Research Article

Augmented Lagrange Based on Modified Covariance Matching Criterion Method for DOA Estimation in Compressed Sensing

Weijian Si, Xinggen Qu, and Lutao Liu

Department of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

Correspondence should be addressed to Lutao Liu; liulutao@msn.com

Received 27 November 2013; Accepted 19 December 2013; Published 11 February 2014

Academic Editors: H. R. Karimi, Z. Yu, and W. Zhang

Copyright © 2014 Weijian Si et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel direction of arrival (DOA) estimation method in compressed sensing (CS) is presented, in which DOA estimation is considered as the joint sparse recovery from multiple measurement vectors (MMV). The proposed method is obtained by minimizing the modified-based covariance matching criterion, which is acquired by adding penalties according to the regularization method. This minimization problem is shown to be a semidefinite program (SDP) and transformed into a constrained quadratic programming problem for reducing computational complexity which can be solved by the augmented Lagrange method. The proposed method can significantly improve the performance especially in the scenarios with low signal to noise ratio (SNR), small number of snapshots, and closely spaced correlated sources. In addition, the Cramér-Rao bound (CRB) of the proposed method is developed and the performance guarantee is given according to a version of the restricted isometry property (RIP). The effectiveness and satisfactory performance of the proposed method are illustrated by simulation results.

1. Introduction

Direction of arrival (DOA) estimation of multiple narrow-band sources has been an interesting research topic in array signal processing. Its applications span many fields including radar, communication systems, and medical imaging [1, 2]. Many effective algorithms are proposed for DOA estimation, which are mainly classified into three categories. The beamforming algorithms such as MVDR [3] and MEM [4] obtain a nonparametric spatial spectrum by optimizing the filter weights. The subspace algorithms such as MUSIC [5] and ESPRIT [6] and their derivatives [7, 8] exploit the orthogonality of signal subspace and noise subspace for DOA estimation. The subspace fitting algorithms including Maximum Likelihood (ML) [9] and Weighted Subspace Fitting (WSF) [10] solve a multidimensional nonlinear optimization problem to obtain a precise estimation, but a good initialization is required to ensure global convergence and computational complexity is very high. All these algorithms focus on two important issues: resolution (i.e., the ability to correctly resolve two closely spaced sources) and precision (i.e., the deviation from true DOA) which are considered to be the

theoretical bases of evaluating certain algorithms. Many high resolution methods suffer from serious performance degradation in the scenarios with low SNR, small number of snapshots, and closely spaced correlated sources. More recently, many research applications involving compressed sensing (CS) [11], especially DOA estimation [12, 13], have become more and more popular in the signal processing. Moreover, the restricted isometry property (RIP) based on the perfect theoretical framework given with modern probability theory by Donoho [11] and Candès et al. [14] provides the performance guarantee in CS. Hence, in this paper, DOA estimation is posed as a joint sparse recovery problem where we recover jointly sparse sources from multiple measurement vectors (MMV) under CS framework.

CS is an emerging area which can break through the limit of Nyquist sampling theorem. On one hand, CS can simultaneously capture and store compressed or sparse source at a rate much lower than that of Nyquist sampling. On the other hand, it can recover the original source using nonadaptive linear projection measurements onto a suitable measurement matrix which satisfies the RIP. In CS, the joint sparse recovery aims to find a common support shared by unknown sparse

vectors from MMV, which is obtained by the measurement matrix. Support denotes the indices of the nonzero elements in the unknown sparse vectors. A sparse solution can be obtained as long as the support is determined.

CS theory has been widely applied to DOA estimation according to the source sparsity, which results from the fact that there are much fewer source directions than all potential directions in the spatial domain. The DOA estimation methods in CS are attractive since they have much better estimation performance than conventional estimation methods. In [15], Gürbüz et al. firstly formulate the DOA estimation problem under CS framework in the time domain. Wang et al. [16] propose a new architecture to estimate DOA by exploiting compressive sampling in the spatial domain. Stoica et al. [17] make full use of covariance matching criterion and present a semiparametric/sparse estimation method and its derivative called LIKES [18] for the separable model. Malioutov et al. [19] present the l_1 -SVD algorithm for DOA estimation which combines the SVD of the data matrix with a sparse recovery method based on l_1 -norm minimization. A new class of subspace-base algorithms such as compressive MUSIC (CS-MUSIC) [20] and subspace-augmented MUSIC (SA-MUSIC) [21] is proposed in recent years.

The RIP and various modified versions of it have been used as a foundation of performance guarantee [21–24] for the joint sparse recovery. The performance guarantee of MUSIC based on joint sparse recovery is given for identifying the unique support in a favorable case [25]. However, in the unfavorable case where the number of measurement vectors is smaller than the sparsity or the covariance matrix tends to lose rank due to correlated sources, performance guarantee fails. Lee et al. [21] propose a new performance guarantee in terms of a version of the RIP under such unfavorable conditions. The performance guarantees of other methods such as greedy algorithms and convex relaxation have been developed to find the sparse solution in [23, 24]. However, the guarantees of these methods cannot be simply extended to the MMV case to obtain a better bound for the sparse solution.

In this paper, we propose a novel augmented Lagrange based on modified covariance matching criterion method called AULMC for DOA estimation in CS. This method utilizes the minimization of the modified covariance matching criterion which is acquired by using regularization method to add penalties in order to obtain the stable sparse parameter estimation, especially in the low sparsity case. This minimization problem is shown to be a semidefinite program (SDP) and transformed into the constrained quadratic programming problem for the sake of reducing computational complexity. The augmented Lagrange function is formed to solve this problem by the use of augmented Lagrange method. AULMC has a number of advantages over the other methods. For example, it provides more precise estimation and higher resolution in the scenarios with low SNR, small number of snapshots, and closely spaced correlated sources, and it does not need priori knowledge about the number of sources or to choose the regularization parameter of the l_1 -optimization framework which is very difficult to select in the DOA estimation. In addition, we give a detailed derivation process of the closed-form expression for the

Cramér-Rao bound (CRB) of the new method and discuss an explicit condition that guarantees performance of the new method. This performance guarantee is given in terms of weaker version of the RIP which is referred to as weak-1 RIP. Simulation results illustrate the performance of the proposed method.

It is worth noting that covariance matching criterion has been used for DOA estimation [26]. In [26], a sparse iterative covariance-base estimation method, abbreviated as SPICE, is proposed. Our approach is different from SPICE because it utilizes modified covariance matching criterion which can guarantee the stability of solution even if the source sparsity is rather low. In the future work, we will focus on the application of our approach to data-driven design [27–30]. Now we briefly summarize the contributions of this work as follows.

- (i) A modified covariance matching criterion is proposed by adding penalties according to the regularization method.
- (ii) The original SDP problem is transformed into the constrained quadratic programming problem. The motivation to transform the problem is that it can reduce computational complexity.
- (iii) Augmented Lagrange based on modified covariance matching criterion method is devised to solve the resulting programming problem.
- (iv) The CRB and performance guarantee of the new method are given in detail.

The rest of this paper is organized as follows. In Section 2, we formulate the DOA estimation problem. Section 3 presents a modified covariance matching criterion. A novel augmented Lagrange based on modified covariance matching criterion method for DOA estimation is described in detail in Section 4; the performance of which is analyzed in Section 5. The performance of the proposed method is evaluated by simulations in Section 6. Conclusions are provided in Section 7.

