the dictionaries. In method 4, all the testing samples need to do transvection with all the atoms in the overcomplete dictionary, which obviously slows down the recognition speed.

5. Conclusion

In order to solve the problem of higher dimension and larger redundancy in constructing the dictionary with the training samples, this paper proposed a SAR image target recognition method based on monogenic signal and sparse representation. The main innovation points of this method is to train

TABLE 2: The recognition accuracy of the proposed method.

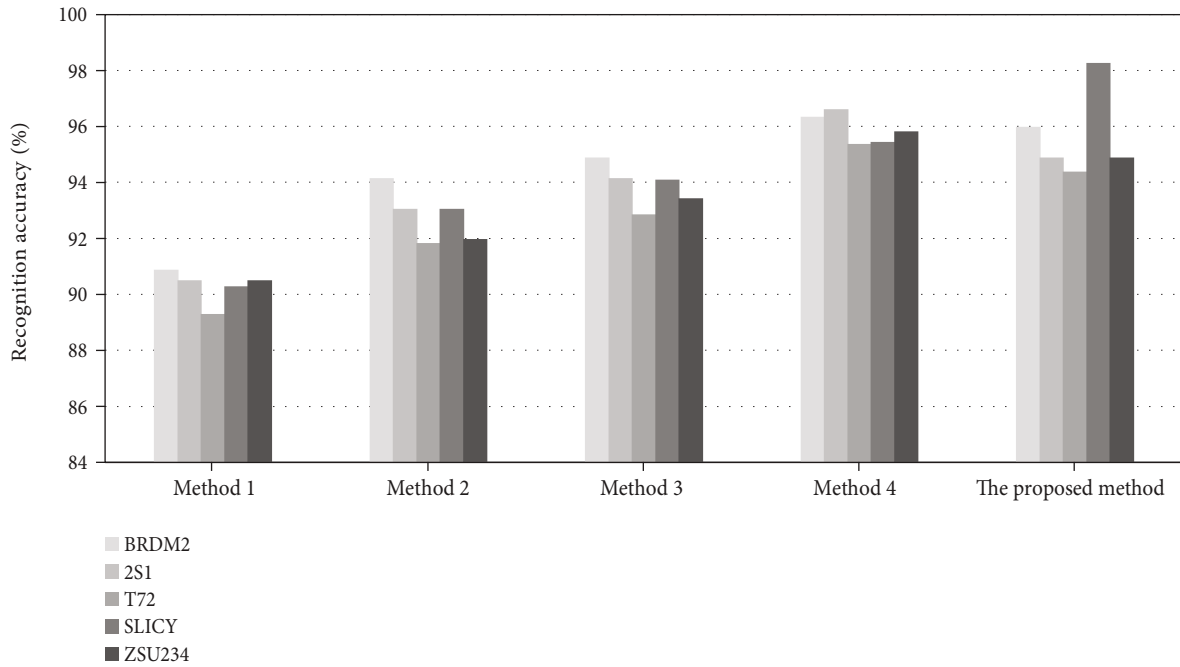|	BRDM2	2S1	T72	SLICY	ZSU234
Total number of samples	274	274	196	288	274
Correct sample number of primary recognition	252	218	165	278	250
Primary recognition accuracy (%)	91.97	79.56	84.18	96.53	91.24
Correct sample number of voting recognition	11	42	20	5	10
Voting recognition accuracy (%)	4.01	15.33	10.20	1.73	3.65
Recognition accuracy (%)	95.98	94.89	94.38	98.26	94.89

FIGURE 5: Recognition accuracy of different recognition methods.
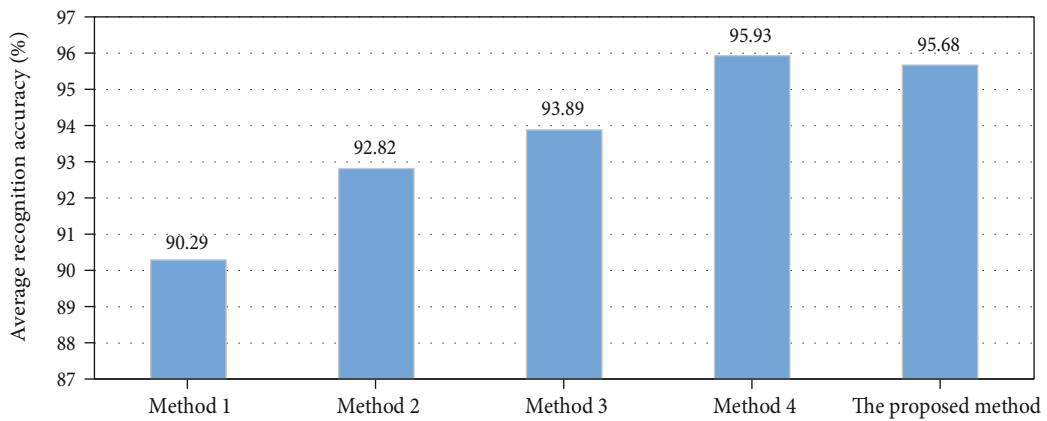

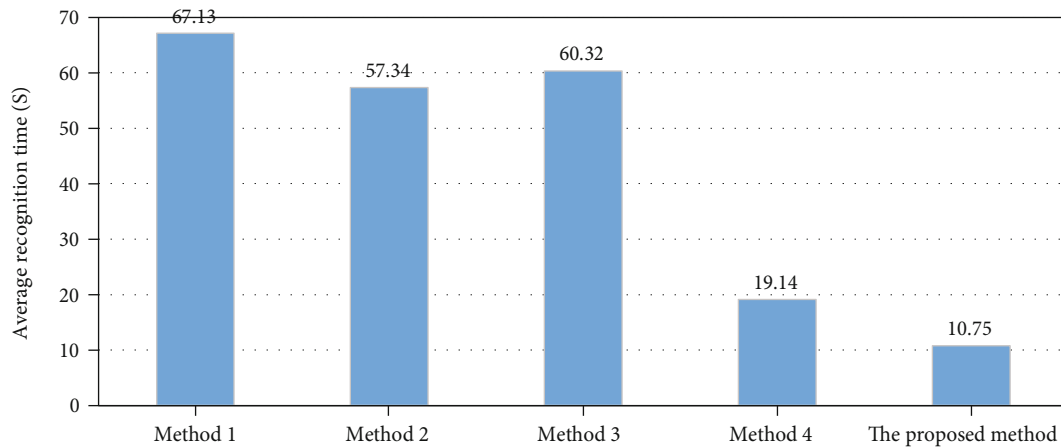FIGURE 6: Recognition accuracy of different recognition methods.

FIGURE 7: Recognition time of different recognition methods.

templates using EMACH filter, generate templates, construct cascade dictionary with the monogenic features of various templates, solve sparse representation coefficients of the testing samples with the proved orthogonal matching pursuit algorithm at each level of dictionary, and determine the category of the testing samples by the reconstruction error of the image. Through experimental results, we have demonstrated that the proposed SAR image target recognition method significantly improved the recognition speed in the precondition of ensuring the recognition rate. Comparison with the traditional dictionary directly generated by G2DPCA feature based on sparse representation indicated that the proposed SAR image recognition method was able to improve the overall accuracy by up to 0.65% and shorten recognition time by up to 43.83%.

Data Availability

The experimental data used herein is from the SAR ground still data of the SAR ground target high resolution provided by the Working Group of the United States DARPA/AFRL MSTAR.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] X. Wang, M. Huang, and L. Wan, "Joint 2D-DOD and 2D-DOA estimation for co-prime EMVS-MIMO radar," *Circuits, Systems, and Signal Processing*, vol. 1, pp. 1–17, 2021.

[2] T. Liu, F. Wen, L. Zhang, and K. Wang, "Off-grid DOA estimation for colocated MIMO radar via reduced-complexity sparse Bayesian learning," *IEEE Access*, vol. 7, pp. 99907–99916, 2019.

[3] L. Wan, X. Kong, and F. Xia, "Joint Range-Doppler-Angle estimation for intelligent tracking of moving aerial targets," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1625–1636, 2018.

[4] A. S. Alphonse and D. Dharma, "A novel monogenic directional pattern (MDP) and pseudo-voigt kernel for facilitating the identification of facial emotions," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 459–470, 2017.

[5] X.-G. Luo, H.-J. Wang, and S. Wang, "Monogenic signal theory based feature similarity index for image quality assessment," *AEU - International Journal of Electronics and Communications*, vol. 69, no. 1, pp. 75–81, 2015.

[6] S. Gai, B. Zhang, C. Yang, and L. Yu, "Speckle noise reduction in medical ultrasound image using monogenic wavelet and Laplace mixture distribution," *Digital Signal Processing*, vol. 72, pp. 192–207, 2018.

[7] Z. Sun, Y. Qiao, B. P. F. Lelieveldt, and M. Staring, "Integrating spatial-anatomical regularization and structure sparsity into SVM: Improving interpretation of Alzheimer's disease classification," *NeuroImage*, vol. 178, pp. 445–460, 2018.

[8] B. Wang, Y. L. Huang, J. Y. Yang, and J. J. Wu, "A Feature extraction method for synthetic aperture radar (SAR) automatic target recognition based on maximum interclass distance," *Science China (Technological Sciences)*, vol. 54, no. 9, pp. 2520–2524, 2011.

[9] H. Xu and G. Srivastava, "Automatic recognition algorithm of traffic signs based on convolution neural network," *Multimedia Tools and Applications*, vol. 79, no. 17-18, pp. 11551–11565, 2020.

[10] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Roubust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[11] Y. Su, X. Gao, and X.-C. Yin, "Fast alignment for sparse representation based face recognition," *Pattern Recognition*, vol. 68, pp. 211–221, 2017.

[12] M. Khorasani, Brain Engineering Research Center, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran, S. Ghofrani, and M. Hazari, "Sparse representation and face recognition," *International Journal of Image Graphics & Signal Processing*, vol. 10, no. 12, pp. 11–20, 2018.

[13] M. Liao and X. Gu, "Face recognition approach by subspace extended sparse representation and discriminative feature learning," *Neurocomputing*, vol. 373, no. 15, pp. 35–49, 2020.

[14] X. Wang, L. Wang, X. Li, and G. Bi, "Nuclear norm minimization framework for DOA estimation in MIMO radar," *Signal Processing*, vol. 135, pp. 147–152, 2017.

[15] J. Shi, F. Wen, and T. Liu, "Nested MIMO radar: coarrays, tensor modeling and angle estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 1, p. 99, 2020.

[16] F. Wen and J. Shi, "Fast direction finding for bistatic EMVS-MIMO radar without pairing," *Signal Processing*, vol. 173, p. 107512, 2020.

[17] G. Dong and G. Kuang, "Classification on the monogenic scale space: application to target recognition in SAR image," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2527–2539, 2015.

[18] G. Dong, H. Liu, G. Kuang, and J. Chanussot, "Target recognition in SAR images via sparse representation in the frequency domain," *Pattern Recognition*, vol. 96, pp. 1–10, 2019.

[19] H. J. Qi, Y. H. Wang, and J. Ding, "SAR target recognition based on multi-information dictionary learning and sparse representation," *Systems Engineering and Electronics*, vol. 37, no. 6, pp. 1280–1287, 2015.

[20] X. X. Ji and G. Zhang, "SAR images target recognition based on sparse representation," *Sensors & Transducers*, vol. 161, no. 12, pp. 67–73, 2013.

[21] X. Ji and G. Zhang, "Contourlet domain SAR image despeckling via self-snake diffusion and sparse representation," *Multimedia Tools and Applications*, vol. 76, no. 4, pp. 5873–5887, 2017.

[22] Y. Chen, M. L. Wong, and H. Li, "Applying ant colony optimization to configuring stacking ensembles for data mining," *Expert Systems with Applications*, vol. 41, no. 6, pp. 2688–2702, 2014.

WILEY | Hindawi

Research Article

Adaptive Reconstruction Algorithm Based on Compressed Sensing Broadband Receiver

Wei-Jian Si,[1,2] Qiang Liu [1,2] and Zhi-An Deng [1,2]

[1]*College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China*
[2]*Key Laboratory of Advanced Marine Communication and Information Technology, Ministry of Industry and Information Technology, Harbin 150001, China*

Correspondence should be addressed to Zhi-An Deng; dengzhian@hrbeu.edu.cn

Existing greedy reconstruction algorithms require signal sparsity, and the remaining sparsity adaptive algorithms can be reconstructed but cannot achieve accurate sparsity estimation. To address this problem, a blind sparsity reconstruction algorithm is proposed in this paper, which is applied to compressed sensing radar receiver system. The proposed algorithm can realize the estimation of signal sparsity and channel position estimation, which mainly consists of two parts. The first part is to use fast search based on dichotomy search, which is based on the high probability reconstruction of greedy algorithm, and uses dichotomy search to cover the number of sparsity. The second part is the signal matching and tracking algorithm, which is mainly used to judge the signal position and reconstruct the signal. Combine the two parts together to realize the blind estimation of the sparsity and the accurate estimation of the number of signals when the number of signals is unknown. The experimental analyses are carried out to evaluate the performance of the reconstruction probability, the accuracy of sparsity estimation, the running time of the algorithm, and the signal-to-noise ratio.

1. Introduction

With the continuous development of radar technology, the electronic countermeasure environment has become increasingly complex, and it will become more complex in the future [1]. First, the analog-to-digital converter (ADC) sampling rate limits the instantaneous bandwidth of the receiver [2]. The use of frequency agile signals requires radar receivers to have increasingly wider instantaneous bandwidths. At present, the development of high-rate ADCs is relatively slow [3–5]. Second, in order to achieve full-probability reception, radar receivers require increasingly hardware, the system is more and more complex, and the volume is getting larger. In order to meet the needs of modern electronic warfare, it is necessary to solve the problems faced by digital channelized receivers [6]. Donoho, Candes et al., Wan et al., and Candes and Tao proved the theory of compressed sensing in 2006 and proposed specific implementation algorithms [7–10]. Compressed sensing is a new sampling method, which is highly concerned in image processing, microwave imaging, wireless communication, radar system, and biomedical engineering [11–18].

In recent years, an analog-to-information converter [19] (AIC) has been proposed based on the compressed sampling theory. The methods to achieve AIC include random demodulation (RD) [20], multicoset sampling (MCS) [21], quadrature compressed sampling (QuadCS) [22], and modulated wideband converter (MWC) [23] structure. The under-Nyquist sampling of the sparse signal can be realized at a sampling rate much lower than the Nyquist theorem, and the compressed sampled data of the signal can be obtained. The modulation broadband converter is currently the most mature multibranch compression sampling structure at. Cohen et al. and Mishali et al. have completed the realization of the hardware structure [24, 25].

Compressed sensing theory is mainly divided into three parts: signal sparse representation, observation matrix, and reconstruction algorithm. Signal sparse representation means that the signal can be sparsely represented in a certain transform domain. The observation matrix reduces the

dimensionality of the signal through domain conversion. Finally, it restores the original signal through a reconstruction algorithm [26]. The reconstruction algorithm is the most important step in compressed sensing. The algorithm can be divided into the following three categories: combination algorithm, convex optimization algorithm, and greedy matching pursuit algorithm [27]. The greedy matching pursuit algorithm has been widely studied due to its fast reconstruction speed and small calculation amount.

Existing greedy matching pursuit algorithms mainly include two types. The first category is greedy matching pursuit algorithms that require sparsity as a priori information, mainly including matching pursuit (MP) [28], orthogonal matching pursuit (OMP) [29], regularized orthogonal matching pursuit (ROMP) [30], compressed sampling matching pursuit (CoSaMP) [31], subspace pursuit (SP) [32], and generalized orthogonal matching pursuit (gOMP) [33] algorithms. The OMP algorithm is one of the simplest matching pursuit algorithms. In each iteration, the atom with the largest inner product of the observation matrix and the residual is selected and added to the support set. Use the least square method to update the residuals. Make the residual continuously approach a certain value to achieve convergence. The OMP algorithm guarantees the orthogonality of iterations, but cannot guarantee the correctness of atoms. So once the wrong atom is introduced, it will affect subsequent iterations. The CoSaMP algorithm adds backtracking ideas when selecting atoms. Each time an atom is selected, not only the atom is added to the support set, but also on the wrong atom to delete. Although the CoSaMP algorithm can improve the reconstruction probability and quickly complete the iterations, as the algorithm increases in sparsity, the reconstruction probability will drop sharply. The gOMP algorithm finds multiple support sets each time, the method of finding at most K times improves the probability of reconstruction. However, the sparsity of the algorithm is greater than K after the iteration.

The other is the adaptive algorithm that does not require sparsity, mainly including stagewise orthogonal matching pursuit (StOMP) [34], stagewise weak orthogonal matching pursuit (SWOMP) [35], and sparsity adaptive matching pursuit (SAMP) [36] algorithms. There are three problems in the aforementioned sparsity adaptive matching pursuit algorithm. The first problem is that the StOMP and SWOMP algorithms' support concentration does not consider the selection of the wrong atom, which leads to the limitation of the reconstruction probability of the algorithm [37]. The second problem is that although the SAMP algorithm has a high reconstruction probability, it cannot accurately estimate the sparsity of the signal during adaptive estimation due to the limitation of the step size [38]. The third problem is that SWOMP algorithm and SAMP algorithm are also susceptible to noise in the signal. Therefore, it is necessary to design a sparsity adaptive matching tracking algorithm that can accurately estimate the signal sparsity and high reconstruction probability in noisy radar signals.

This paper proposes a sparseness adaptive matching tracking algorithm for radar receiver system based on compressed sensing. Aiming at the problem of low reconstruction probability of greedy matching pursuit algorithm. A backtracking-based orthogonal matching pursuit BOMP algorithm is designed using backtracking ideas. In each iteration, the wrong atoms selected in the support set will be eliminated. Compared with other greedy algorithms that require sparsity, it has a higher reconstruction probability. Aiming at the problem of sparsity estimation error. Utilizing the principle of binary search, we devised a method sparsity search. First, set an estimated value of sparsity continuously and change the estimated value of sparsity through judgment criteria. Finally, get an accurate estimated sparsity. The estimated sparsity values are all based on high probability reconstruction. At the same time, due to the characteristics of binary search, it can quickly find the true value within a limited number of steps. For the noise problem in the signal, joining residuals associated with the threshold, the ability dynamically adapts to noise. The matching pursuit algorithm proposed in this paper is compared with similar sparsity adaptive algorithms, which improved the accuracy of output signal-to-noise ratio, reconstruction probability, and sparsity estimation.

The arrangement of this article is as follows. In the second section, the system model of the compressed sensing broadband receiver based on MWC is introduced. In the third section, an orthogonal matching pursuit algorithm based on backtracking is designed, which improves the reconstruction probability. In the fourth section, a sparsity adaptive reconstruction algorithm based on compressed sensing is proposed. In the fifth section, the performance of the algorithm is verified by simulation. Compare the reconstruction probability, sparsity, and running time of several adaptive algorithms under one-dimensional signal. Under the radar signal, the algorithm SNR and sparsity estimation and reconstruction probability are also analyzed. Finally, our conclusions are given in the sixth section.

2. System Model

In this section, the system model of MWC compressed sampling structure and sparse reconstruction algorithm will be introduced. The structure design and working principle of MWC compressed sampling wideband digital receiver are introduced, including the introduction of mathematical expressions and a brief description of the compressed sensing reconstruction algorithm.

2.1. MWC Structure. The radar signal in space is often continuous. In this article, for the convenience of introduction, the signal is set as a discrete signal under the Nyquist sampling rate. This setting will not affect the result. The received signal $x[n]$ is first mixed with the pseudorandom sequence $\tilde{p}_m[n]$ to obtain the mixed signal $\tilde{x}_m[n]$, and the mixed signal is then passed through the corresponding low-pass filter $h[n]$ to obtain the filtered signal $w_m[n]$. Finally, the signal is extracted by M_p time and sent to the low-speed ADC for samples to get $y_m[k]$. The discrete compressed sampling structure based on MWC receiver is shown in Figure 1.
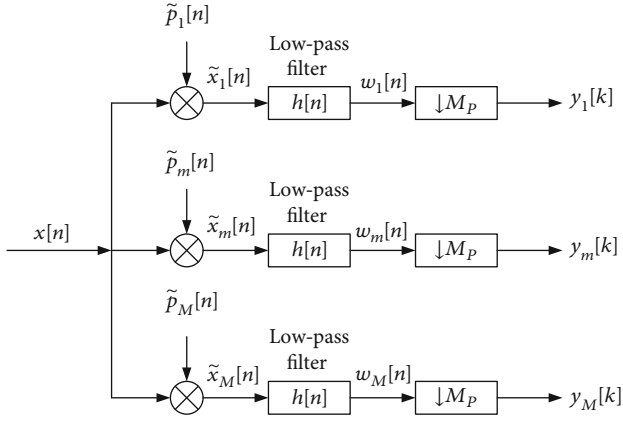
FIGURE 1: Block diagram of MWC.

The discrete time domain of the radar received signal can be expressed as follows:

$$x[n] = s[n] + \eta[n], 0 \le n \le N. \tag{1}$$

Among them, $s[n]$ is the effective signal component, $\eta[n]$ is the Gaussian white noise component in the signal, and N represents the length of the signal. $x[n]$ is a complex discrete signal; in the Nyquist sampling rate $f_{NYQ} = 1/T_{NYQ}$, T_{NYQ} represents the period of Nyquist sampling, and the bandwidth range is $\mathscr{F}_{NYQ} \triangleq [-f_{NYQ}/2, f_{NYQ}/2]$. The discrete-time Fourier transform (DTFT) is $x[n]$ and expressed as follows:

$$X\left(e^{j2\pi fT_{NYQ}}\right) = \sum_{n=0}^{N-1} x[n]e^{-j2\pi fnT_{NYQ}}. \tag{2}$$

It can be seen from Figure 1 that the structure contains M parallel branches. Taking the mth path as an example, the period of a periodic pseudorandom sequence $p_m[n]$ is T_p, each period contains $M_p = T_p f_{NYQ}$ elements. The sequence $p_m[n]$ can be expressed as follows:

$$p_m[n] = \begin{cases} \tilde{p}_m[n] & 0 \le n \le M_p - 1 \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

The discrete Fourier series of the sequence $\tilde{p}_m[n]$ can be expressed as follows:

$$\tilde{P}_m[n] = \frac{1}{M_p} \sum_{l=0}^{M_p-1} P_m(l)e^{j(2\pi/M_p)nl}. \tag{4}$$

l represents the index number and $0 \le l \le M_p - 1$, and $P_m(l)$ is the discrete Fourier transform coefficient of the sequence $p_m[n]$. The mixing rate of the signal and the sequence is $f_p = 1/T_p = f_{NYQ}/M_p$. The sequence $p_m[n]$ will be divided $\mathscr{F}_{NYQ}$ into M_p segments. The interval of each segment is the same f_p. The position of each segment is called an index. The range of the baseband spectrum segment is

$\mathscr{F}_p \triangleq [-f_p/2, f_p/2]$. In order to avoid edge effects, there is a need to meet $f_p \ge B$, where B represents the bandwidth of $s[n]$. The discrete-time Fourier transform (DTFT) of the mixed signal $\tilde{x}_m[n]$ is expressed as follows:

$$\begin{aligned} \tilde{X}_m\left(e^{j2\pi fT_{NYQ}}\right) &= \sum_{n=0}^{N-1} x[n] \cdot \tilde{p}_m[n]e^{-j2\pi fnT_{NYQ}} \\ &= \sum_{n=0}^{N-1} x[n] \cdot \frac{1}{M_p} \sum_{l=0}^{M_p-1} P_m(l)e^{j(2\pi/M_p)nl}e^{-j2\pi fnT_{NYQ}} \\ &= \frac{1}{M_p} \sum_{l=0}^{M_p-1} P_m(l)X\left(e^{j2\pi fT_{NYQ}(f-lf_p)}\right). \end{aligned} \tag{5}$$

The mixed signal $\tilde{x}_m[n]$ is filtered by a low-pass filter $h[n]$. The frequency response is $H(e^{-j2\pi fT_{NYQ}})$. The cutoff frequency of the low-pass filter is $f_p/2$. The discrete-time Fourier transform (DTFT) filtered signal $w_m[n]$ is expressed as follows:

$$\begin{aligned} W_m\left(e^{j2\pi fT_{NYQ}}\right) &= \sum_{n=0}^{N-1} \tilde{x}_m[n] \cdot h[n]e^{-j2\pi fnT_{NYQ}} \\ &= \tilde{X}_m\left(e^{j2\pi fT_{NYQ}}\right)H\left(e^{j2\pi fT_{NYQ}}\right) \\ &= \begin{cases} \frac{1}{M_p} \sum_{l=0}^{M_p-1} P_m(l)X\left(e^{j2\pi fT_{NYQ}(f-lf_p)}\right), f \in F_p \\ 0, f \notin F_p \end{cases}. \end{aligned} \tag{6}$$

After the signal passes through the low-pass filter, the spectral components are within $\mathscr{F}_p \triangleq [-f_p/2, f_p/2]$. But the data rate of the signal is still f_{NYQ}. The data rate can be reduced by decimating the filtered signal. Use low-speed ADC sampling to complete the decimation of the signal. Decimation factor is $M_p = f_{NYQ}/f_p$. The discrete-time Fourier transform (DTFT) of the sampled signal $y_m[k]$ is expressed as follows:

$$\begin{aligned} Y_m\left(e^{j2\pi fT_s}\right) &= \sum_{k=0}^{K-1} y_m[k]e^{-j2\pi fnT_s} \\ &= \frac{1}{M_p} \sum_{l=0}^{M_p-1} P_m(l)X\left(e^{j2\pi fT_{NYQ}(f-lf_p)}\right), f \in F_s. \end{aligned} \tag{7}$$

The output can be written in the form of a matrix; the expression is as follows:

$$y(f) = Cz(f), f \in F_s, \tag{8}$$

where $y(f)$ is a column vector of length M. The mth element is $Y_m(e^{j2\pi fT_s})$, $z(f)$ is a column vector of length M_p, and

the lth element is $X(e^{j2\pi f T_{NYQ}}(f-lp))$. The linear combination can be expressed as follows:

$$P_m(l) = \frac{1}{M_p} \sum_{l=0}^{M_p-1} p_m[n]e^{-j(2\pi/M_p)nl}. \tag{9}$$

The observation matrix C is $M \times M_p$ dimensional and can be expressed as follows:

$$C = \frac{PF}{M_p}, \tag{10}$$

where P represents $M \times M_p$ dimensional random matrix, F represents $M_p \times M_p$ dimensional discrete Fourier transform matrix, $F_l = [1, e^{j2\pi l/M_p}, \cdots, e^{j2\pi(M_p-1)l/M_p}]^T$ is a column vector, and $(\bullet)^T$ represents transpose. The expanded form of the formula can be obtained.

$$\begin{pmatrix} Y_1\left(e^{j2\pi f T_s}\right) \\ Y_2\left(e^{j2\pi f T_s}\right) \\ \vdots \\ Y_M\left(e^{j2\pi f T_s}\right) \end{pmatrix} = \frac{1}{M_p} \begin{bmatrix} p_{1,0} & \cdots & p_{1,M_p-1} \\ \vdots & \ddots & \vdots \\ p_{M,0} & \cdots & p_{M,M_p-1} \end{bmatrix}$$

$$\cdot \begin{bmatrix} | & \cdots & | & \cdots & | \\ F_0 & \cdots & F_l & \cdots & F_{M_p-1} \\ | & \cdots & | & \cdots & | \end{bmatrix} \tag{11}$$

$$\times \begin{bmatrix} X\left(e^{j2\pi f T_{NYQ} f}\right) \\ \vdots \\ X\left(e^{j2\pi f T_{NYQ}(f-lf_p)}\right) \\ \vdots \\ X\left(e^{j2\pi f T_{NYQ}(f-(M_p-1)f_p)}\right) \end{bmatrix}.$$

The above formula is the frequency domain model of MWC compressed sensing receiver, and the feasibility of this structure is proved through mathematical deduction. Therefore, the compressed sensing of signals can be realized through the MWC structure.

2.2. Reconstruction Algorithm. The sampled signal model obtained by using the MWC structure is the infinite measurement vector (IMV) model. In theory, it takes countless iterations to recover the original signal. Xu et al. proposed continuous finiteness in literature [20]. The continuous-to-finite (CTF) algorithm can convert the IMV model of the MWC compressed sample signal into a multiple measurement vector (MMV) model. The signal under this model can be reconstructed from the original signal through the tra-

ditional reconstruction algorithm structure. The block diagram of CTF algorithm implementation is shown in the figure.

As can be seen from Figure 2, first, use the sampled signal $y[k]$ to get the matrix Q, and the expression is as follows:

$$Q = \int_{f \in F_s} y(f)y^H(f)df = \sum_{k=0}^{K} y(k)y^T(k), \tag{12}$$

which $(\bullet)^H$ represents the conjugate transpose. Then, decompose Q to get frame V. The expression is as follows:

$$Q = VV^H. \tag{13}$$

Obtain the joint support set S according to the sparse reconstruction of the vector V. The greedy matching pursuit algorithm can find the index set S where the signal exists. Using the sampling signal and the observation matrix, the time-domain waveform $z_s[n]$ of each channel in the baseband can be obtained. The expression is as follows:

$$\hat{z}_i[n] = \begin{cases} A_i^\dagger y[n] & i \in S, \\ 0 & i \notin S, \end{cases} \tag{14}$$

where $(\bullet)^\dagger$ represents the pseudoinverse. The reconstructed signal $\hat{x}[n]$ can be finally obtained. The time domain expression is as follows:

$$\hat{x}[n] = \hat{x}[nT_{NYQ}] = \sum_{i \in S} (\tilde{z}_i[n] * h_i[n])e^{j2\pi i f_p n T_{NYQ}}. \tag{15}$$

$h[n]$ is an ideal interpolation filter, and the rate is the Nyquist sampling rate f_{NYQ}. The greedy algorithm part will be studied below.