2. Problem Formulation

Consider L narrow-band sources $\hat{s}_1, \hat{s}_2, \dots, \hat{s}_L$ impinging on a sensor array that consisted of P omnidirectional sensors from far field. At time instant t , the array received source can be given by

$$\mathbf{x}(t) = \sum_{k=1}^L \mathbf{a}(\theta_k) \hat{s}_k(t) + \mathbf{n}(t), \quad (1)$$

where $\mathbf{n}(t) \in \mathbb{C}^{P \times 1}$ denotes a noise term, $\theta_k \in \Omega$ is the unknown direction, of the k th source where Ω denotes the entire spatial range and $\mathbf{a}(\theta_k)$ is the $P \times 1$ steering vector. Although the DOA estimation of the single snapshot, which is a typical single measurement vector (SMV) model, has its value, the number of snapshots is larger than one in the most practical applications. Correspondingly, the multiple snapshots model is a MMV model.

Let $\{\tilde{\theta}_k\}_{k=1}^K$ denote a fine grid which covers Ω where there exist K ($K \gg L$) potential directions of the sources $s_1, s_2, \dots, s_K$ so that the true directions $\{\theta_k\}_{k=1}^L$ are aligned or are close to the grids. This means that if $\tilde{\theta}_{k_1}, \tilde{\theta}_{k_2}, \dots, \tilde{\theta}_{k_L}$ are equal to $\theta_1, \theta_2, \dots, \theta_L$, respectively, we have

$$s_k = \begin{cases} \hat{s}_l, & k = k_l \ (l = 1, 2, \dots, L), \\ 0, & \text{elsewhere.} \end{cases} \quad (2)$$

Hence, the multiple snapshots model can be written as the following sparse form:

$$\mathbf{x}(t) = \sum_{k=1}^K \mathbf{a}(\tilde{\theta}_k) s_k(t) + \mathbf{n}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad (3)$$

$$t = t_1, t_2, \dots, t_M,$$

where $\mathbf{A} = [\mathbf{a}(\tilde{\theta}_1) \ \mathbf{a}(\tilde{\theta}_2) \ \dots \ \mathbf{a}(\tilde{\theta}_K)]$ is the $P \times K$ manifold matrix corresponding to all the potential directions which is also referred to as an overcomplete dictionary in CS. $\mathbf{s}(t) = [s_1(t) \ s_2(t) \ \dots \ s_K(t)]^T$ is a L -sparse vector since it has at most L nonzero elements in K elements, and L is defined as sparsity where the operator $[\cdot]^T$ denotes transport. $\{\mathbf{s}(t_i)\}_{i=1}^M$ are jointly L -sparse if they share a common support. Hence, the matrix $\mathbf{S} = [\mathbf{s}(t_1) \ \mathbf{s}(t_2) \ \dots \ \mathbf{s}(t_M)] \in \mathbb{C}^{K \times M}$ has no more than L nonzero rows in order to be called row L -sparse. The MMV problem is that of identifying the row support of the unknown matrix $\mathbf{S}$ from the matrix $\mathbf{Y} \in \mathbb{C}^{N \times M}$ that consisted of MMV which is given by

$$\mathbf{Y} = [\mathbf{y}(t_1) \ \mathbf{y}(t_2) \ \dots \ \mathbf{y}(t_M)] = \Phi \mathbf{A} \mathbf{S} + \Phi \mathbf{N} \quad (4)$$

with a common measurement matrix Φ of the size $N \times P$ with $N < P$ where N is the number of nonadaptive linear projection measurements, such as random Gaussian matrix or random partial Fourier matrix, and noise matrix $\mathbf{N}$.

3. Modified Covariance Matching Criterion

In this section, the modified covariance matching criterion is developed in the CS scenario. A conventional covariance matrix of the compressed measurement source with the size $N \times N$ is given by

$$\mathbf{R}_y = E[\mathbf{y}(t) \mathbf{y}^*(t)] = \Phi \mathbf{A} \mathbf{R}_s \mathbf{A}^* \Phi^* + \Phi \mathbf{R}_n \Phi^*, \quad (5)$$

where $\mathbf{R}_s = E[\mathbf{s}(t) \mathbf{s}^*(t)]$ is a $K \times K$ covariance matrix of the sparse source whose off-diagonal elements denote the source correlation and diagonal elements denote the source powers. Since the powers are one-to-one corresponding to all the potential directions and our focus is on the source angle parameter estimation, $\mathbf{R}_s$ can be reduced to a diagonal matrix $\mathbf{R}_s(\tilde{\theta}) = \text{diag}(\tilde{\theta}_1 \ \tilde{\theta}_2 \ \dots \ \tilde{\theta}_K)$. According to (5), the measurement matrix can change the covariance matrix of the noise to render the noise colored even if the noise is white. Therefore, this adverse factor must be considered before recovering jointly sparse sources. In addition, the operators $(\cdot)^*$ and $E[\cdot]$ denote conjugate transpose and expectation, respectively.

Since $\Phi \mathbf{R}_n \Phi^*$ is a positive definite Hermitian matrix, a prewhitened method is given by the Cholesky decomposition. Let $\mathbf{B}$ be the Cholesky factor that satisfies

$$(\Phi \mathbf{R}_n \Phi^*)^{-1} = \mathbf{B}^* \mathbf{B}, \quad (6)$$

where $\mathbf{B} \in \mathbb{C}^{N \times N}$ is an upper triangular matrix of positive line. Hence, a prewhitened process is implemented by multiplying $\mathbf{y}(t)$ by $\mathbf{B}$ in order to obtain a pure source whose covariance matrix is written as

$$\tilde{\mathbf{R}}_y = \mathbf{B} \Phi \mathbf{A} \mathbf{R}_s \mathbf{A}^* \Phi^* \mathbf{B}^* + \mathbf{I}_N = \mathbf{C} \mathbf{R}_s \mathbf{C}^* + \mathbf{I}_N, \quad (7)$$

where $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K]$ and $\mathbf{I}_N$ is an identity matrix of the size $N \times N$. It is important to note that the unknown covariance matrix of the noise is transformed into the known identity matrix after the prewhitened process. Therefore, the prewhitened method improves the robustness to the noise. Then, the covariance matrix of the compressed measurement source denoising is realized by

$$\mathbf{R} = \tilde{\mathbf{R}}_y - \mathbf{I}_N = \mathbf{C} \mathbf{R}_s \mathbf{C}^*. \quad (8)$$

The parameter $\tilde{\theta}$ can be estimated by a class of the covariance matching estimation techniques (COMET) based on covariance matching criterion [31]. This parameter estimation method is well known to be a large-sample approximation of ML method and provide a more attractive solution than ML estimator.

The principle of COMET is that of using the right data to minimize its data model by the weighted least-squares (WLS) method. However, the lower the source sparsity is, the more likely it is to be ill-posed for the covariance matrix estimation error meaning that the optimal solution obtained directly by minimizing the conventional covariance matching criterion is instable. Hence, we employ regularization method to add penalties in order to sufficiently exploit prior knowledge to reduce the solution space for determining the stable optimal solution. The modified covariance matching criterion is proposed as the following form:

$$\begin{aligned} & [\text{vec}(\hat{\mathbf{R}}) - \text{vec}(\mathbf{R}(\tilde{\theta}))]^* \mathbf{\Gamma}^* \mathbf{P}^{-1} \mathbf{\Gamma} [\text{vec}(\hat{\mathbf{R}}) - \text{vec}(\mathbf{R}(\tilde{\theta}))] \\ & + \alpha \text{Tr}[\mathbf{R}^{-1}(\tilde{\theta}) - \hat{\mathbf{R}}^{-1}] + \beta \text{Tr} \left[\frac{\mathbf{R}^{-1}(\tilde{\theta}) \hat{\mathbf{R}}^{-1}}{4} \right], \end{aligned} \quad (9)$$