3. BOMP

This section proposes a backtracking orthogonal match pursuit (BOMP) algorithm. The algorithm mainly improves the OMP algorithm by three parts: threshold filtering, maximum selection, and backtracking idea. First, threshold filtering is used to calculate the inner product of the residual and the matrix. Filter out all atoms whose inner product is greater than the threshold u_t and add these atoms to the support set. Secondly, the backtracking idea is used to find the least square solution using the support set. Through backtracking idea, only some atoms in the support set are selected to form a new support set. Finally, use the maximum selection. In order to ensure that the number of atoms' final output is the same as the sparsity, t atoms are selected to form a new support set during the tth backtracking. It is equivalent to only adding one atom for each cycle of support concentration. After the cycle ends, there are only t atoms in the support set. Specific steps are as follows:

As seen in Algorithm 1, we need to enter $\mathbf{A}$, y, and K. Find the atom whose inner product is greater than the threshold u_t each time. Combine these atoms with the support set Λ_{t-1} obtained in the previous cycle to obtain a new
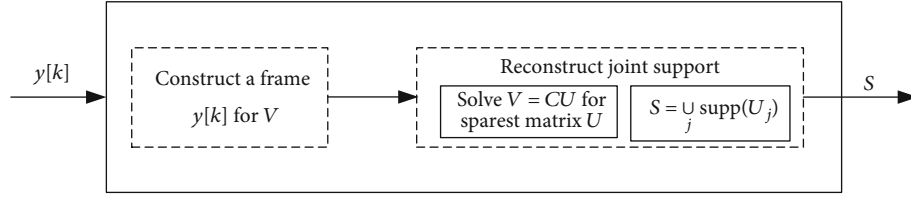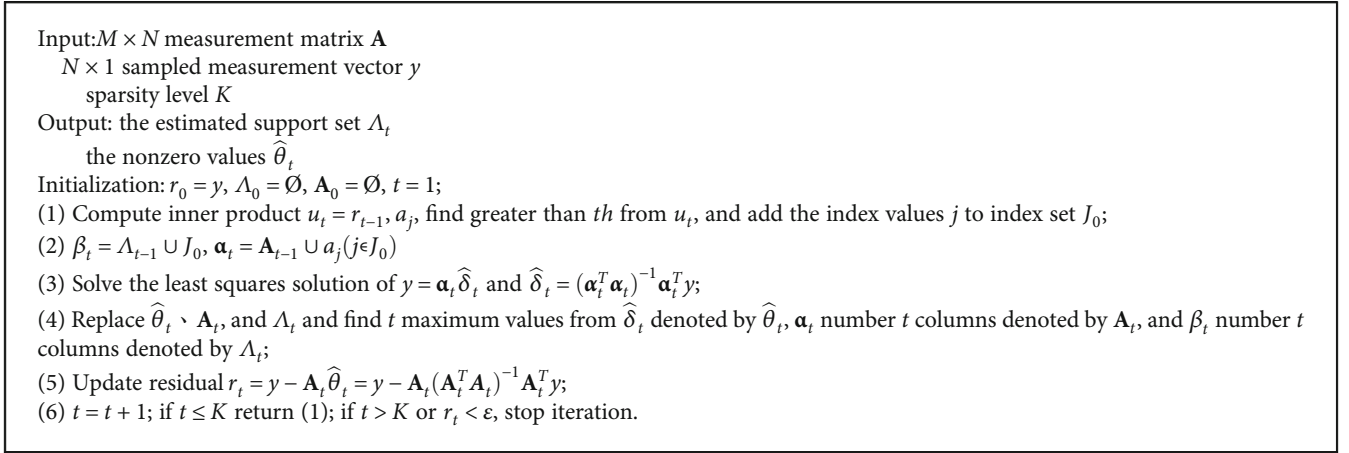
FIGURE 2: Block diagram of CTF algorithm.

Input: $M \times N$ measurement matrix $\mathbf{A}$
   $N \times 1$ sampled measurement vector y
     sparsity level K
Output: the estimated support set Λ_t
     the nonzero values $\widehat{\theta}_t$
Initialization: $r_0 = y$, $\Lambda_0 = \varnothing$, $\mathbf{A}_0 = \varnothing$, $t = 1$;
(1) Compute inner product $u_t = r_{t-1}, a_j$, find greater than th from u_t, and add the index values j to index set J_0;
(2) $\beta_t = \Lambda_{t-1} \cup J_0$, $\boldsymbol{\alpha}_t = \mathbf{A}_{t-1} \cup a_j(j\epsilon J_0)$
(3) Solve the least squares solution of $y = \boldsymbol{\alpha}_t \widehat{\delta}_t$ and $\widehat{\delta}_t = (\boldsymbol{\alpha}_t^T \boldsymbol{\alpha}_t)^{-1} \boldsymbol{\alpha}_t^T y$;
(4) Replace $\widehat{\theta}_t$ 、 $\mathbf{A}_t$, and Λ_t and find t maximum values from $\widehat{\delta}_t$ denoted by $\widehat{\theta}_t$, $\boldsymbol{\alpha}_t$ number t columns denoted by $\mathbf{A}_t$, and β_t number t columns denoted by Λ_t;
(5) Update residual $r_t = y - \mathbf{A}_t \widehat{\theta}_t = y - \mathbf{A}_t(\mathbf{A}_t^T \mathbf{A}_t)^{-1} \mathbf{A}_t^T y$;
(6) $t = t + 1$; if $t \leq K$ return (1); if $t > K$ or $r_t < \varepsilon$, stop iteration.

ALGORITHM 1: BOMP algorithm.

support set β_t. Use the matrix and residuals corresponding to the support set to find the least squares solution of $\widehat{\delta}_t$. Replace $\widehat{\theta}_t$, $\mathbf{A}_t$, and Λ_t and find the residual r_t according to $\widehat{\theta}_t$ and $\mathbf{A}_t$. Finally, judge whether the iteration is completed through t and r_t. Set $\varepsilon = 10^{-6}$. Set the threshold th $= t_s \|r_s\|_2 / \|y\|_2$, where r_s is the residual calculated last time, $\|\bullet\|_2$ is the 2 norms. $t_s \geq 2$ because the found atom needs to be greater than twice the average.

In order to verify the reconstruction probability of the BOMP algorithm, choose OMP, CoSaMP, and gOMP algorithms to compare with the algorithm proposed in this paper. The original signal x uses a Gaussian random signal, and the signal length is $N = 256$. The observation matrix $\boldsymbol{\Phi}$ is a Gaussian random matrix, and the observation value $M = 128$. The interval of sparsity K is [35, 70], and the step size is 5. Each sparsity is measured 1000 times. The same signal is reconstructed by four algorithms. The reconstruction probability under different sparsity is shown in Figure 3.

As shown in Figure 3, when the sparsity is greater than 30, the reconstruction probability of the OMP algorithm begins to gradually decrease. The reconstruction probability of CoSaMP algorithm, gOMP algorithm, and BOMP algorithm is still good. When the sparsity is greater than 50, the reconstruction probability of CoSaMP algorithm, gOMP algorithm, and BOMP algorithm begins to decrease. The reconstruction probability of BOMP algorithm is about 15% higher than that of gOMP algorithm. It can be seen that the BOMP algorithm improves the reconstruction probability. Regarding the reconstruction probability, the CoSaMP algorithm eliminates the wrong atom every time the atom
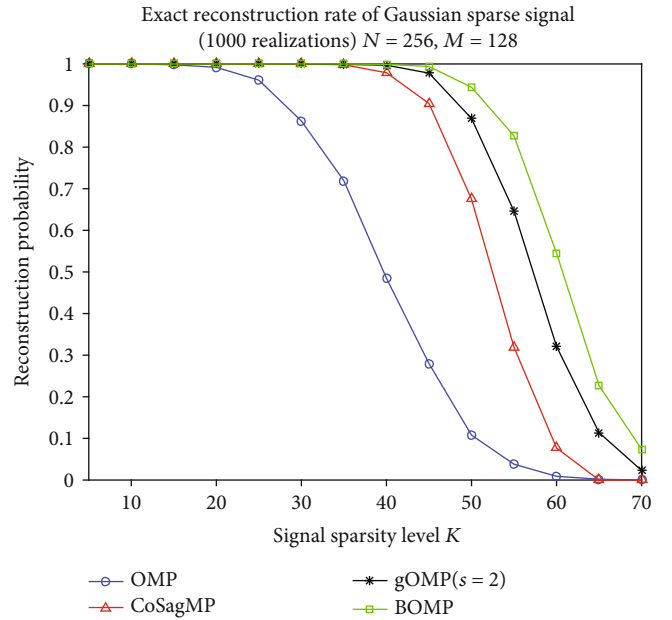

FIGURE 3: Reconstruction probability diagram under different sparsity.

is selected to prevent the wrong atom from affecting the selection of subsequent atoms, so the reconstruction probability is better than the OMP algorithm. The gOMP algorithm selects multiple atoms each time. According to the conditions for stopping iteration, it can be known that theoretically at most SK atoms can be selected. The correct atom

---

Input: $M \times N$ measurement matrix $\mathbf{A}$
  $N \times 1$ sampled measurement vector y
Output: sparsity level t
    the estimated support set Λ_t
    the nonzero values $\widehat{\theta}_t$
Initialization: $i = 0, n = <\log_2 M - 1 > (<\bullet>\text{is round}), K_0 = 2^{n-1}$
(1) $i = i + 1, r_0 = y, \Lambda_0 = \emptyset, A_0 = \emptyset, t = 1$;
(2) Compute inner product $u_t = r_{t-1}, a_j$, find greater than th from u_t, and add the index values j to index set J_0;
(3) $\beta_t = \Lambda_{t-1} \cup J_0, \mathbf{\alpha}_t = \mathbf{A}_{t-1} \cup a_j (j \epsilon J_0)$; if $\beta_t = \Lambda_{t-1}$, stop iteration; otherwise, solve the least squares solution of $y = \mathbf{\alpha}_t \widehat{\delta}_t, \widehat{\delta}_t = (\mathbf{\alpha}_t^T \mathbf{\alpha}_t)^{-1} \mathbf{\alpha}_t^T y$;
(4) Replace $\widehat{\theta}_t \diagdown \mathbf{A}_t$, and Λ_t and find t maximum values from $\widehat{\delta}_t$ denoted by $\widehat{\theta}_t$, $\mathbf{\alpha}_t$ number t columns denoted by $\mathbf{A}_t$, and β_t number t columns denoted by Λ_t;
(5) Update residual $r_t = y - \mathbf{A}_t \widehat{\theta}_t = y - \mathbf{A}_t (\mathbf{A}_t^T \mathbf{A}_t)^{-1} \mathbf{A}_t^T y$;
(6) $t = t + 1$; if $t \leq K_0$ return (2); if $t > K_0$ or $r_t < \varepsilon$ proceed to (7);
(7) If $t < K_0$, stop iteration; if $i = n + 1$, stop iteration; if $i = n$, proceed to (8); if $t = K_0$ proceed to (9);
(8) If $r_t > \varepsilon$, $K_0 = K_0 + 1$, proceed to (1); otherwise, $K_0 = K_0$, stop iteration;
(9) If $r_t > \varepsilon$, $K_0 = K_0 + 2^{n-1-i}$, proceed to (1); otherwise, $K_0 = K_0 - 2^{n-1-i}$, proceed to (1).

ALGORITHM 2: DSAMP algorithm.

set is a subset of the supporting set, thus further improving the probability of reconstruction. For the BOMP algorithm, it is finally achieved by selecting the best atom while continuously removing the wrong atom. Improve the probability of reconstruction.

4. DSAMP

This paper proposes a dichotomy-based sparsity adaptive matching pursuit (DSAMP) algorithm for compressed sensing broadband receivers. The algorithm mainly contains two parts. The first part is a fast atom search based on dichotomy. First, select the number of sparsity is 2^{n-1}, in order to achieve sparse coverage under high reconstruction probability. The number of atoms selected for the first time is n. After the preestimated sparsity, the unknown sparsity problem becomes a known sparsity problem. Perform the residual calculation on the basis of this sparsity. Use criteria to determine the calculation results, then use the dichotomy to change the preestimated sparsity according to the evaluation results. The iteration is repeated until the dichotomy satisfies the stop iteration condition or the residual calculation of the nth atomic selection is completed. The sparsity when the iteration is stopped is the sparsity finally estimated by the algorithm. Find the true sparsity through this method. The second part is the BOMP algorithm, which guarantees that the reconstruction algorithm under a given sparsity has a high reconstruction probability. At the same time, it can ensure that the sparsity is the same as the number of atoms in the support set.

Proposition 1. *For the greedy matching pursuit algorithm, the estimated sparsity K_0, the real sparsity K, when $K_0 < K$ and $K_0 \in K$, the residual $r_t > \varepsilon$; when $K_0 \geq K$ and $K \in K_0$, the residual $r_t \leq \varepsilon$.*

Proof. According to the principle of the greedy matching pursuit algorithm, r_t is monotonically decreasing when the atom

selection is correct in the iteration process, and then according to the iteration stop condition, it is obvious that the proposition is correct.

In Algorithm 2, for the recovery algorithm, the number of channels M must be at least twice the sparsity K to recover the original signal with high probability. Therefore, limit the sparsity range of the algorithm $[0, 2^n]$, where $n = <\log_2 M - 1 > (<\bullet>\text{is round})$. In the first cycle, the number of atoms is set to the midpoint of the entire range, which is 2^{n-1}. Steps (2)~(6) are BOMP algorithm, and it is mainly to calculate the residual under the current sparsity and iterate under the set sparsity. Steps (7)~(9) are to set K_0, and use Proposition 1 to set K_0. $t < K_0$ indicates that the algorithm has completed convergence ahead of time. The real sparsity is less than the sparsity; at this time, the number of cycles is the estimated sparsity. When $i = n+1$, it means that the last calculation is completed, the estimated sparsity and support set are found, and the iteration can be stopped. When $i = n$, the sparsity K_0 has been selected for the last time. Satisfying $r_t < \varepsilon$ indicates that the sparsity is the estimated sparsity. It can directly output all residuals, sparsity, and support set. When $r_t < \varepsilon$ is not satisfied, K_0 plus 1 is the estimated sparsity. It is necessary to recalculate the residual, sparsity, and support set under the new sparsity. When $t = K_0$, the algorithm has iterated K_0 times. $r_t > \varepsilon$ indicates that the algorithm has not reached the condition to stop iteration. It needs to update the estimated sparsity $K_0 = K_0 + 2^{n-1-i}$ and continue to iterate. In other cases, the iteration continues after updating the estimated sparsity $K_0 = K_0 - 2^{n-1-i}$.

5. Simulation Results

In this section, the performance of the algorithm is verified through simulation of the algorithm. This article simulates the algorithm in two situations. The first is to compare the performance of the algorithm proposed in this paper with several other algorithms in the case of one-dimensional
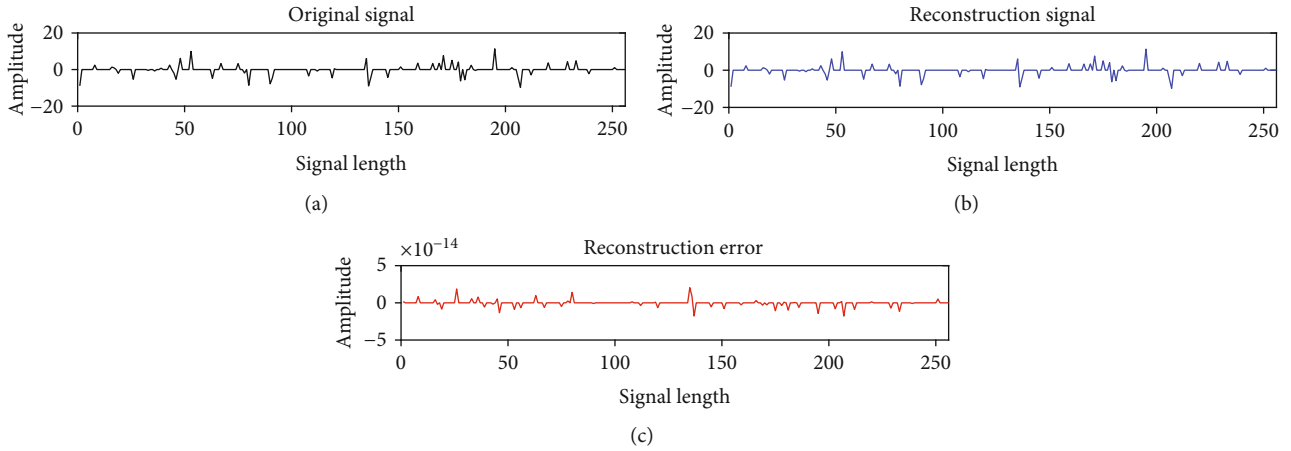
(a)

(b)

(c)

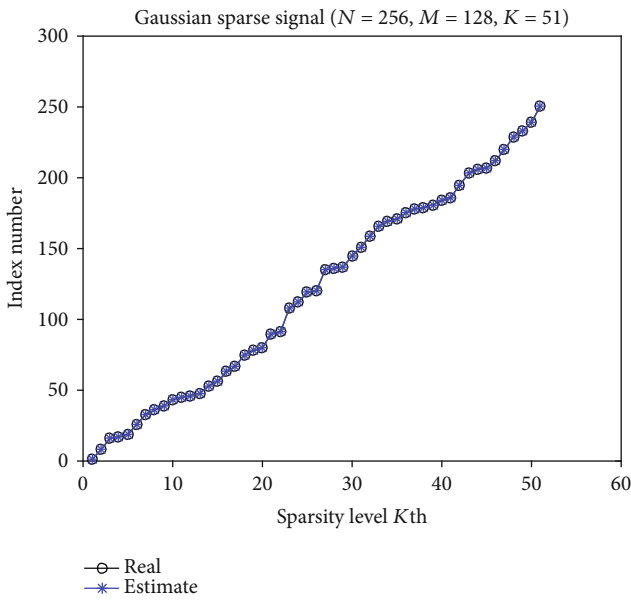FIGURE 4: One-dimensional signal reconstruction of BSAMP.

FIGURE 5: Sparse location diagram.

signals, mainly to verify the accuracy and running time of the algorithm for the estimation of reconstruction probability and sparsity. The other is to simulate under radar signals, which mainly verifies that the algorithm can adapt to multiple sparse signals under low signal-to-noise ratio and finally simulates the signal-to-noise ratio of several algorithms.

5.1. One-Dimensional Signal Simulation. This experiment mainly verifies the correctness of the one-dimensional signal, including signal reconstruction error and sparsity estimation. The signal x is a Gaussian random signal, the signal length is $N = 256$, the observation value is $M = 128$, and the sparsity is $K = 51$. The sparse locations are randomly selected. The observation matrix Φ is a Gaussian random matrix. Figure 4 is a one-dimensional signal reconstruction diagram of the DSAMP algorithm, including the original signal, reconstructed signal, and reconstruction error. Figure 5 is a

diagram of the true position and estimated position of the sparse signal in Figure 4.

As shown in Figure 4, it can be seen that the error between the reconstructed signal and the original signal is $\varepsilon_r < 1 \times 10^{-13}$. It can be seen that the algorithm can accurately reconstruct the original signal.

Figure 5 shows the position and estimated position of the sparse signal when the sparsity is $K = 51$, respectively. It can be seen that the estimated sparsity is the same as the real sparsity of the signal, and the estimated position and the real position are also the same. The algorithm can accurately estimate the sparsity and the location of sparsity.

5.2. Sparsity Estimation Probability Diagram. In this experiment, it is mainly compared with the sparsity adaptive algorithms such as StOMP, SWOMP, and SAMP to illustrate the advantages of the algorithm proposed in this paper. The signal x is a Gaussian random signal, the signal length is $N = 256$, the observation value is $M = 128$, the interval of sparsity K interval is $[1, 66]$, and the step size is 5. The sparse locations are randomly selected. The observation matrix Φ is a Gaussian random matrix. The sparse position is randomly selected and performs 1000 Monte Carlo experiments at each sparsity. The matrix is a Gaussian random matrix. Figure 6 is the reconstruction probability diagram of 1000 Monte Carlo experiments under different sparsity. Figure 7 is the average sparsity estimation diagram of 1000 Monte Carlo experiments under different sparsity. Figure 8 is the average running time diagram of 1000 Monte Carlo experiments under different sparsity.

As shown in Figure 6, when the sparsity is greater than 40, the reconstruction probability of the StOMP algorithm and the SWOMP algorithm begins to decrease. The reconstruction probability of SAMP algorithm and DSAMP algorithm is still 100%. When the sparsity is greater than 56, the StOMP and SWOMP algorithms cannot be reconstructed. The reconstruction probability of SAMP algorithm and DSAMP algorithm begins to decline, but the reconstruction probability of SAMP algorithm and DSAMP algorithm is similar.
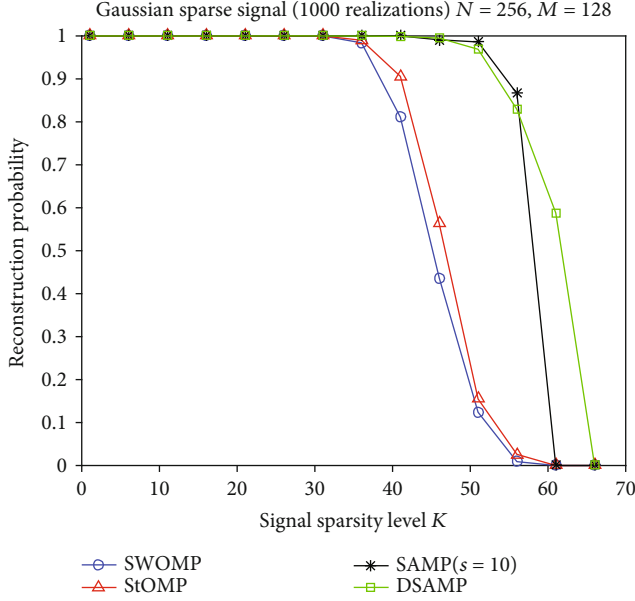
FIGURE 6: Reconstruction probability diagram under different sparsity.
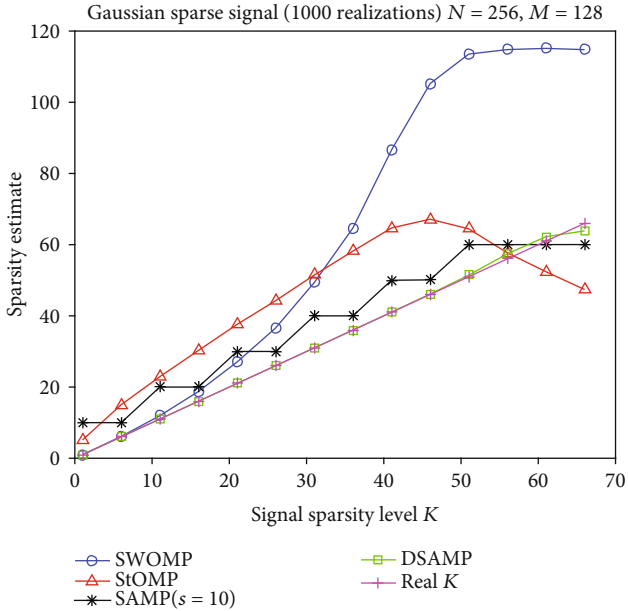

FIGURE 7: Estimated sparsity diagram under different sparsity.

FIGURE 8: Average running time graph under non-sparseness.

As shown in Figure 7, the real K represents the real sparsity set, and the remaining four lines represent the corresponding algorithms. When the sparsity is less than 45, the SWOMP algorithm can accurately estimate the sparsity when the sparsity is small, and as the sparsity increases, the error of the estimated sparsity number gradually increases. The number of sparsities estimated by the StOMP algorithm changes with the number of real sparsities, but there is an error with the real sparsity. When the sparsity is greater than 45, the sparsity estimated by the SWOMP algorithm has exceeded 100, which is relatively close to the observed value M. When the sparsity is greater than 45, the amount of spar-
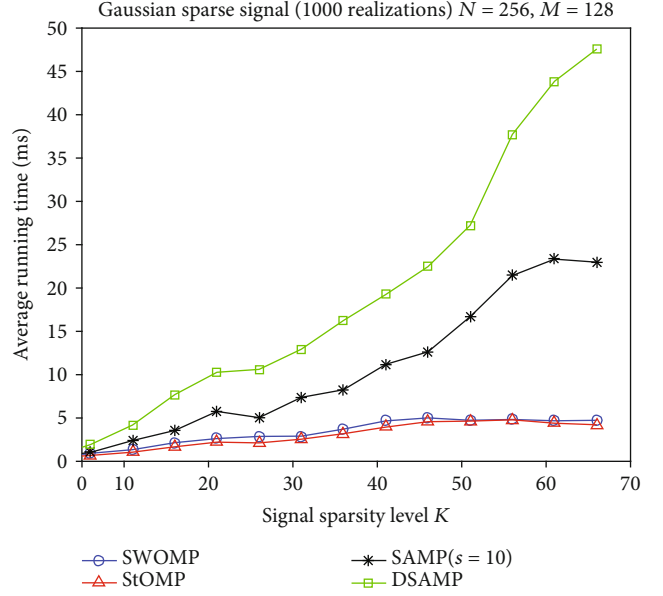
sities estimated by the StOMP algorithm continues to decrease. From Figure 6 can be seen that the reconstruction probability of these two algorithms is decreasing under this sparsity. When the sparsity is less than 51, due to the setting of the step size of the SAMP algorithm, the amount of sparsity estimates each time is an integer multiple of 10. The DSAMP algorithm can accurately estimate the number of signal sparsity. When the sparsity is greater than 51, the number of sparsity estimated by SAMP algorithm is not changed, and the sparsity estimated by DSAMP algorithm also has errors. It can be seen from Figure 6 that the reconstruction probability of these two algorithms is rapidly decreasing under this sparsity.

As shown in Figure 8, the average running time of the StOMP algorithm and the SWOMP algorithm is relatively short, the average running time of the DSAMP algorithm is the longest, and the average running time of the SAMP algorithm is between several algorithms. This is because the StOMP algorithm and the SWOMP algorithm can select multiple atoms in each iteration and can quickly complete the iteration. The SAMP algorithm can select multiple atoms in each iteration, but the iterative atoms are repeatedly selected in the iteration, so the average running time is longer than the StOMP algorithm and the SWOMP algorithm. The DSAMP algorithm only adds one atom per iteration, so the average running time is the longest.

5.3. Radar Signal Simulation. In this section of the experiment, radar signals and MWC compressed sampling wideband digital receiver are used to simulate and verify the algorithm proposed in this paper. Set the Nyquist sampling frequency $f_{NYQ} = 20$ GHz of the signal, and the frequency of the signal is $f_1 = 1.1$ GHz, $f_2 = 3.37$ GHz, and $f_3 = 5.13$ GHz, the signal 1 is a linear frequency modulation signal, the frequency modulation width is 50 MHz, signals 2 and 3 are regular signals, the pulse width of the signal is $t = 1.02$
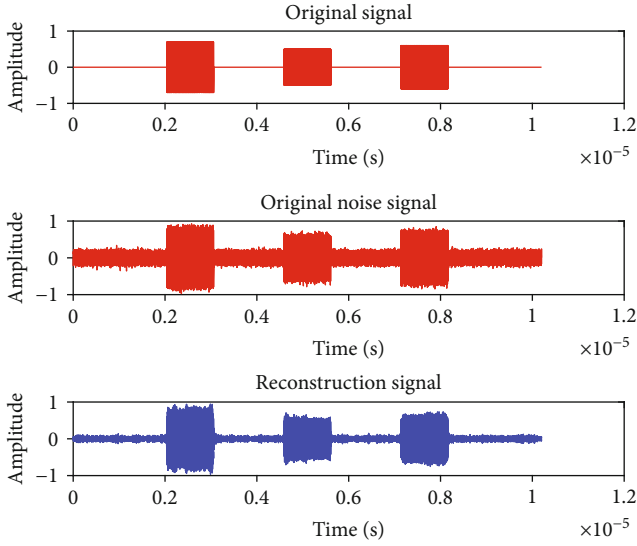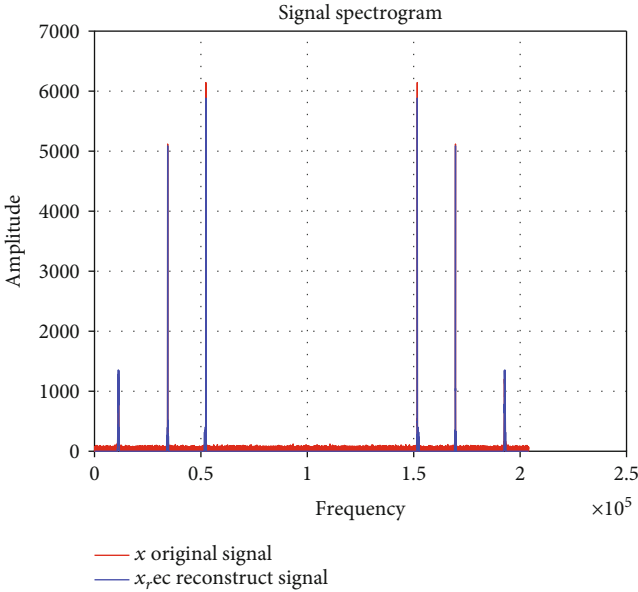
FIGURE 9: Radar signal reconstruction diagram.