where parameter $\alpha \geq 0$ controls the solution smoothness (guarantee precision), parameter $\beta \geq 0$ controls the solution scale (guarantee sparsity), the inverses of the matrix $\mathbf{\Gamma}$, the sample covariance matrix $\hat{\mathbf{R}}$, and $\mathbf{R}(\tilde{\theta})$ are assumed to exist, and the matrix $\mathbf{P}^{-1}$ is the inverse of the covariance matrix of the residuals, $\tilde{\mathbf{e}} = \mathbf{\Gamma} \text{vec}(\hat{\mathbf{R}} - \mathbf{R}(\tilde{\theta}))$. Since $\hat{\mathbf{R}}$ is equal to $\mathbf{R}(\tilde{\theta})$ as $M \rightarrow \infty$, it follows from [32] that $\text{vec}(\hat{\mathbf{R}} - \mathbf{R}(\tilde{\theta}))$ satisfies the asymptotic normal distribution with mean zero and variance $N^{-1} \mathbf{R}^T(\tilde{\theta}) \otimes \hat{\mathbf{R}}$. In addition, the operators $\text{Tr}[\cdot]$, $\text{vec}(\cdot)$, and $\otimes$ denote matrix trace, column stacking operation,

and Kronecker product, respectively. Then, the matrix $\mathbf{P}$ can be given by

$$\begin{aligned}\mathbf{P} &= E[\tilde{\mathbf{e}}(\tilde{\boldsymbol{\theta}})\tilde{\mathbf{e}}^*(\tilde{\boldsymbol{\theta}})] \\ &= \Gamma E[\text{vec}(\widehat{\mathbf{R}} - \mathbf{R}(\tilde{\boldsymbol{\theta}}))\text{vec}^*(\widehat{\mathbf{R}} - \mathbf{R}(\tilde{\boldsymbol{\theta}}))]\Gamma^* \quad (10) \\ &= \Gamma N^{-1}\mathbf{R}^T(\tilde{\boldsymbol{\theta}}) \otimes \widehat{\mathbf{R}}\Gamma^* = \Gamma N^{-1}\mathbf{D}\Gamma^*.\end{aligned}$$

We consider the normalized data model in (9) and choose the regularization parameters $\alpha = \beta = N$ based on perceptual criterion [33]. By substituting (10) into (9), we have

$$\begin{aligned}\text{vec}^*(\widehat{\mathbf{R}} - \mathbf{R}(\tilde{\boldsymbol{\theta}}))\mathbf{D}^{-1}\text{vec}(\widehat{\mathbf{R}} - \mathbf{R}(\tilde{\boldsymbol{\theta}})) \\ + \text{Tr}[\mathbf{R}^{-1}(\tilde{\boldsymbol{\theta}}) - \widehat{\mathbf{R}}^{-1}] + \frac{\text{Tr}[\mathbf{R}^{-1}(\tilde{\boldsymbol{\theta}})\widehat{\mathbf{R}}^{-1}]}{4}.\end{aligned} \quad (11)$$

By the properties of vec , $\otimes$, and Tr , the data model can be further simplified to

$$\begin{aligned}\text{vec}^*(\widehat{\mathbf{R}} - \mathbf{R}(\tilde{\boldsymbol{\theta}}))\mathbf{D}^{-1}\text{vec}(\widehat{\mathbf{R}} - \mathbf{R}(\tilde{\boldsymbol{\theta}})) \\ + \text{Tr}[\mathbf{R}^{-1}(\tilde{\boldsymbol{\theta}})(\widehat{\mathbf{R}} - \mathbf{R}(\tilde{\boldsymbol{\theta}}))\widehat{\mathbf{R}}^{-1}] + \frac{\text{Tr}[\mathbf{R}^{-1}(\tilde{\boldsymbol{\theta}})\widehat{\mathbf{R}}^{-1}]}{4} \\ = \left\| \widehat{\mathbf{R}}^{-1/2} \left(\widehat{\mathbf{R}} - \mathbf{R}(\tilde{\boldsymbol{\theta}}) + \frac{\mathbf{I}_N}{2} \right) \mathbf{R}^{-1/2}(\tilde{\boldsymbol{\theta}}) \right\|^2,\end{aligned} \quad (12)$$

where $\|\cdot\|$ denotes the Frobenius norm for matrices or the Euclidean norm for vectors. The data model in (12) is considered to be the modified covariance matching criterion. It can be seen from (12) that a positive definite Hermitian matrix is added to the covariance matrix estimation error by applying penalties according to regularization method in order to guarantee the stability of solution.

4. DOA Estimation

In this section, we will utilize the minimization of the modified covariance matching criterion to estimate parameter $\tilde{\boldsymbol{\theta}}$. Let

$$\tilde{\boldsymbol{\theta}}_s = \arg \min_{\tilde{\boldsymbol{\theta}}} \left\| \widehat{\mathbf{R}}^{-1/2} \left(\widehat{\mathbf{R}} - \mathbf{R}(\tilde{\boldsymbol{\theta}}) + \frac{\mathbf{I}_N}{2} \right) \mathbf{R}^{-1/2}(\tilde{\boldsymbol{\theta}}) \right\|^2 \quad (13)$$

be the optimal solution of $\tilde{\boldsymbol{\theta}}$ in the structure model of (8). Our goal is to utilize the modified covariance matching criterion for an estimate that is asymptotically equal to $\tilde{\boldsymbol{\theta}}_s$. By the properties of the trace and the Hermitian matrix, a derivation process is shown as follows, where we omit the dependence on $\tilde{\boldsymbol{\theta}}$ for notational convenience:

$$\begin{aligned}f &= \text{Tr} \left[\widehat{\mathbf{R}}^{-1} \left(\widehat{\mathbf{R}} - \mathbf{R} + \frac{\mathbf{I}}{2} \right) \mathbf{R}^{-1} \left(\widehat{\mathbf{R}} - \mathbf{R} + \frac{\mathbf{I}}{2} \right) \right] \\ &= \text{Tr} \left[\widehat{\mathbf{R}}^{-1}\mathbf{R} - \widehat{\mathbf{R}}^{-1} + \left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2}\widehat{\mathbf{R}}^{-1/2} \right) \right. \\ &\quad \left. \cdot \mathbf{R}^{-1/2} \left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2}\widehat{\mathbf{R}}^{-1/2} \right) \mathbf{R}^{-1/2} \right].\end{aligned} \quad (14)$$

Due to

$$\text{Tr}[\widehat{\mathbf{R}}^{-1}\mathbf{R}] = \sum_{k=1}^K (\mathbf{c}_k^* \widehat{\mathbf{R}}^{-1} \mathbf{c}_k) \tilde{\theta}_k \quad (15)$$

we can deduce that the minimization of f is equal to the minimization of h :

$$\begin{aligned}h &= \text{Tr} \left[\left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2}\widehat{\mathbf{R}}^{-1/2} \right) \mathbf{R}^{-1/2} \left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2}\widehat{\mathbf{R}}^{-1/2} \right) \mathbf{R}^{-1/2} \right] \\ &\quad + \sum_{k=1}^K (\mathbf{c}_k^* \widehat{\mathbf{R}}^{-1} \mathbf{c}_k) \tilde{\theta}_k.\end{aligned} \quad (16)$$

Then, we will demonstrate that the minimization of h in (16) with respect to $\{\tilde{\theta}_k\}_{k=1}^K$ is an SDP. To prove this fact, One assumes

$$\widehat{\mathbf{R}}^{1/2} + \frac{1}{2}\widehat{\mathbf{R}}^{-1/2} = [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \cdots \quad \mathbf{r}_N]. \quad (17)$$