FIGURE 10: Radar signal spectrum diagram.

us, the noise is Gaussian white noise, the signal-to-noise ratio SNR = 10 dB, and the signal amplitude is $A_1 = 0.7$, $A_2 = 0.5$, and $A_3 = 0.6$. The parameter settings of the MWC compressed sampling broadband digital receiver: the number of sampling channels is $M = 128$, the periodic pseudorandom sequence adopts a Bernoulli pseudorandom sequence of ±1, the period length is $M_P = 255$, the signal sampling time is $T = 10.2$ us, and the original sampling point number is $N = 204000$ points. Figure 9 is the radar signal reconstruction diagram of the compressed sampling broadband receiver of MWC by the DSAMP algorithm, including the original signal, the noise signal, and the reconstructed signal. Figure 10 is the corresponding spectrogram.
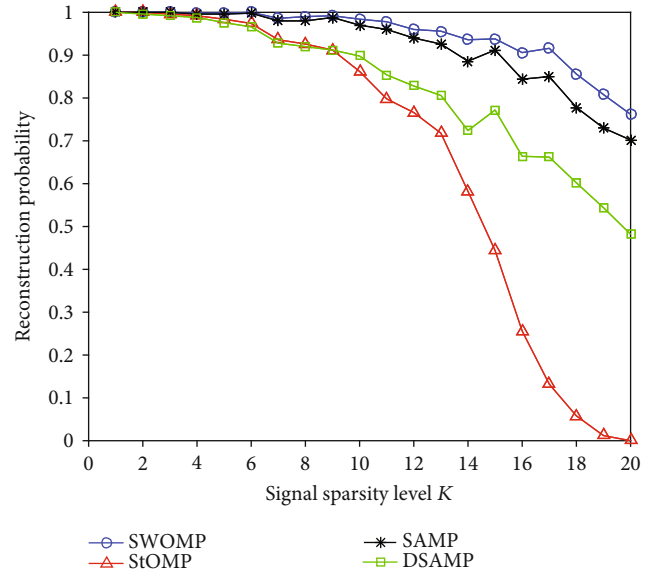

FIGURE 11: Reconstruction probability diagram under different sparsity.

As shown in Figure 9, the algorithm can estimate 3 signals and reduce the noise of the output signal in the reconstruction process. As shown in Figure 10, the algorithm can accurately estimate the position of the 3 signals. Since the real signal is used in the simulation, the same spectrum will appear in the position of the symmetrical spectrum. It can be seen that the spectrum on the far left and right is the LFM signal. The noise of the four reconstructed signals in the middle is higher than that of the noise-added signal, and all noises beyond the sparsity are eliminated. Therefore, the algorithm has a certain effect on the improvement of signal-to-noise ratio.

In order to prevent multiple signals from appearing in a channel at the same time, the frequencies of the signals are equally spaced. The initial frequency of the signal is, respectively, $f_0 = 1.1555$ GHz, the interval between each signal is 0.315 GHz, the interval of the number of signals is [1, 20], and the step is one. The signal amplitude of is all set to $A_1 = 0.7$. Perform 1000 Monte Carlo experiments under each signal-to-noise ratio. Other conditions are the same as in Figure 9. Figure 11 is a reconstruction probability diagram of different numbers of signals. Figure 12 is an RMSE diagram of sparsity. The simulations use root mean square error (RMSE) to analyze the sparsity estimation performance, expressed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}(K_i - K)^2}, \quad (16)$$

where N is the times of independent Monte Carlo simulations, K_i denotes the sparsity estimation, and K denotes real sparsity.

Figures 11 and 12 are the results of the same experiment. The basis for judging the successful reconstruction of the
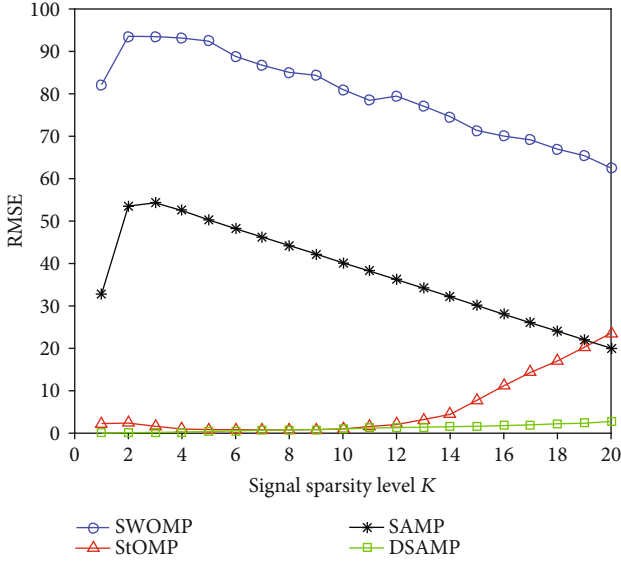
Figure 12: Sparsity estimation RMSE diagram.

Figure 13: Simulation diagram of radar signal SNR.

algorithm in Figure 11 is the intersection of the estimated sparsity position and the real position. If the intersection is the same as the real position, the reconstruction is considered successful. It can be seen from Figure 11 that the reconstruction probability of SWOMP algorithm and SAMP algorithm is better than DSAMP. The reconstruction probability of StOMP algorithm drops sharply after the number of signals is greater than 10. It can be seen from Figure 12 that the RMSE of the SWOMP algorithm and the SAMP algorithm is very large. It can be seen from Figure 12 that the RMSE of the SWOMP algorithm and the SAMP algorithm is very large. After the number of signals is greater than 3, it decreases. The sparsity estimated by the SWOMP algorithm is close to the observed value M. When the number of signals in the SAMP algorithm is greater than 2, the estimated sparsity is all 60. The RMSE of the StOMP algorithm rises sharply after the number of signals is greater than 10. The RMSE of the DSAMP algorithm has been kept within 5. The SWOMP algorithm and SAMP algorithm can still maintain a reconstruction probability above 0.7 when the number of signals is 20. This is because the two algorithms continue to add atoms to the atomic set until the stop iteration condition is met. Although it can be reconstructed, it introduces a large number of error atoms. When the number of signals in the StOMP algorithm is greater than 10, the reconstruction probability drops rapidly and the RMSE rises rapidly, indicating that the algorithm begins to fail at this time. The DSAMP algorithm maintains a small RMSE when the reconstruction probability decreases. This is because as the signal sparsity increases, the accuracy of the algorithm sparsity estimation decreases, but the estimated sparsity is still close to the true sparsity.

Set the number of signals to one signal, the frequency of the signal is selected randomly between $[0, 10\,\mathrm{GHz}]$. The signal-to-noise ratio SNR interval is $[-15, 20]$, and the step size is 5. Perform 1000 Monte Carlo experiments under each signal-to-noise ratio. Other conditions are the same
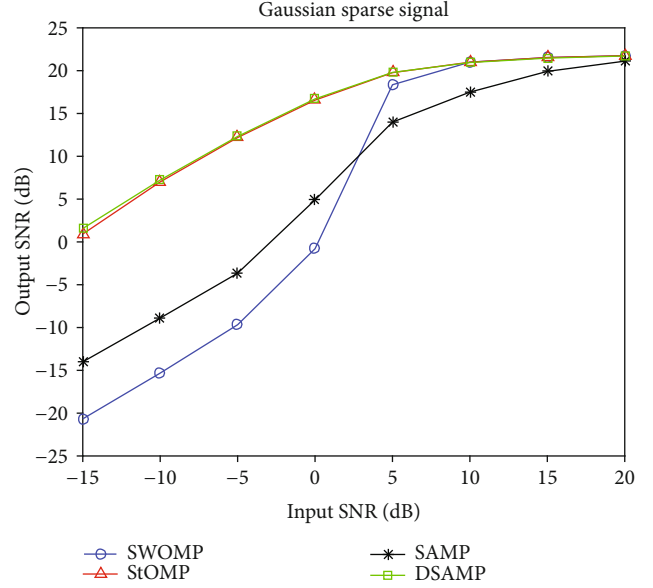
as in Figure 9. Figure 13 is a simulation diagram of the algorithm SNR.

As shown in Figure 13, the SWOMP algorithm and the SAMP algorithm are not effective in improving the output signal-to-noise ratio under low signal-to-noise ratio conditions, while the DSAMP algorithm and StOMP algorithm have a greater improvement in the output signal-to-noise ratio under low signal-to-noise ratio conditions. It can be seen from Figures 11 and 12 that the SWOMP algorithm and the SAMP algorithm are in the case of low signal-to-noise ratio. When the observation matrix is mixed, the noise of all frequency bands is mixed into the baseband, which causes the noise of the baseband signal to be strengthened, and the number of signals recovered during reconstruction is greater than the sparsity. Therefore, the output SNR is lower than the input SNR under the condition of low SNR. For DSAMP algorithm and StOMP algorithm, the sparsity estimation is more accurate at low SNR. It is equivalent to filtering out all the out-of-band noise of the signal, so the signal-to-noise ratio is greatly improved. For a high signal-to-noise ratio, the accuracy of the sparsity estimation of the four algorithms increases, so the output signal-to-noise ratio is improved. The DSAMP algorithm and the StOMP algorithm have a relatively close output signal-to-noise ratio. This is because the two algorithms have the same threshold setting when selecting the threshold. The DSAMP algorithm can have the advantage of improving the signal-to-noise ratio.

6. Conclusion

Adaptive sparsity estimation is necessary for wideband receivers based on MWC compressed sampling in radar systems. This paper proposes a sparsity adaptive matching pursuit algorithm based on dichotomy. In the algorithm proposed in this paper, there is no need to know the sparsity of the signal in advance. We estimate the sparsity of the signal

in a limited number of iterations through the dichotomy. It can be seen through simulation that an accurate and adaptive estimation of the sparsity of the signal is achieved. The orthogonal matching pursuit based on backtracking improves the accuracy of supporting concentrated atoms and finally improves the reconstruction probability of the algorithm. Due to the algorithm's accurate estimation of a single signal, the noise introduced by error atoms is reduced during reconstruction. Improve the signal-to-noise ratio of the algorithm to the signal. As the sparsity increases, the algorithm can still maintain a better reconstruction probability and more accurate sparsity estimation. In the future, the running time of the algorithm can be improved to reduce the calculation time of the algorithm.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] C.-T. Liu, R.-J. Wu, Z.-X. He, X.-F. Zhao, H.-C. Li, and P.-Z. Wang, "Modeling and analyzing interference signal in a complex electromagnetic environment," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, Article ID 1, 2016.

[2] X. Su, Q. X. Fu, Y. Z. Li, and N. Liu, "The hotspots and development trend of wideband digital channelized receiver technology," *Aerospace Electronic Warfare*, vol. 30, no. 1, pp. 32–35, 2014.

[3] T. Huang, Y. Liu, X. Xu, Y. C. Eldar, and X. Wang, "Analysis of frequency agile radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 66, no. 23, pp. 6228–6240, 2018.

[4] L. Wan, L. Sun, K. Liu, X. Wang, Q. Lin, and T. Zhu, "Autonomous vehicle source enumeration exploiting non-cooperative UAV in software defined internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.

[5] J. Skapa and M. Pola, "Ultra wide band receiver design," in *2016 26th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 304–308, Kosice, Slovakia, 2016.

[6] T. Chen, L. Z. Liu, and L. M. Guo, "Wideband signal detection based on MWC discrete compressed sampling structure," *Transactions of Nanjing University of Aeronautics and Astronautics*, vol. 34, no. 2, pp. 105–114, 2017.

[7] D. L. Donoho, "Compressed sensing," *IEEE Trans Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[9] L. Wan, Y. Sun, L. Sun, Z. Ning, and J. J. P. C. Rodrigues, "Deep learning based autonomous vehicle super resolution DOA estimation for safety driving," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, 2020.

[10] E. J. Candes and T. Tao, "Decoding by linear programming," *EEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[11] Z. Shi, C. Zhou, Y. Gu, N. A. Goodman, and F. Qu, "Source estimation using coprime array: a sparse reconstruction perspective," *IEEE Sensors Journal*, vol. 17, no. 3, pp. 755–765, 2017.

[12] R. Li, X. Duan, X. Guo, W. He, and Y. Lv, "Adaptive compressive sensing of images using spatial entropy," *Computational Intelligence and Neuroscience*, vol. 2017, Article ID 9059204, 9 pages, 2017.

[13] Q. Cheng, A. Alomainy, and Y. Hao, "Compressive millimeter-wave phased array imaging," *IEEE Access*, vol. 4, pp. 9580–9588, 2017.

[14] Y. H. Wang, X. Li, K. Xu, F. B. Ren, and H. Yu, "Data-Driven Sampling Matrix Boolean Optimization for Energy-Efficient Biomedical Signal Acquisition by Compressive Sensing," *IEEE Transactions on Biomedical Circuits ï¼† Systems*, vol. 11, no. 2, pp. 255–266, 2017.

[15] Z. Qin, J. Fan, Y. Liu, Y. Gao, and G. Y. Li, "Sparse representation for wireless communications: a compressive sensing approach," *IEEE Signal Processing Magazine*, vol. 35, no. 3, pp. 40–58, 2018.

[16] L. Sun, L. T. Wan, and X. P. Wang, "Learning-based resource allocation strategy for industrial IoT in UAV-enabled MEC systems," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[17] C. L. An, Z. Y. Zhao, M. Diao, and H. Y. Gao, "A method for direction finding of uncorrelated and coherent signals coexisted based on arbitrary array," *Journal of Harbin Engineering University*, vol. 34, no. 4, pp. 517–523, 2013.

[18] C. Zhou, Y. Gu, S. He, and Z. Shi, "A robust and efficient algorithm for coprime array adaptive beamforming," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1099–1112, 2018.

[19] A. Wadhwa, U. Madhow, and N. R. Shanbhag, "Slicer architectures for analog-to-information conversion in channel equalizers," *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1234–1246, 2017.

[20] W. B. Xu, Y. P. Cui, Y. Wang, S. Y. Wang, and J. R. Lin, "A hardware implementation of random demodulation analog-to-information converter," *ICE Electronics Express*, vol. 13, no. 16, pp. 1–6, 2016.

[21] R. Grigoryan, T. L. Jensen, and T. Larsen, "Computational complexity reduction in nonuniform compressed sensing by multicoset emulation," *Signal Processing*, vol. 131, pp. 492–501, 2017.

[22] F. Xi, S. Chen, Y. D. Zhang, and Z. Liu, "Gridless quadrature compressive sampling with interpolated array technique," *Signal Processing*, vol. 133, pp. 1–12, 2017.

[23] M. Mishali and Y. C. Elder, "From theory to practice: sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375–391, 2010.

[24] K. M. Cohen, C. Attias, B. Farbman, I. Teselniker, and Y. C. Eldar, "Channel estimation in UWB channels using compressed sensing," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1966–1970, Florence, Italy, 2014.

[25] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Sub-Nyquist acquisition hardware for wideband communication," in *2010 IEEE Workshop On Signal Processing Systems*, pp. 156–161, San Francisco, CA, USA, 2010.

[26] C. Zhou, Y. Gu, Y. D. Zhang, Z. Shi, T. Jin, and X. Wu, "Compressive sensing-based coprime array direction-of-arrival estimation," *IET Communications*, vol. 11, no. 11, pp. 1719–1724, 2017.

[27] L. Li, Y. Fang, L. Liu, H. Peng, J. Kurths, and Y. Yang, "Overview of compressed sensing: sensing model, reconstruction algorithm, and its applications," *Applied Sciences*, vol. 10, no. 17, p. 5909, 2020.

[28] S. G. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1994.

[29] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[30] D. Neenell and R. Vetshynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[31] D. Neenell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[32] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[33] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 6202–6216, 2012.

[34] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[35] T. Blumensath and M. E. Davies, "Stagewise weak gradient pursuits," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4333–4346, 2009.

[36] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pp. 581–587, Pacific Grove, CA, USA, 2008.

[37] L. Sun, L. Wan, K. Liu, and X. Wang, "Cooperative-evolution-based WPT resource allocation for large-scale cognitive industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5401–5411, 2020.

[38] X. L. D, H. Xing, B. B. Gu, B. Chen, and S. M. Qiu, "SAMP-RB compressed sensing reconstruction algorithm based on variable step," *Application Research of Computers*, vol. 35, no. 4, pp. 1084–1087, 2018.

WILEY | Hindawi

Research Article

High-Precision Mutual Coupling Coefficient Estimation for Adaptive Beamforming

Ziang Feng [ID],[1] **Guoping Hu** [ID],[2] **Qichao Ge** [ID],[1] **and Hao Zhou** [ID][2]

[1]*Graduate College, Air Force Engineering University, Xi'an 710051, China*
[2]*Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China*

Correspondence should be addressed to Qichao Ge; geqichao927@163.com

Here, a high-precision mutual coupling coefficient estimation method is proposed that is more suitable for adaptive beamforming than traditional algorithms. According to the relationship between the designed transition matrix and the signal, the proposed algorithm selects the transition matrix corresponding to the high-power signal. The high-precision estimation of the mutual coupling coefficient is obtained by using the selected transition matrix estimation, which yields relatively good estimation accuracy for the mutual coupling coefficient when the desired signal-to-noise ratio (SNR) is low and relatively robust adaptive beamforming with unknown mutual coupling. Simulation results demonstrate the validity of the proposed method.

1. Introduction

Signal processing technology, such as direction of arrival (DOA) and robust adaptive beamforming (RAB), has been widely used in radar, sonar, communication, etc. [1–7]. As a kind of array error, mutual coupling seriously affects the performance of various signal processing algorithms [8–11]. To avoid the influence of mutual coupling, a middle subarray-based (MSB) approach is proposed in [12], and a maximum interelement spacing constraint (MISC) array is designed in [13]. In addition, many algorithms for calibrating mutual coupling have been proposed. Since the mutual coupling matrix (MCM) can be modeled as a banded symmetric Toeplitz matrix for a uniform linear array (ULA), a subspace-based method is proposed in [14], and a fourth-order cumulant- (FOC-) based method is proposed in [15]. In addition, an iterative autocalibration algorithm based on the eigendecomposition of the sampling covariance matrix for a uniform circular array (UCA) is proposed to calibrate unknown mutual coupling since the MCM has a complex symmetric circular Toeplitz structure in a UCA in [16]. Furthermore, a joint DOA estimation and mutual coupling self-calibration for ULA-based bistatic multiple-input-multiple-output (MIMO) radar is proposed in [17]. Based on [16], a parameter estimation method for direction-dependent mutual coupling is proposed in [18]. However, the MSB approach reduces the degree of freedom (DOF) of the array, and the complex structure of the MISC array increases the difficulty of signal processing, while the subspace-based method in [14] constructs a high-dimensional matrix, and the algorithms proposed in [16, 18] both need an iterative process. To reduce the computational complexity in estimating mutual coupling, two low-complexity algorithms for direction-dependent mutual coupling and direction-independent mutual coupling were proposed in [19, 20], respectively.

To improve the robustness of adaptive beamforming in the presence of unknown mutual coupling, a middle subarray-plus-reconstruction-based (MSRB) method combining the MSB algorithm with interference-plus-noise covariance matrix (INCM) reconstruction [21] is proposed in [22]. However, the MSRB approach requires a large array aperture for high performance. Similarly, the desired signal steering vector with unknown mutual coupling is calibrated by using the specific structure of the MCM, and the new beamformer is obtained by combining a diagonal loading beamformer with the desired signal steering vector estimation [11]. To reduce the computational load, a subspace-plus-reconstruction-based (SRB) beamformer is then designed by

incorporating the subspace-based mutual coupling coefficient estimation method [14], and INCM reconstruction is designed in [10]. Obviously, the algorithms proposed in [20] can be utilized to design beamformers to further reduce the computational load. However, the algorithm proposed in [20] cannot obtain a high accuracy for the mutual coupling coefficients when there is a high power difference between signals since the accuracy of the subspace-based method is positively correlated with the signal-to-noise ratio (SNR).

To further improve the robustness of adaptive beamforming to unknown mutual coupling, a novel subspace-based algorithm is proposed to estimate mutual coupling coefficients and is utilized to design a novel adaptive beamformer. Different from the algorithms proposed in [20], we add the process of selecting several suitable transitional matrices and calculating their inverses. After estimating a group of transitional matrices, we select the transitional matrix corresponding to the maximum spectral peak to estimate the mutual coupling coefficient vector and the MCM. Then, the signal steering vector is calibrated with the estimated MCM. Finally, by combining the estimated MCM and the INCM reconstruction, we propose a novel adaptive beamforming algorithm. According to simulation, compared with several existing approaches, the proposed mutual coupling estimation method has a higher estimation accuracy, especially in the case of SNR differences between several signals. The designed beamformer is more robust than the existing beamformer to mutual coupling.

2. Signal Mode

We assume that a ULA of N sensors is impinged by L narrowband uncorrelated signals, and the noise is additive white Gaussian noise. We assume that the signal and noise are statistically independent. When the direction-independent mutual coupling effect is considered, the received snapshot at the kth time instant can be expressed as

$$\mathbf{x}(k) = \mathbf{x}_s(k) + \mathbf{n}(k) = \mathbf{C} \sum_{l=1}^{L} s_l(k)\mathbf{a}(\theta_l) + \mathbf{n}(k), \qquad (1)$$

where $\mathbf{x}_s(k)$ and $\mathbf{n}(k)$ stand for the $N \times 1$ vector of the signal and noise, respectively. θ_l is the lth signal DOA, and $s_l(k)$ and $\mathbf{a}(\theta_l) \in \mathbb{C}^{N \times 1}$ are the corresponding complex envelope and steering vector, respectively. $\mathbf{C} \in \mathbb{C}^{N \times N}$ denotes the MCM, which can be molded as a banded symmetric Toeplitz matrix for a ULA since the mutual coupling coefficients between the elements are inversely proportional to their distance. In general, we assume that the mutual coupling coefficient becomes zero when the spacing of two elements exceeds an interelement spacing of P; hence, the mutual coupling coefficient vector and the MCM can be defined as

$$\begin{aligned} \mathbf{c} &= [c_0, c_1, c_2, \cdots, c_{P-1}]^T, \\ \mathbf{C} &= \text{Toeplitz}(\mathbf{c}_N, \mathbf{c}_N), \end{aligned} \qquad (2)$$

where $c_0 = 1$ and $\mathbf{c}_N = [\mathbf{c}^T, 0]^T \in \mathbb{C}^{N \times 1}$.

Recall that in Equation (1), the covariance matrix of $\mathbf{x}(k)$ can be given by

$$\mathbf{R} = E[\mathbf{x}(k)\mathbf{x}^H(k)] = \mathbf{C}\mathbf{A}\mathbf{R}_s\mathbf{A}^H\mathbf{C}^H + \sigma_n^2\mathbf{I}_N, \qquad (3)$$

where $E[\cdot]$ and $(\cdot)^H$ represent the expectation and conjugate transpose, respectively. $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_L] \in \mathbb{C}^{N \times L}$ denotes the manifold matrix, and $\mathbf{R}_s = \text{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_L^2) \in \mathbb{R}^{L \times L}$ is the covariance matrix of signals, where σ_l^2 is the power of the lth signal. σ_n^2 is the power of the noise, and $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ is an identity matrix.

In practice, the sampling covariance matrix is usually used in lieu of the covariance matrix, and the sampling covariance matrix $\hat{\mathbf{R}}_x$ can be expressed as

$$\hat{\mathbf{R}}_x = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}(k)\mathbf{x}^H(k), \qquad (4)$$

where K stands for the number of snapshots. After eigendecomposing $\hat{\mathbf{R}}_x$, we can obtain

$$\hat{\mathbf{R}}_x = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H + \mathbf{U}_n\mathbf{\Lambda}_n\mathbf{U}_n^H, \qquad (5)$$

where $\mathbf{U}$ is the eigenvector matrix and $\mathbf{\Lambda}$ denotes the corresponding eigenvalue matrix. $\mathbf{\Lambda}_s$ and $\mathbf{\Lambda}_n$ are diagonal matrices that contain L large eigenvalues and the remaining small eigenvalues, respectively. $\mathbf{U}_s$ and $\mathbf{U}_n$ are the corresponding eigenvector matrices. In general, $\mathbf{U}_s$ is called the signal subspace, and $\mathbf{U}_n$ is called the noise subspace. Additionally, $\{\mathbf{C}\mathbf{a}(\theta_l)\}_{l=1}^{L}$ can span the signal subspace.

For instance, when DOA estimation is performed using a subspace algorithm, the spatial spectral function shown below is usually used [13]:

$$P(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}(\theta)}. \qquad (6)$$

Obviously, if the MCM is unknown, the spectral peaks do not correspond to the true DOA of the signals since $\mathbf{U}_n$ is orthogonal to $\mathbf{C}\mathbf{a}(\theta)$, not $\mathbf{a}(\theta)$. That is, if $\mathbf{C}$ can be estimated, then the DOA of signals can be estimated.

3. Proposed Algorithm

3.1. Mutual Coupling Coefficient Estimation.
According to the banded symmetric Toeplitz structure of MCM $\mathbf{C}$, we can obtain [19]

$$\mathbf{C}\mathbf{a}(\theta) = \mathbf{T}(\theta)\mathbf{c}, \qquad (7)$$

where

$$\mathbf{T}(\theta) = [\mathbf{E}_1\mathbf{a}(\theta), \mathbf{E}_2\mathbf{a}(\theta), \cdots, \mathbf{E}_P\mathbf{a}(\theta)], \qquad (8)$$

where

$$[\mathbf{E}_p]_{ij} = \begin{cases} 1, & \text{if} [\mathbf{C}]_{ij} = c_p, \\ 0, & \text{otherwise} \end{cases} \quad p = 0, 1, \cdots, P-1. \quad (9)$$

Based on the orthogonality between $\mathbf{Ca}(\theta_l)$ and the noise subspace $\mathbf{U}_n$, we can obtain

$$\mathbf{a}^H(\theta_l)\mathbf{C}^H\mathbf{U}_n\mathbf{U}_n^H\mathbf{Ca}(\theta_l) = 0. \quad (10)$$

Recall that Equations (7) and (10) can be rewritten as

$$\mathbf{c}^H\mathbf{Q}(\theta_l)\mathbf{c} = 0, \quad (11)$$

where $\mathbf{Q}(\theta)$ is a transitional matrix and is defined as

$$\mathbf{Q}(\theta) = \mathbf{T}^H(\theta)\mathbf{U}_n\mathbf{U}_n^H\mathbf{T}(\theta), \quad (12)$$

which is a $P \times P$ matrix, and its dimensions are smaller than those of the transitional matrix in [11, 13].

When $P \leq N - L$, $\mathbf{Q}(\theta)$ is a nonsingular matrix for a general θ since the ranks of $\mathbf{T}(\theta)$ and $\mathbf{U}_n$ are P and $N - L$, respectively. However, if θ is one DOA of the incident signals, $\mathbf{Q}(\theta)$ is a singular matrix, and its determinant is zero [23]. When $\mathbf{Q}(\theta)$ is a singular matrix, we can find that Equation (11) holds and that $\mathbf{c}$ is an eigenvector of the matrix $\mathbf{Q}(\theta)$ corresponding to the eigenvalue zero, since $\mathbf{c}$ is a nonzero vector. That is to say, the degree of freedom of the proposed algorithm is $N - P$.

Namely, the mutual coupling coefficient vector $\mathbf{c}$ can be estimated by

$$\mathbf{c}_l = \frac{\mathbf{v}_{l\,\min}}{v_{l\,\min\,1}}, \quad (13)$$

$$\widehat{\mathbf{c}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{c}_l, \quad (14)$$

where $\mathbf{v}_{l\,\min}$ denotes the eigenvector corresponding to the minimum eigenvalue of $\mathbf{Q}(\theta_l)$, and $v_{l\,\min\,1}$ is the first entry of $\mathbf{v}_{l\,\min}$.

From Equation (12), we can find that the matrix $\mathbf{Q}(\theta_l)$ can be easily calculated by the known DOA of the incident signals, but the DOAs are unknown. Hence, we construct a new spectral function

$$P_{\det}(\theta) = \frac{1}{\det[\mathbf{Q}(\theta)]}, \quad (15)$$

where $\det[\cdot]$ stands for the determinant of a matrix. Finally, $\mathbf{Q}(\theta_l)$ can be obtained through spectral peak searching since the first L peaks correspond to $\{\mathbf{Q}(\theta_l)\}_{l=1}^{L}$.

Since the noise subspace is estimated by the sampling covariance matrix, the correlation between the signal steering vector and the signal subspace improves with increasing SNR. Namely, as the SNR increases, the orthogonality of the signal and noise subspace becomes more obvious. That

is, at a low SNR, the signal steering vector and the noise subspace may not be orthogonal, i.e.,

$$\mathbf{a}^H(\theta_r)\mathbf{C}^H\mathbf{U}_n\mathbf{U}_n^H\mathbf{Ca}(\theta_r) \neq 0, \quad (16)$$

where θ_r stands for the DOA corresponding to a signal with a low SNR.

Further, we can obtain

$$\mathbf{c}^H\mathbf{Q}(\theta_r)\mathbf{c} \neq 0. \quad (17)$$

Distinctly, $\mathbf{c}$ is not the eigenvector of the matrix $\mathbf{Q}(\theta_r)$. In other words, if there are two signals, one with a high SNR and the other with a low SNR, such as an interference signal and a desired signal simultaneously incident on an array, the estimation accuracy decreases when Equations (13) and (14) are used to estimate the mutual coupling coefficients. In this case, we can eigendecompose only $\mathbf{Q}_{\max}$ corresponding to the maximum peak of Equation (15); hence, the estimation of $\mathbf{c}$ can be given by

$$\widehat{\mathbf{c}} = \frac{\mathbf{v}_{\mathrm{m-min}}}{v_{\mathrm{m-min}\,1}}, \quad (18)$$

where $\mathbf{v}_{\mathrm{m-min}}$ is the eigenvector corresponding to the minimum eigenvalue of the matrix $\mathbf{Q}_{\max}$, and $v_{\mathrm{m-min}\,1}$ is the first entry of $\mathbf{v}_{\mathrm{m-min}}$.