The h in (16) can be rewritten as

$$h = \sum_{k=1}^N \mathbf{r}_k^* \mathbf{R}^{-1} \mathbf{r}_k + \sum_{k=1}^K (\mathbf{c}_k^* \widehat{\mathbf{R}}^{-1} \mathbf{c}_k) \tilde{\theta}_k. \quad (18)$$

By the Schur complement, let $\{x_k\}_{k=1}^N$ be auxiliary variables satisfying

$$\begin{bmatrix} x_k & \mathbf{r}_k^* \\ \mathbf{r}_k & \mathbf{R} \end{bmatrix} \geq 0. \quad (19)$$

The equivalent form of this minimization problem is expressed as

$$\begin{aligned}\min_{x_k, \tilde{\theta}_k} \quad & \sum_{k=1}^N x_k + \sum_{k=1}^K (\mathbf{c}_k^* \widehat{\mathbf{R}}^{-1} \mathbf{c}_k) \tilde{\theta}_k \\ \text{s.t.} \quad & \tilde{\theta}_k \geq 0, \quad k = 1, 2, \dots, K,\end{aligned} \quad (20)$$

$$\begin{bmatrix} x_k & \mathbf{r}_k^* \\ \mathbf{r}_k & \mathbf{R} \end{bmatrix} \geq 0, \quad k = 1, 2, \dots, N.$$

It is easy to see that (20) is an SDP [34]. Many software packages can solve an SDP, but solving (16) as an SDP is not a good choice because SDP solvers have generally rather high computational complexity for the values of N , M , and K in the DOA estimation. To solve it effectively, we transform it into the constrained quadratic programming problem for reducing computational complexity, as described next.

Since a consistent estimation of $\tilde{\boldsymbol{\theta}}$ is given by (15), we can reformulate the minimization of h in (16) as the following constrained minimization by the Schur inequality of the trace:

$$\begin{aligned}\min_{\theta_k \geq 0} \quad & \text{Tr} \left[\left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2}\widehat{\mathbf{R}}^{-1/2} \right) \mathbf{R}^{-1/2} \left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2}\widehat{\mathbf{R}}^{-1/2} \right) \mathbf{R}^{-1/2} \right] \\ \text{s.t.} \quad & \mathbf{W}\mathbf{z} = \mathbf{1}_K,\end{aligned} \quad (21)$$

where $\mathbf{W}$ is a $K \times K$ diagonal matrix of $[\mathbf{C}^* \widehat{\mathbf{R}}^{-1} \mathbf{C}]^{-1/2}$ and $\mathbf{z} = [z_1 \ z_2 \ \cdots \ z_K]^T$ is a K -dimensional vector with $z_i = \theta_i^{-1/2}$, $i = 1, 2, \dots, K$. By substituting (8) into (21), the objective function in (21) can be rewritten as

$$\begin{aligned} & \text{Tr} \left[\left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2} \widehat{\mathbf{R}}^{-1/2} \right) (\mathbf{C}^*)^{-1/2} \mathbf{R}_s^{-1/2} \mathbf{C}^{-1/2} \right. \\ & \quad \cdot \left. \left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2} \widehat{\mathbf{R}}^{-1/2} \right) (\mathbf{C}^*)^{-1/2} \mathbf{R}_s^{-1/2} \mathbf{C}^{-1/2} \right] \\ & = \text{Tr} \left[\mathbf{C}^{-1/2} \left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2} \widehat{\mathbf{R}}^{-1/2} \right) (\mathbf{C}^*)^{-1/2} \mathbf{R}_s^{-1/2} \right. \\ & \quad \cdot \left. \mathbf{C}^{-1/2} \left(\widehat{\mathbf{R}}^{1/2} + \frac{1}{2} \widehat{\mathbf{R}}^{-1/2} \right) (\mathbf{C}^*)^{-1/2} \mathbf{R}_s^{-1/2} \right]. \end{aligned} \quad (22)$$

Based on the following equation:

$$\text{Tr} [\mathbf{M} \mathbf{T}(\mathbf{d}) \mathbf{N} \mathbf{T}(\mathbf{d})] = \mathbf{d}^* (\mathbf{M}^T \odot \mathbf{N}) \mathbf{d}, \quad (23)$$

where $\odot$ denotes the Schur-Hadamard product, $\mathbf{M}$ and $\mathbf{N}$ are both $K \times K$ matrices, $\mathbf{T}(\mathbf{d}) = \text{diag}(d_1 \ d_2 \ \cdots \ d_K)$, and $\mathbf{d} = [d_1 \ d_2 \ \cdots \ d_K]^T$, the objective function (22) can be written as

$$\text{Tr} [\mathbf{Q} \mathbf{R}_s^{-1/2} \mathbf{Q} \mathbf{R}_s^{-1/2}] = \mathbf{z}^* \mathbf{V} \mathbf{z}, \quad (24)$$

where $\mathbf{Q} = \mathbf{C}^{-1/2} (\widehat{\mathbf{R}}^{1/2} + (1/2) \widehat{\mathbf{R}}^{-1/2}) (\mathbf{C}^*)^{-1/2}$ and $\mathbf{V} = \mathbf{Q}^T \odot \mathbf{Q}$. Hence, the minimization problem in (21) is transformed into the following form:

$$\begin{aligned} \min_{\mathbf{z}} \quad & f(\mathbf{z}) = \mathbf{z}^* \mathbf{V} \mathbf{z} \\ \text{s.t.} \quad & \Pi(\mathbf{z}) = \mathbf{W} \mathbf{z} - \mathbf{1}_K = \mathbf{0}_K \end{aligned} \quad (25)$$

which is a typical constrained quadratic programming problem.

It is well known that an important class of methods for solving the constrained quadratic programming problem is to form the auxiliary function. To solve (25), we form the following augmented Lagrange function with respect to (25):

$$\mathbf{p}_\sigma(\mathbf{z}, \bar{\boldsymbol{\mu}}) = f(\mathbf{z}) + \bar{\boldsymbol{\mu}}^* \Pi(\mathbf{z}) + \frac{1}{2} \sigma \|\Pi(\mathbf{z})\|^2, \quad (26)$$

where $\bar{\boldsymbol{\mu}}$ is the asymptotical solution of the Lagrange multiplier of (25) and σ is a penalty factor.

By setting the gradient and the Hessian matrix of (26) with respect to $\mathbf{z}$ to zero, we have

$$\begin{aligned} \nabla_{\mathbf{z}} \mathbf{p}_\sigma(\mathbf{z}, \bar{\boldsymbol{\mu}}) &= \nabla_{\mathbf{z}} l(\mathbf{z}, \bar{\boldsymbol{\mu}}) + \sigma \mathbf{q}(\mathbf{z}) \Pi(\mathbf{z}), \\ \nabla_{\mathbf{z}\mathbf{z}}^2 \mathbf{p}_\sigma(\mathbf{z}, \bar{\boldsymbol{\mu}}) &= \nabla_{\mathbf{z}\mathbf{z}}^2 l(\mathbf{z}, \bar{\boldsymbol{\mu}}) + \sigma \sum_{i=1}^K \Pi_i(\mathbf{z}) \nabla^2 \Pi(\mathbf{z}) \\ &\quad + \sigma \mathbf{q}(\mathbf{z}) \mathbf{q}^*(\mathbf{z}), \end{aligned} \quad (27)$$

where $l(\mathbf{z}, \bar{\boldsymbol{\mu}}) = f(\mathbf{z}) + \bar{\boldsymbol{\mu}}^* \Pi(\mathbf{z})$ and $\mathbf{q}(\mathbf{z}) = \nabla \Pi^*(\mathbf{z}) = [\nabla \Pi_1(\mathbf{z}) \ \nabla \Pi_2(\mathbf{z}) \ \cdots \ \nabla \Pi_K(\mathbf{z})]$. One assumes that

$$\mathbf{b}(\mathbf{z}, \bar{\boldsymbol{\mu}}) = \nabla_{\mathbf{z}\mathbf{z}}^2 l(\mathbf{z}, \bar{\boldsymbol{\mu}}) + \sigma \sum_{i=1}^K \Pi_i(\mathbf{z}) \nabla^2 \Pi(\mathbf{z}). \quad (28)$$