Obviously, when multiple incident SNRs are similar, the use of Equation (18) to estimate the mutual coupling coefficients results in errors. However, in most cases, the SNRs of incident signals are not the same. For example, in RAB, due to the simultaneous existence of interference signals and the desired signal, the estimation accuracy of mutual coupling coefficients is higher when using Equation (18) than when using Equations (13) and (14).

3.2. Adaptive Beamforming with Unknown Mutual Coupling. Based on Equation (18), when there are one desired signal and L interference signals, the received data can be calibrated as

$$\widehat{\mathbf{x}}(k) = \widehat{\mathbf{C}}^{-1}\mathbf{x}(k) = \widehat{\mathbf{C}}^{-1}\mathbf{C}\left[s_0(k)\mathbf{a}(\theta_0) + \sum_{l=1}^{L}s_l(k)\mathbf{a}(\theta_l)\right]$$

$$+ \widehat{\mathbf{C}}^{-1}\mathbf{n}(k) \approx s_0(k)\mathbf{a}(\theta_0) + \sum_{l=1}^{L}s_l(k)\mathbf{a}(\theta_l) + \widehat{\mathbf{C}}^{-1}\mathbf{n}(k). \quad (19)$$

where $\widehat{\mathbf{C}} \triangleq \mathrm{Toeplitz}(\widehat{\mathbf{c}}_N, \widehat{\mathbf{c}}_N)$, $\widehat{\mathbf{c}}_N = [\widehat{\mathbf{c}}^T, 0]^T \in \mathbb{C}^{N\times 1}$, θ_0 is the direction of the desired signal, and $s_0(k)$ is the corresponding signal complex envelope.

Recall that in Equation (3), the covariance matrix can be rewritten as

$$\mathbf{R}_{\widehat{x}} = E[\widehat{\mathbf{x}}(k)\widehat{\mathbf{x}}^H(k)] = \sigma_0^2\mathbf{a}(\theta_0)\mathbf{a}^H(\theta_0)$$

$$+ \sum_{l=1}^{L}\sigma_l^2\mathbf{a}(\theta_l)\mathbf{a}^H(\theta_l) + \sigma_n^2\widehat{\mathbf{C}}^{-1}\widehat{\mathbf{C}}^{-H}. \quad (20)$$

Distinctly, after the received data are calibrated, the noise covariance matrix is no longer $\sigma_n^2\mathbf{I}_M$ but is $\sigma_n^2\widehat{\mathbf{C}}^{-1}\widehat{\mathbf{C}}^{-H}$; namely, the white Gaussian noise becomes nonwhite Gaussian noise, which seriously affects the performance of the beamformer. Therefore, we compensate for $\mathbf{R}_{\widehat{x}}$ to obtain

$$\tilde{\mathbf{R}}_{\widehat{x}} = \widehat{\mathbf{R}}_{\widehat{x}} - \widehat{\sigma}_n^2\widehat{\mathbf{C}}^{-1}\widehat{\mathbf{C}}^{-H} + \widehat{\sigma}_n^2\mathbf{I}_M, \qquad (21)$$

where $\widehat{\sigma}_n^2$ is the noise power estimate, which can be expressed as

$$\widehat{\sigma}_n^2 = \frac{1}{M-L-1}\sum_{m=L+2}^{M}\widehat{\lambda}_m, \qquad (22)$$

where $\widehat{\lambda}_m$ $(m=1,2,\cdots,M)$ in descending order are the eigenvalues of the sample covariance matrix $\widehat{\mathbf{R}}_{\widehat{x}}$. Similarly, $\widehat{\mathbf{R}}_{\widehat{x}}$ is given by

$$\widehat{\mathbf{R}}_{\widehat{x}} = \frac{1}{K}\sum_{k=1}^{K}\widehat{\mathbf{x}}(k)\widehat{\mathbf{x}}^H(k). \qquad (23)$$

Further, the INCM can be reconstructed by [24]

$$\tilde{\mathbf{R}}_{i+n} = \tilde{\mathbf{R}}_{\widehat{x}} - \widehat{\sigma}_0^2\mathbf{a}(\theta_0)\mathbf{a}^H(\theta_0) + \gamma\mathbf{I}_M, \qquad (24)$$

where γ is a diagonal loading factor set to further reduce the impact of nonwhite Gaussian noise and $\widehat{\sigma}_0^2$ can be calculated as

$$\widehat{\sigma}_0^2 = \frac{1}{\mathbf{a}^H(\theta_0)\tilde{\mathbf{R}}_{\widehat{x}}^{-1}\mathbf{a}(\theta_0)}. \qquad (25)$$

Under the minimum variance distortionless response principle for RAB, the weight vector is usually expressed as [25]

$$\mathbf{w}_{\mathrm{opt}} = \frac{\mathbf{R}_{i+n}^{-1}\mathbf{a}(\theta_0)}{\mathbf{a}^H(\theta_0)\mathbf{R}_{i+n}^{-1}\mathbf{a}(\theta_0)}, \qquad (26)$$

where $\mathbf{R}_{i+n}$ denotes the ideal INCM without mutual coupling. Using $\tilde{\mathbf{R}}_{i+n}$ in lieu of $\mathbf{R}_{i+n}$ yields

$$\widehat{\mathbf{w}}_{\mathrm{opt}} = \frac{\tilde{\mathbf{R}}_{i+n}^{-1}\mathbf{a}(\theta_0)}{\mathbf{a}^H(\theta_0)\tilde{\mathbf{R}}_{i+n}^{-1}\mathbf{a}(\theta_0)}. \qquad (27)$$

However, Equation (11) cannot be used to estimate the MCM when $M-L-1 < P$, since $\mathbf{Q}(\theta)$ is a singular matrix for any θ. Namely, the DOF of the proposed algorithm is limited for a fixed ULA.

The main computational load of the designed beamformer is caused by estimating mutual coupling coefficients. The process of estimating mutual coupling coefficients consists of two parts: one is the construction of spectral function, and the other is the search of spectral peaks. Hence, the computational load of the proposed method is approximately O

$(M^3) + S \times O(P^3)$, where $O(M^3)$ is due to the construction of spectral function, $S \times O(P^3)$ is caused by the search of spectral peaks, and S is the number of spectral peak searches.

In other words, the proposed beamformer can be summarized as follows:

Step 1. Estimate the covariance matrix $\mathbf{R}_{\widehat{x}}$ using Equation (4), and eigendecompose it using Equation (5).

Step 2. Construct the transformation matrix and the transitional matrix using Equations (7) and (12), respectively.

Step 3. Estimate the transitional matrix $\mathbf{Q}_{\max}$ corresponding to the highest SNR signal by searching for the highest peak of Equation (15), and estimate the mutual coupling coefficient vector using Equation (18).

Step 4. Reconstruct the INCM by calibrating the received data and compensating for calibration errors using Equation (24).

Step 5. Calculate the weight using Equation (27).

4. Simulations

In this section, we implement several simulations to validate the effectiveness and superiority of the proposed algorithm. A ULA with $N = 12$ elements spaced a half-wavelength apart is used. The additive noise in the elements is modeled as spatially and temporally independent complex Gaussian noise with zero mean and unit variance. We assume that the mutual coupling coefficient vector is

$$\mathbf{c} = \left[1, 0.90e^{-j\pi/3}, 0.75e^{j\pi/4}, 0.45e^{-j\pi/10}, 0.15e^{-j\pi/6}\right]^T. \qquad (28)$$

According to the previous analysis, an iterative algorithm and two subspace-based methods without iteration are selected for comparison, and the three algorithms are called Liao's method [14], Elbir's method [18], and Wen's method [20]. In each trial, the number of angular sectors is selected as 12, and $\varepsilon = 10^{-4}$ in Elbir's method. In addition, the Cramér-Rao bound (CRB) of the real and imaginary parts of the mutual coupling coefficients is provided in simulations. Note that the calculation of CRB is performed as in [26], and the CRB is modified to correspond to the real and imaginary parts of the mutual coupling coefficients.

Example 1. The directions of the desired signal and two interferences are assumed to be 10°, -20°, and 40°, respectively, and the corresponding SNR and interference-to-noise ratio (INR) are set to 10 dB and 40 dB, respectively. The number of snapshots is 500, and 100 Monte Carlo runs are performed. In this experiment, the results of the estimation of mutual coupling coefficients are listed in Table 1 (a single run). The proposed method successfully estimates all mutual coefficients with high accuracy. Figure 1 displays the root mean square error (RMSE) of the real and imaginary parts of all mutual coupling coefficients versus SNR and the number of snapshots, where the RMSE is calculated by

$$\mathrm{RMSE} = \sqrt{\frac{1}{M(P-1)}\sum_{m=1}^{M}\sum_{p=1}^{P-1}(\widehat{w}_m^p - w^p)^2}, \qquad (29)$$

TABLE 1: Mutual coupling coefficient estimation (SNR = 10 dB).

Mutual coupling coefficients	Real part (α)		Imaginary part (β)	
	$\widehat{\alpha}$	α	$\widehat{\beta}$	β
c_1	0.4488	0.4500	-0.7766	-0.7794
c_2	0.5288	0.5303	0.5284	0.5303
c_3	0.4264	0.4280	-0.1392	0.1391
c_4	0.1303	0.1299	-0.0742	-0.0750

where M is the number of Monte Carlo runs, w^p denotes the real or imaginary part of c_p, and $\widehat{w}_m^p$ stands for the estimated value of w^p in the mth trial.

Figure 1(a) displays the RMSE values of the estimation of the real and imaginary parts of all the mutual coupling coefficients versus the SNR. The RMSE of the real and imaginary parts of all the mutual coupling coefficients versus the number of snapshots is shown in Figure 1(b). As we can see, the mutual coupling coefficient estimation accuracy of the proposed algorithm is always close to the CRB; when the difference in the signal power is large, the performance of the proposed method is closer to the CRB, and the larger the difference is, the more obvious the advantage of the proposed algorithm. However, when the power of each signal is similar, the performance of the proposed method is worse than that of Liao's method and Wen's algorithm but is still better than that of Elbir's algorithm. These results occur because when all the signal powers are similar, the mutual coupling coefficients estimated by all the transitional matrices are close to each other, and the mutual coupling coefficients obtained using more information have higher accuracies. Additionally, when the SNR of the desired signal is significantly higher than the INR of the interference signal, the proposed method still yields better performance.

Example 2. The performance of the proposed beamformer and several classical robust adaptive beamformers, such as LSMI, MSB, MSRB, and SRB, is investigated in this example. In addition, the simulation compares a simplified SRB (SSRB) beamformer obtained utilizing Wen's method to replace the mutual coupling coefficient estimation algorithm in the SRB beamformer. The output signal-to-interference-plus-noise ratio (SINR) versus the SNR with the number of snapshots fixed at 500 and the output SINR versus the number of snapshots with the SNR fixed at 10 dB are analyzed. Figure 2(a) displays the output SINR of different beamformers versus SNR, while Figure 2(b) shows the output SINR of those approaches versus the number of snapshots.

In Figure 2, the output SINR of the proposed beamformer is always close to the optimal SINR. In addition, at a high SNR, its performance is close to that of the SRB beamformer and SSRB beamformer, which is obviously better than that of the other algorithms. However, at a low SNR, the performance of the designed beamformer is obviously higher than that of the SRB beamformer and the SSRB beamformer due

to the high-precision estimation of the mutual coupling matrix. The performance of the MSRB beamformer is limited by the array aperture and is always lower than that of the proposed beamformer. In addition, the designed beamformer has a faster convergence rate than the SRB beamformer and SSRB beamformer and is always close to the optimal output SINR.

Example 3. The performance of the designed beamformer with the unknown direction error is analyzed in this example. It is assumed that there is a random direction error of the desired signal, and it is uniformly distributed in $[-2°, 2°]$ in each trial. Namely, the direction error of the desired signal changes from run to run but remains fixed in one trial. The results are shown in Figure 3.

Compared with Figure 2, the performance of the proposed beamformer, the SRB beamformer, and the SSRB beamformer is basically unchanged in Figure 3 because these three beamformers can estimate the desired signal direction when estimating the mutual coupling coefficients. The performance of the MSRB beamformer is only slightly degraded, but that of the MSB beamformer is seriously degraded because INCM reconstruction is utilized in the MSRB beamformer but not in the latter. Note that the convergence rate of the proposed beamformer is also very fast compared to those of other beamformers.

Example 4. The performance of the designed beamformer with incoherent local scattering.

The influence of incoherent local scattering is considered in this experiment. Generally, incoherent local scattering always occurs and seriously affects the performance of the beamformer. Here, we assume that in the case of incoherent local scattering, the received desired signal can be expressed as

$$\mathbf{x}_s(k) = \mathbf{a}(\theta_0)s_0(k) + \sum_{q=1}^{3} s_q(k)\mathbf{a}(\theta_q), \qquad (30)$$

where $\theta_q(q=1,2,3)$ denotes the direction of the qth local scattering signal and is subject to uniform distribution in $[\theta_0 - 2°, \theta_0 + 2°]$ and $s_q(k)$ and $\mathbf{a}(\theta_q)$ are the corresponding signal waveform and the steering vector, respectively. The simulation results are shown in Figure 4.