Applying the Newton method, we obtain

$$\nabla_{\mathbf{z}\mathbf{z}}^2 \mathbf{p}_{\sigma_k}(\mathbf{z}_k, \bar{\boldsymbol{\mu}}_k)(\mathbf{z}_{k+1} - \mathbf{z}_k) = -\nabla_{\mathbf{z}} \mathbf{p}_{\sigma_k}(\mathbf{z}_k, \bar{\boldsymbol{\mu}}_k). \quad (29)$$

For notational convenience, we assume that $\mathbf{q}_k = \mathbf{q}(\mathbf{z}_k)$, $\nabla f_k = \nabla f(\mathbf{z}_k)$, $\Pi_k = \Pi(\mathbf{z}_k)$, $\mathbf{b}_k = \mathbf{b}(\mathbf{z}_k, \bar{\boldsymbol{\mu}}_k)$, $\mathbf{d}_k = \mathbf{z}_{k+1} - \mathbf{z}_k$ and we get the following equation by (29):

$$\mathbf{b}_k \mathbf{d}_k + \sigma_k \mathbf{q}_k \mathbf{q}_k^* \mathbf{d}_k + \nabla_{\mathbf{z}} l(\mathbf{z}_k, \bar{\boldsymbol{\mu}}_k) + \sigma_k \mathbf{q}_k \Pi_k = \mathbf{0}. \quad (30)$$

Assuming that the inverse of $\mathbf{b}_k$ exists, by left-multiplying (30) by $(1/\sigma_k) \mathbf{q}_k^* \mathbf{b}_k^{-1}$, we have

$$\begin{aligned} & \left(\frac{\mathbf{I}_K}{\sigma_k} + \mathbf{q}_k^* \mathbf{b}_k^{-1} \mathbf{q}_k \right) \mathbf{q}_k^* \mathbf{d}_k \\ & = -\mathbf{q}_k^* \mathbf{b}_k^{-1} \mathbf{q}_k \Pi_k - \frac{1}{\sigma_k} \mathbf{q}_k^* \mathbf{b}_k^{-1} \nabla_{\mathbf{z}} l(\mathbf{z}_k, \bar{\boldsymbol{\mu}}_k). \end{aligned} \quad (31)$$

It follows from (31) that

$$\mathbf{q}_k^* \mathbf{b}_k = -\Pi_k + \frac{\boldsymbol{\alpha}_k}{\sigma_k}, \quad (32)$$

where $\boldsymbol{\alpha}_k$ satisfies

$$\left(\frac{\mathbf{I}_K}{\sigma_k} + \mathbf{q}_k^* \mathbf{b}_k^{-1} \mathbf{q}_k \right) \boldsymbol{\alpha}_k = -\mathbf{q}_k^* \mathbf{b}_k^{-1} \nabla_{\mathbf{z}} l(\mathbf{z}_k, \bar{\boldsymbol{\mu}}_k) + \Pi_k. \quad (33)$$

By substituting (32) into (30), we have

$$\mathbf{d}_k = -\mathbf{b}_k^{-1} (\mathbf{q}_k \boldsymbol{\alpha}_k + \nabla_{\mathbf{z}} l(\mathbf{z}_k, \bar{\boldsymbol{\mu}}_k)). \quad (34)$$

Both the multiplier factor and the penalty factor are determined with difficulty for utilizing the augmented Lagrange function to solve the constrained quadratic programming problem. Hence, an updated sequence for the multiplier factor is given in terms of Proposition 1.

Proposition 1. One assumes that $\boldsymbol{\mu}(\mathbf{z})$ is the optimal solution of the problem $\min_{\boldsymbol{\mu} \in \mathcal{C}^K} \|\mathbf{q}(\mathbf{z}) \boldsymbol{\mu} + \nabla f(\mathbf{z})\|^2$. Then, the following equation holds:

$$\boldsymbol{\mu}(\mathbf{z}_k) = \bar{\boldsymbol{\mu}}_k + \boldsymbol{\alpha}_k + \boldsymbol{\beta}_k \mathbf{d}_k, \quad (35)$$

where $\boldsymbol{\beta}_k = (\mathbf{q}_k^* \mathbf{q}_k)^{-1} \mathbf{q}_k^* \mathbf{b}_k$.

Proof. See Appendix A. $\square$

It can be deduced from Proposition 1 that $\bar{\boldsymbol{\mu}}_k + \boldsymbol{\alpha}_k$ is referred to as the next iteration of $\bar{\boldsymbol{\mu}}_k$. We apply a heuristic update sequence for the penalty factor to achieve a stable solution. If the k th iterative solution $\mathbf{z}_k$ is closer to the feasible region than the previous solution $\mathbf{z}_{k-1}$, the penalty factor is decreased. Inversely, we increase the penalty factor when $\mathbf{z}_k$ is not closer to the feasible region.

The specific steps of solving by the augmented Lagrange method are given as follows.

Initialization: set $k = 1$, $\mathbf{z}_1 \in \mathcal{C}^K$, $\sigma_1 > 0$, $\tau \in (0, 1)$, and $\bar{\boldsymbol{\mu}}_1 = -(\mathbf{q}_1^* \mathbf{q}_1)^{-1} \mathbf{q}_1^* \nabla f_1$.

(1) Calculate α_k and $\mathbf{d}_k$ in terms of (33) and (34). If $\|\mathbf{d}_k\| \approx 0$, $\mathbf{z}_k$ is the KKT point of the problem (25); then stop iteration.

(2) We use Armijo search method to find the maximum of t_k satisfying

$$\mathbf{p}_{\sigma_k}(\mathbf{z}_k + t_k \mathbf{d}_k, \bar{\boldsymbol{\mu}}_k) \leq \mathbf{p}_{\sigma_k}(\mathbf{z}_k, \bar{\boldsymbol{\mu}}_k) + \tau t_k \nabla \mathbf{p}_{\sigma_k}^*(\mathbf{z}_k, \bar{\boldsymbol{\mu}}_k) \mathbf{d}_k. \quad (36)$$

(3) Set $\mathbf{z}_{k+1} = \mathbf{z}_k + t_k \mathbf{d}_k$ and update the multiplier factor and penalty factor, respectively:

$$\begin{aligned} \bar{\boldsymbol{\mu}}_{k+1} &= \bar{\boldsymbol{\mu}}_k + \alpha_k, \\ \sigma_{k+1} &= \begin{cases} 2\sigma_k & \text{if } \|\Pi(\mathbf{z}_{k+1})\| > \|\Pi(\mathbf{z}_k)\|, \\ \frac{1}{2}\sigma_k & \text{if } \|\Pi(\mathbf{z}_{k+1})\| \leq \|\Pi(\mathbf{z}_k)\|. \end{cases} \end{aligned} \quad (37)$$

(4) Set $k = k + 1$ and return to step (1).

5. Performance Analysis

5.1. Cramér-Rao Bound. In this subsection, the closed-form expression for the CRB of the proposed method with complex white Gaussian noise after the prewhitened process is illustrated. The bound of the noise variance estimation can be computed separately as $\text{CRB}_{\eta} = 1/NM$ (see [35]). The remaining parameters consist of an unknown vector $\boldsymbol{\varphi} = [\tilde{\boldsymbol{\theta}}^T \mathbf{s}^T]^T$. It is not an easy task to get the CRB of the unknown parameters. However, fortunately, [36, 37] have provided a critical inspiration for the derivation in this paper.

First, the likelihood function is given by

$$\begin{aligned} L\left(\frac{\mathbf{y}}{\mathbf{s}(t_j)}, \tilde{\boldsymbol{\theta}}\right) &= \frac{1}{(2\pi)^{MN}(1/2)^{MN}} \\ &\cdot \exp \left\{ -\sum_{j=1}^M (\mathbf{y}(t_j) - \mathbf{B}\Phi\mathbf{A}\mathbf{s}(t_j))^* \right. \\ &\quad \left. \times (\mathbf{y}(t_j) - \mathbf{B}\Phi\mathbf{A}\mathbf{s}(t_j)) \right\}. \end{aligned} \quad (38)$$

Thus, the log-likelihood function is

$$\begin{aligned} \ln L\left(\frac{\mathbf{y}}{\mathbf{s}(t_j)}, \tilde{\boldsymbol{\theta}}\right) &= \text{const} - \sum_{j=1}^M (\mathbf{y}(t_j) - \mathbf{B}\Phi\mathbf{A}\mathbf{s}(t_j))^* \\ &\quad \times (\mathbf{y}(t_j) - \mathbf{B}\Phi\mathbf{A}\mathbf{s}(t_j)). \end{aligned} \quad (39)$$

Then, the partial derivatives of (39) with respect to $\tilde{\boldsymbol{\theta}}$, $\mathbf{s}_r(t_j) = \text{Re}\{\mathbf{s}(t_j)\}$ and $\mathbf{s}_i(t_j) = \text{Im}\{\mathbf{s}(t_j)\}$ are given by

$$\begin{aligned} \frac{\partial \ln L}{\partial \tilde{\boldsymbol{\theta}}} &= 2 \sum_{j=1}^M \text{Re} \left\{ \mathbf{S}_j^* \mathbf{U}^* \Phi^* \mathbf{B}^* \mathbf{n}(t_j) \right\}, \\ \frac{\partial \ln L}{\partial \mathbf{s}_r(t_j)} &= 2 \text{Re} \left\{ \mathbf{A}^* \Phi^* \mathbf{B}^* \mathbf{n}(t_j) \right\}, \\ \frac{\partial \ln L}{\partial \mathbf{s}_i(t_j)} &= 2 \text{Im} \left\{ \mathbf{A}^* \Phi^* \mathbf{B}^* \mathbf{n}(t_j) \right\}, \end{aligned} \quad (40)$$

where $\mathbf{S}_j = \text{diag}(\mathbf{s}(t_j))$ and $\mathbf{U} = [d\mathbf{a}(\tilde{\theta}_1)/d\tilde{\theta}_1 \quad d\mathbf{a}(\tilde{\theta}_2)/d\tilde{\theta}_2 \quad \cdots \quad d\mathbf{a}(\tilde{\theta}_K)/d\tilde{\theta}_K]$. Following [35, 37], we can obtain the Fisher information matrix (FIM) as follows:

$$\begin{aligned} \text{FIM} &= \mathbf{\Lambda} - [\Delta_{1r} \quad \Delta_{1i} \quad \Delta_{2r} \quad \cdots \quad \Delta_{Mi} \quad \Delta_{Mr}] \\ &\quad \cdot \begin{bmatrix} \mathbf{G}_r & -\mathbf{G}_i & 0 & \cdots & 0 \\ \mathbf{G}_i & \mathbf{G}_r & & \ddots & \vdots \\ 0 & & \ddots & & 0 \\ \vdots & & & \ddots & \mathbf{G}_r & -\mathbf{G}_i \\ 0 & \cdots & 0 & \mathbf{G}_i & \mathbf{G}_r \end{bmatrix} \cdot \begin{bmatrix} \Delta_{1r} \\ \Delta_{1i} \\ \Delta_{2r} \\ \vdots \\ \Delta_{Mr} \\ \Delta_{Mi} \end{bmatrix}, \end{aligned} \quad (41)$$

where

$$\begin{aligned} E \left\{ \frac{\partial \ln L}{\partial \tilde{\boldsymbol{\theta}}} \left(\frac{\partial \ln L}{\partial \tilde{\boldsymbol{\theta}}} \right)^T \right\} &= \mathbf{\Lambda} \\ &= 2 \sum_{j=1}^M \text{Re} \left\{ \mathbf{S}_j^* \mathbf{U}^* \Phi^* \mathbf{B}^* \mathbf{B} \Phi \mathbf{U} \mathbf{S}_j \right\}, \\ E \left\{ \frac{\partial \ln L}{\partial \mathbf{s}_r(t_k)} \left(\frac{\partial \ln L}{\partial \mathbf{s}_i(t_p)} \right)^T \right\} &= -\mathbf{G}_i^{-1} = -\text{Im} \left\{ \mathbf{G}^{-1} \right\} \\ &= -\text{Im} \left\{ 2\mathbf{A}^* \Phi^* \mathbf{B}^* \mathbf{B} \Phi \mathbf{A} \right\} \delta_{kp}, \\ E \left\{ \frac{\partial \ln L}{\partial \mathbf{s}_r(t_k)} \left(\frac{\partial \ln L}{\partial \mathbf{s}_r(t_p)} \right)^T \right\} &= E \left\{ \frac{\partial \ln L}{\partial \mathbf{s}_i(t_k)} \left(\frac{\partial \ln L}{\partial \mathbf{s}_i(t_p)} \right)^T \right\} \\ &= \mathbf{G}_r^{-1} = \text{Re} \left\{ \mathbf{G}^{-1} \right\} \\ &= \text{Re} \left\{ 2\mathbf{A}^* \Phi^* \mathbf{B}^* \mathbf{B} \Phi \mathbf{A} \right\} \delta_{kp}, \\ E \left\{ \frac{\partial \ln L}{\partial \mathbf{s}_r(t_j)} \left(\frac{\partial \ln L}{\partial \tilde{\boldsymbol{\theta}}} \right)^T \right\} &= \text{Re} \left\{ \Delta_j \right\} = \Delta_{jr} \\ &= \text{Re} \left\{ 2\mathbf{A}^* \Phi^* \mathbf{B}^* \mathbf{B} \Phi \mathbf{U} \mathbf{S}_j \right\}, \\ &\quad j = 1, 2, \dots, M, \end{aligned}$$

$$\begin{aligned}
E \left\{ \frac{\partial \ln L}{\partial \mathbf{s}_i(t_j)} \left(\frac{\partial \ln L}{\partial \tilde{\boldsymbol{\theta}}} \right)^T \right\} &= \text{Im} \{ \Delta_j \} = \Delta_{ji} \\
&= \text{Im} \{ 2\mathbf{A}^* \boldsymbol{\Phi}^* \mathbf{B}^* \mathbf{B} \boldsymbol{\Phi} \mathbf{U} \mathbf{s}_j \}, \\
j &= 1, 2, \dots, M.
\end{aligned} \quad (42)$$

It is well known that

$$\begin{bmatrix} \Delta_r^T & \Delta_i^T \end{bmatrix} \begin{bmatrix} \mathbf{G}_r & -\mathbf{G}_i \\ \mathbf{G}_i & \mathbf{G}_r \end{bmatrix} \begin{bmatrix} \Delta_r \\ \Delta_i \end{bmatrix} = \text{Re} \{ \Delta^* \mathbf{G} \Delta \}. \quad (43)$$

It can be deduced from (41) and (43) that

$$\begin{aligned}
\text{CRB}(\tilde{\boldsymbol{\theta}}) &= \text{FIM}^{-1} \\
&= \frac{1}{2} \left\{ \sum_{j=1}^M \text{Re} \{ \mathbf{S}_j^* \mathbf{U}^* \mathbf{H}^* (\mathbf{I} - \mathbf{H} \mathbf{A} (\mathbf{H} \mathbf{A})^+) \mathbf{H} \mathbf{U} \mathbf{s}_j \} \right\}^{-1},
\end{aligned} \quad (44)$$

where $\mathbf{H} = \mathbf{B} \boldsymbol{\Phi}$ is a $N \times P$ matrix and $(\cdot)^+$ denotes pseudoinverse. Note that the CRB in CS is affected not only by the conventional factors, for example, SNR, array structure and signal relative location, but also by the measurement matrix.