Figure 4 displays the simulation result of each tested beamformer with incoherent local scattering. Compared with Figure 2, since local scattering signals disturb the sampling covariance matrix, the performance of all tested beamformers deteriorates to varying degrees, especially at high SNRs. Since the incoherent scattering signal causes serious errors in the estimation of the mutual coupling coefficient at a high SNR, the performance of all algorithms is seriously degraded at a high SNR. However, at a low SNR, the influence of the incoherent scattered signal is small, and the proposed algorithm only utilizes the highest power signal to estimate mutual coupling coefficients; thus, the performance of the proposed beamformer is always close to the optimal output SINR. In addition, the convergence rate of
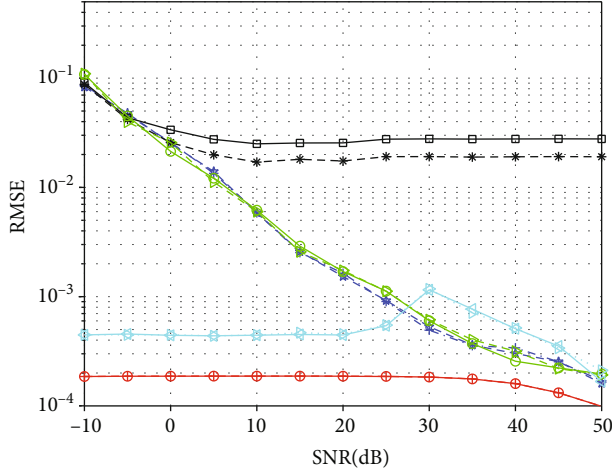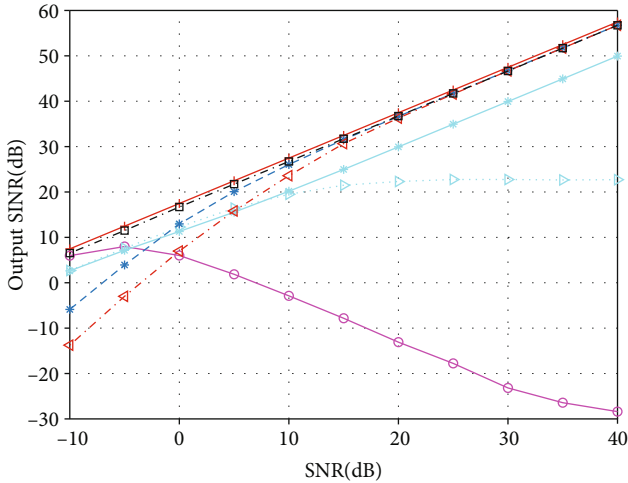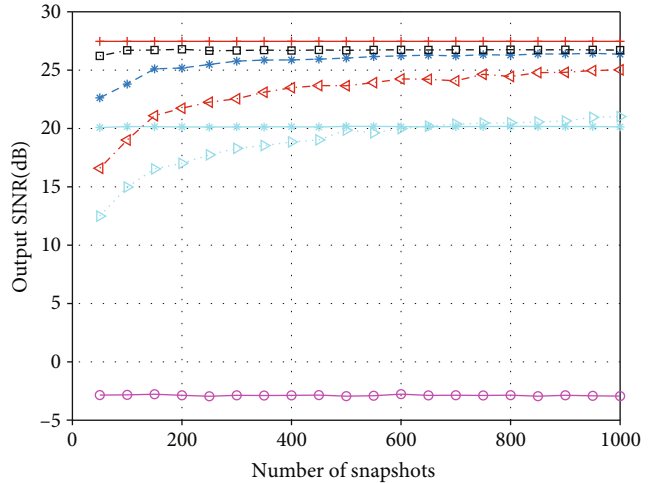
(a)
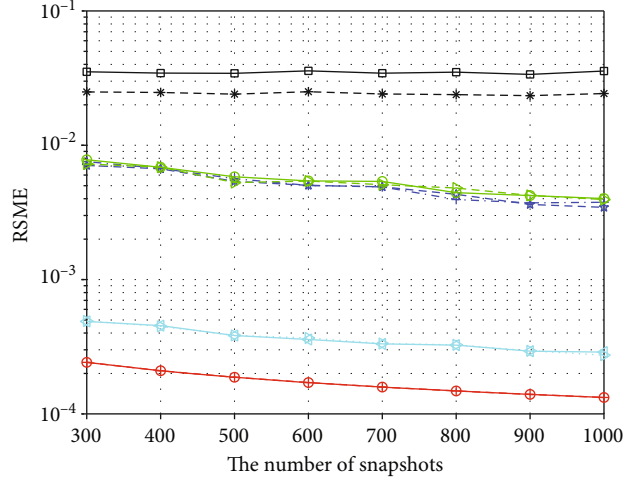
(b)

FIGURE 1: (a) RMSE of the mutual coupling coefficients versus SNR, $K = 500$, INR = 40 dB. (b) RMSE of the mutual coupling coefficients versus the number of snapshots, SNR = 10 dB, INR = 40 dB.

(a)

(b)

FIGURE 2: (a) Output SINR versus SNR, INR = 40 dB, $K = 500$. (b) Output SINR versus the number of snapshots, SNR = 10 dB, INR = 40 dB.

the proposed beamformer is very fast compared to those of other beamformers.

5. Conclusion

In this paper, a modified subspace-based mutual coupling coefficient estimation algorithm for beamforming is proposed. The main contribution of this manuscript includes two parts: (1) to improve the mutual coupling coefficient

estimation accuracy, we propose a strategy to improve the estimation accuracy by choosing appropriate transition matrices; (2) to improve the robustness of the beamformer to unknown mutual coupling, a beamformer is designed combining the calibrated steering vector and the interference-plus-noise covariance matrix reconstruction method. The proposed beamformer has superior performance than exiting algorithms especially when there is a big power gap between different interferences. Simulations demonstrate
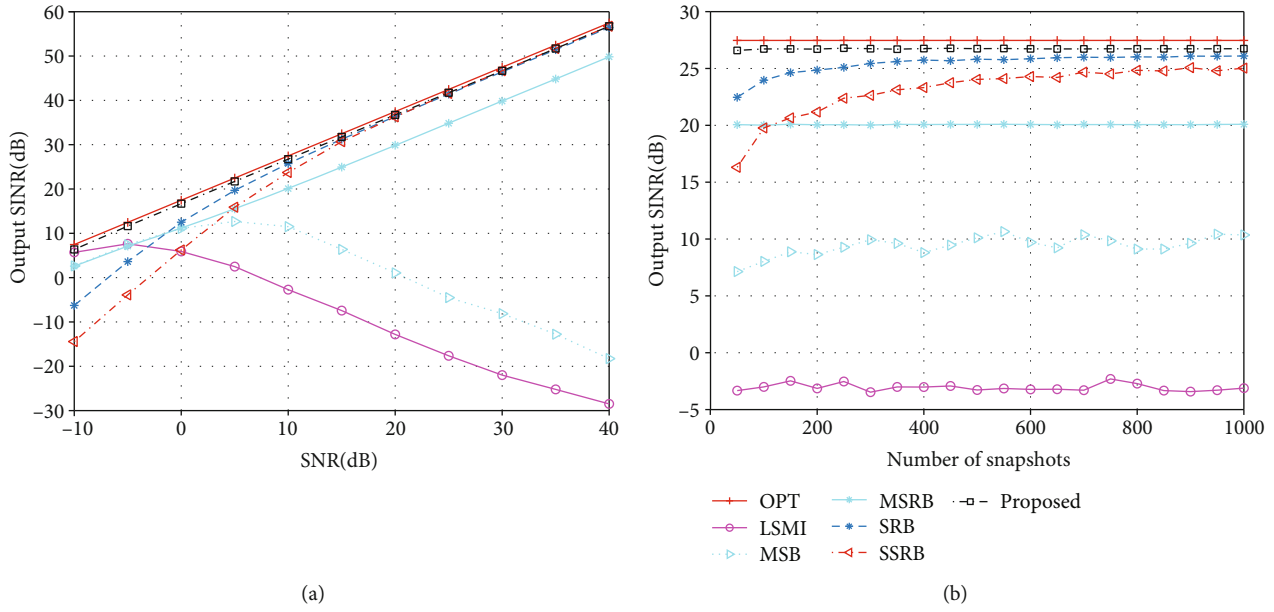
(a)

(b)

FIGURE 3: (a) Output SINR versus SNR with the unknown direction error, INR = 40 dB, $K = 500$. (b) Output SINR versus the number of snapshots with the unknown direction error, SNR = 10 dB, INR = 40 dB.
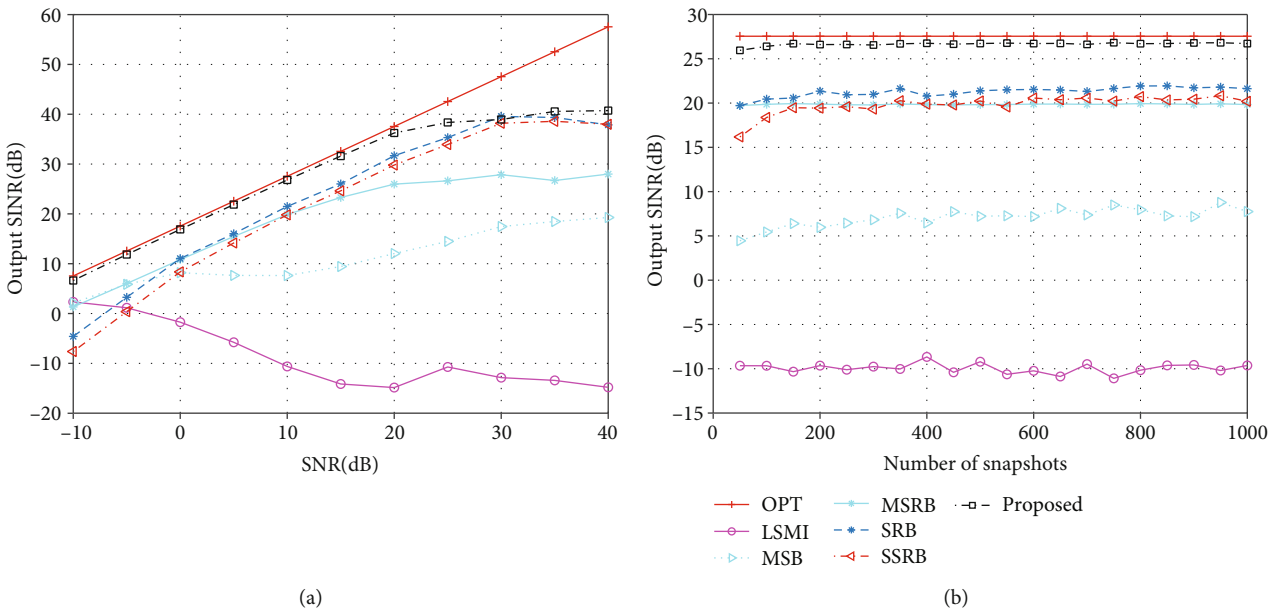

(a)

(b)

FIGURE 4: (a) Output SINR versus SNR with incoherent local scattering, INR = 40 dB, $K = 500$. (b) Output SINR versus the number of snapshots with incoherent local scattering, SNR = 10 dB, INR = 40 dB.

the superiority of the modified subspace-based mutual coupling coefficient estimation algorithm and the robustness of the designed beamformer to unknown mutual coupling.

Data Availability

The data sources from the references are all marked in this paper. And the parameters designed are all described in this paper. Therefore, the data in this paper has been fully covered and can be obtained.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] W. Zhang and S. A. Vorobyov, "Joint robust transmit/receive adaptive beamforming for MIMO radar using probability-constrained optimization," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 112–116, 2016.

[2] S. D. Somasundaram, "Wideband robust Capon beamforming for passive sonar," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 2, pp. 308–322, 2013.

[3] Z. Xiang and M. Tao, "Robust beamforming for wireless information and power transmission," *IEEE Wireless Communications Letters*, vol. 1, no. 4, pp. 372–375, 2012.

[4] L. Wan, X. Kong, and F. Xia, "Joint range-Doppler-angle estimation for intelligent tracking of moving aerial targets," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1625–1636, 2018.

[5] F. Wen and J. Shi, "Fast direction finding for bistatic EMVS-MIMO radar without pairing," *Signal Processing*, vol. 173, article 107512, p. 107512, 2020.

[6] J. Shi, F. Wen, and T. Liu, "Nested MIMO radar: coarrays, tensor modeling and angle estimation," *IEEE Transactions on Aerospace and Electronic Systems*, 2020.

[7] X. Wang, L. Wang, X. Li, and G. Bi, "Nuclear norm minimization framework for DOA estimation in MIMO radar," *Signal Process*, vol. 135, pp. 147–152, 2017.

[8] B. Friedlander and A. Weiss, "Direction finding in the presence of mutual coupling," *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 3, pp. 273–284, 1991.

[9] J. Dai, X. Bao, N. Hu, C. Chang, and W. Xu, "A recursive RARE algorithm for DOA estimation with unknown mutual coupling," *IEEE Antennas and Wireless Propagation Letters*, vol. 13, pp. 1593–1596, 2014.

[10] Z. Zheng, K. Liu, W.-Q. Wang, Y. Yang, and J. Yang, "Robust adaptive beamforming against mutual coupling based on mutual coupling coefficients estimation," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9124–9133, 2017.

[11] B. Liao and S.-C. Chan, "Adaptive beamforming for uniform linear arrays with unknown mutual coupling," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 464–467, 2012.

[12] Z. Ye and C. Liu, "Non-sensitive adaptive beamforming against mutual coupling," *IET Signal Processing*, vol. 3, no. 1, pp. 1–6, 2009.

[13] Z. Zheng, W.-Q. Wang, Y. Kong, and Y. D. Zhang, "MISC array: a new sparse array design achieving increased degrees of freedom and reduced mutual coupling effect," *IEEE Transactions on Signal Processing*, vol. 67, no. 7, pp. 1728–1741, 2019.

[14] B. Liao, Z.-G. Zhang, and S.-C. Chan, "DOA estimation and tracking of ULAs with mutual coupling," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 891–905, 2012.

[15] B. Liao and S.-C. Chan, "A cumulant-based method for direction finding in uniform linear arrays with mutual coupling," *IEEE antennas and wireless propagation letters*, vol. 13, pp. 1717–1720, 2014.

[16] M. Wang, X. Ma, S. Yan, and C. Hao, "An autocalibration algorithm for uniform circular array with unknown mutual coupling," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 12–15, 2015.

[17] F. Wen, Z. Zhang, K. Wang, G. Sheng, and G. Zhang, "Angle estimation and mutual coupling self-calibration for ULA-based bistatic MIMO radar," *Signal Processing*, vol. 144, pp. 61–67, 2018.

[18] A. M. Elbir, "Direction finding in the presence of direction-dependent mutual coupling," *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 1541–1544, 2017.

[19] Q. Ge, Y. Zhang, and Y. Wang, "A low complexity algorithm for direction of arrival estimation with direction-dependent mutual coupling," *IEEE Communications Letters*, vol. 24, no. 1, pp. 90–94, 2019.

[20] F. Wen, J. Wang, J. Shi, and G. Gui, "Auxiliary vehicle positioning based on robust DOA estimation with unknown mutual coupling," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5521–5532, 2020.

[21] Y. Gu and A. Leshem, "Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3881–3885, 2012.

[22] Z. Li, Y. Zhang, Q. Ge, and Y. Guo, "Middle subarray interference covariance matrix reconstruction approach for robust adaptive beamforming with mutual coupling," *IEEE Communications Letters*, vol. 23, no. 4, pp. 664–667, 2019.

[23] C. M. S. See and A. B. Gershman, "Direction-of-arrival estimation in partly calibrated subarray-based sensor arrays," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 329–338, 2004.

[24] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *Signal Processing*, vol. 93, no. 12, pp. 3264–3277, 2013.

[25] Z. Zheng, T. Yang, W.-Q. Wang, and H. C. So, "Robust adaptive beamforming via simplified interference power estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 3139–3152, 2019.

[26] M. Wang and A. Nehorai, "Coarrays, MUSIC, and the Cramér-Rao bound," *IEEE Transactions on Signal Processing*, vol. 65, no. 4, pp. 933–946, 2017.