5.2. Performance Guarantee. In CS, the RIP has been deeply studied for the joint sparse recovery by minimizing the l_1 norm. We say that the matrix $\mathbf{C} \in \mathbb{C}^{N \times K}$ obeys the RIP of the order L if there exists a constant $\delta \in (0, 1)$ satisfying

$$(1 - \delta) \|\mathbf{s}\|_2^2 \ll \|\mathbf{C}\mathbf{s}\|_2^2 \ll (1 + \delta) \|\mathbf{s}\|_2^2. \quad (45)$$

Therefore, all submatrices of $\mathbf{C}$ with L columns are uniformly well conditioned. The restricted isometry constant (RIC) of the order L , described as $\delta_L(\mathbf{C})$, is the smallest δ that satisfies (45) and $\delta_L(\mathbf{C})$ satisfies

$$\delta_L(\mathbf{C}) = \max_{|J'|=L} \|\mathbf{C}_{J'}^* \mathbf{C}_{J'} - \mathbf{I}_L\|, \quad (46)$$

where J' is a L -dimensional subsupport of $J = [1, 2, \dots, K]$ and $\mathbf{C}_{J'}$ denotes the submatrix of $\mathbf{C}$ with columns indexed by J' . Note that the condition satisfying the RIP is so demanding that its applications are limited. Therefore, we should make use of a new version of the RIP, which is called the weak-1 RIP [38], to control the size of the recovery error. The weak-1 RIP is given in the following form:

$$(1 - \alpha) \|\mathbf{v}\|_2^2 \ll \|\mathbf{C}\mathbf{v}\|_2^2 \ll (1 + \alpha) \|\mathbf{v}\|_2^2 \quad (47)$$

for all $\mathbf{v}$ supported on U , where the cardinality of the set U is $L+1$. If the matrix $\mathbf{C}$ satisfies the weak-1 RIP, it can be deduced that $0 \leq \alpha \leq \phi_{L+1}(\mathbf{C}_U)$, where $\phi_k(\mathbf{C}_U)$ denotes the k th largest eigenvalue of $\mathbf{C}_U$. The corresponding weak-1 RIC is given by

$$\alpha_{L+1}^w(\mathbf{C}; U) = \min \phi_{L+1}(\mathbf{C}_U). \quad (48)$$

In this paper, when the estimation quality is imperfect, especially in the unfavorable case, the support is no longer identified by the algebraic property of $\mathbf{R}$. Hence, a new performance guarantee is given in terms of the weak-1 RIP in the following proposition.

Proposition 2. One assumes that $\mathbf{F} = \widehat{\mathbf{R}}^{1/2} + (1/2)\widehat{\mathbf{R}}^{-1/2}$ and $\mathbf{R}^{-1} = \boldsymbol{\Psi} \mathbf{R}_s^{-1}(\tilde{\boldsymbol{\theta}}) \boldsymbol{\Psi}^*$ where $\boldsymbol{\Psi} = [\psi_1 \ \psi_2 \ \dots \ \psi_K]$. $\tilde{\boldsymbol{\theta}}$ is an estimation that is asymptotically equal to $\tilde{\boldsymbol{\theta}}_s$ such that $|\text{tr}(\mathbf{R}^{-1}(\tilde{\boldsymbol{\theta}})) - \text{tr}(\mathbf{R}^{-1}(\tilde{\boldsymbol{\theta}}_s))| \leq \eta$ for $\eta \in (0, 1/2\|\mathbf{F}\|_2^2)$. Let J_0 and $J_1 = \{j_i\}_{i=1}^L$ be the L -dimensional supports that consisted of the indexes of L largest elements in $\tilde{\boldsymbol{\theta}}_s$ and $\tilde{\boldsymbol{\theta}}$, respectively. If the matrix $\mathbf{R}$ satisfies

$$\alpha_{L+1}^w(\mathbf{R}; J_0) > \kappa \quad (49)$$

for

$$\kappa \geq \frac{L\|\mathbf{F}\|_2^2 \sum_{j \in J_0} \psi_j^* \psi_j}{1 - 2\eta\|\mathbf{F}\|_2^2} \quad (50)$$

the support can be identified.

Proof. See Appendix B. $\square$

It follows from Proposition 2 that the performance guarantee of the proposed method requires a mild condition in the unfavorable case. However, in the favorable case, the performance guarantee only requires a much milder condition, $\alpha_{L+1}^w(\mathbf{R}; J_0) > 0$, which is an algebraic condition.

6. Simulation Results

In this section, the performance of the proposed method is illustrated by several simulation results and compared with that of CS-MUSIC, SPICE, and CRB under the condition that the number of sources is unknown. We consider the spatial signal impinging on the uniform linear array (ULA) with interspacing $\lambda/2$ where λ denotes the wavelength of source. In the ULA case, the steering vector corresponding to the DOA equal to θ_k is given by

$$\mathbf{a}(\theta_k) = [1 \ e^{-j\pi \sin(\theta_k)} \ \dots \ e^{-j(P-1)\pi \sin(\theta_k)}]^T, \quad (51)$$

where the number of the array elements is set to be $P = 8$.

In the simulation, the average root mean square error (RMSE) of the DOA estimation with 50 Monte Carlo runs is defined as the significant performance index:

$$\text{RMSE} = \left[\sum_{m=1}^{50} \sum_{l=1}^L \frac{(\bar{\theta}_{lm} - \theta_l)^2}{50L} \right]^{1/2}, \quad (52)$$

where $\bar{\theta}_{lm}$ is the estimate of θ_l in the m th run.

The resolution of the grid is closely related to the precision of the DOA estimation. A coarse grid can lead to poor precision, but a too fine grid increases computational complexity. Therefore, an adaptive grid refinement method is used to

FIGURE 1: Superimposed spatial spectra of CS-MUSIC, SPICE, and AULMC in 10 Monte Carlo runs, where the red vertical dashed lines denote the true DOAs. (a) CS-MUSIC, (b) SPICE and (c) AULMC.

balance the tradeoff between precision and computational complexity. In the simulation, we make a coarse grid with 1° step in the range of -90° to 90° and perform a local fine grid in the vicinity of locations obtained by using the coarse grid.

In the first simulation, we display the superimposed spatial spectra of three algorithms in 10 Monte Carlo runs in the scenario with low SNR, small number of snapshots, and five sources impinging from $[-40.3^\circ -25^\circ 5.2^\circ 15.2^\circ 38.1^\circ]$, respectively where two most closely spaced sources are correlated and the remaining sources are uncorrelated. The spatial spectra are shown in Figure 1 with 3 dB SNR and 50 snapshots. The following facts can be acquired from Figure 1

FIGURE 2: RMSE of the DOA estimation versus SNR for 50 snapshots.

FIGURE 3: RMSE of the DOA estimation versus number of snapshots for 5 dB SNR.

as follows: the spatial spectrum obtained by CS-MUSIC suffers from severe interference at the true directions, especially at two most closely spaced correlated sources whose bias is clearly seen from the insert. Two most closely spaced correlated sources can be resolved by SPICE (note that peaks in the spatial spectrum are much larger than 2, but they are cut off at 2 to use the same scale as the other figures in Figure 1), but SPICE can yield false peaks and slight bias in the vicinity of the correlated sources and uncorrelated sources, respectively. The proposed method AULMC yields a nearly ideal spatial spectrum and provides a precise estimation for all the sources. In summary, AULMC outperforms CS-MUSIC and SPICE in terms of the spatial spectrum in the scenario with low SNR, small number of snapshots, and closely spaced correlated sources.

We analyze the RMSE of three algorithms under different conditions in the second simulation. The source model is the same as the first simulation. Figure 2 shows the RMSE as a function of SNR of all the algorithms and CRB in 50 Monte Carlo runs for the fixed number of snapshots 50, whereas the RMSE versus number of snapshots is shown in Figure 3 for the fixed SNR 5 dB in 50 Monte Carlo runs. Based on Figures 2

FIGURE 4: RMSE of the DOA estimation versus angle separation, where the SNR is 13 dB and the number of snapshots is 100.

TABLE 1: Comparison of computation time.

Number of snapshots	Time (sec)			
	AULMC	CS-MUSIC	SPICE	SDP solver
50	0.0651	0.0569	0.0601	0.7526
75	0.0776	0.0604	0.0714	0.8871
100	0.0860	0.0640	0.0819	1.0268
125	0.0974	0.0691	0.0923	1.3018

and 3, we can draw the conclusions that the RMSE of AULMC is smaller than those of other two algorithms and AULMC has the more significant performance advantages than the other two algorithms, especially in the scenarios with low SNR or small number of snapshots. One possible explanation is that AULMC can give the stable estimation in every Monte Carlo run. It can be also seen that the RMSE is close to the CRB with the increase of SNR and the number of snapshots.

In Figure 4, we display the relation between the RMSE and angle separation of correlated sources which can illustrate the resolving ability. Let two correlated sources at angles 20° and $20^\circ + \Delta\theta$, where the step of $\Delta\theta$ is 1° , be impinged on the ULA. The SNR is 13 dB and the number of snapshots is 100. It can be seen from Figure 4 that when angle separation is 2° , AULMC fails; however, AULMC can still provide a precise estimation as long as the angle separation is no less than 3° and has higher resolution than the other two algorithms.

Finally, the computation time of different algorithms versus number of snapshots is shown in Table 1 for comparing the efficiency of these algorithms and SDP solver. Two correlated sources impinge on the ULA at 20° and 25° . The SNR is fixed at 13 dB. The computation time is obtained by the MATLAB 7.8 (R2009a) on a 2.8 GHz 4 GB PC. For AULMC, the computation time is mainly spent on the iterations of augmented Lagrange.

It can be seen from Table 1 that the computation time of SDP solver is the longest, and although the computation time of AULMC is longer than that of other two algorithms, it is comparable. Moreover, it is worth noting that the

performance of AULMC is much better than that of CS-MUSIC or SPICE.

7. Conclusion

A novel augmented Lagrange based on modified covariance matching criterion method for DOA estimation is proposed in CS. It is proved that the problem of minimizing the modified covariance matching criterion is an SDP, which can be transformed into the constrained quadratic programming problem solved by the augmented Lagrange method. A detailed derivation for the CRB and a theoretical performance guarantee for identifying the support are provided. Simulation results show that AULMC outperforms CS-MUSIC and SPICE in terms of the spatial spectrum and has more precise estimation as well as higher resolution, especially in the scenarios with low SNR, small number of snapshots, and closely spaced correlated sources.

Appendices

A. Proof of Proposition 1

Due to $\mu(\mathbf{z}_k) = -(\mathbf{q}_k^* \mathbf{q}_k)^{-1} \mathbf{q}_k^* \nabla f_k$, we have

$$\mathbf{q}_k^* \mathbf{q}_k \mu(\mathbf{z}_k) = -\mathbf{q}_k^* \nabla f_k. \quad (\text{A.1})$$

With (34) and (A.1), we get

$$\mathbf{q}_k^* \mathbf{q}_k [\mu(\mathbf{z}_k) - \bar{\mu}_k - \alpha_k] = \mathbf{q}_k^* \mathbf{b}_k \mathbf{d}_k. \quad (\text{A.2})$$

Therefore, it is implied by (A.2) that (35) holds and the proof of Proposition 1 is completed.

B. Proof of Proposition 2

The support J_0 can be identified if

$$\begin{aligned} & \max_{j \in J_0} \frac{\text{Tr} [\mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}}) \mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}})]}{\text{Tr} [\mathbf{F}^2]} \\ & < \min_{j \in J \setminus J_0} \frac{\text{Tr} [\mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}}) \mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}})]}{\text{Tr} [\mathbf{F}^2]}. \end{aligned} \quad (\text{B.1})$$

By the Cauchy-Schwarz inequality of the trace, we have

$$\begin{aligned} & \left| \frac{\text{Tr} [(\mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}}))^2] - \text{Tr} [(\mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}}_s))^2]}{\text{Tr} [\mathbf{F}^2]} \right| \\ & \leq \frac{\text{Tr} [\mathbf{F}^2] \cdot |\text{Tr} [\mathbf{R}^{-1} (\hat{\boldsymbol{\theta}})] - \text{Tr} [\mathbf{R}^{-1} (\hat{\boldsymbol{\theta}}_s)]|}{\text{Tr} [\mathbf{F}^2]} \quad (\text{B.2}) \\ & = |\text{Tr} [\mathbf{R}^{-1} (\hat{\boldsymbol{\theta}})] - \text{Tr} [\mathbf{R}^{-1} (\hat{\boldsymbol{\theta}}_s)]| \leq \eta. \end{aligned}$$

Then, for all $j \in J_0$ we have

$$\begin{aligned} \frac{\text{Tr} \left[\left(\mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}}) \right)^2 \right]}{\text{Tr} [\mathbf{F}^2]} &\leq \eta + \frac{\text{Tr} \left[\left(\mathbf{F} \mathbf{R}^{-1/2} (\tilde{\boldsymbol{\theta}}_s) \right)^2 \right]}{\text{Tr} [\mathbf{F}^2]} \\ &\leq \eta + \frac{\text{Tr} [\mathbf{F}^2] \text{Tr} [\mathbf{R}^{-1} (\tilde{\boldsymbol{\theta}}_s)]}{\text{Tr} [\mathbf{F}^2]} \\ &= \eta + \frac{\sum_{j \in J_0} \boldsymbol{\psi}_j^* \boldsymbol{\psi}_j}{\tilde{\theta}_j} \\ &\leq \eta + \frac{L \sum_{j \in J_0} \boldsymbol{\psi}_j^* \boldsymbol{\psi}_j}{\alpha_{L+1}^w (\mathbf{R}; J_0)}. \end{aligned} \quad (\text{B.3})$$

Thus, an upper bound on the left-hand side of (B.1) is given by

$$\max_{j \in J_0} \frac{\text{Tr} \left[\left(\mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}}) \right)^2 \right]}{\text{Tr} [\mathbf{F}^2]} \leq \eta + \frac{L \sum_{j \in J_0} \boldsymbol{\psi}_j^* \boldsymbol{\psi}_j}{\alpha_{L+1}^w (\mathbf{R}; J_0)}. \quad (\text{B.4})$$

Similarly, we can obtain

$$\begin{aligned} \frac{\text{Tr} \left[\left(\mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}}) \right)^2 \right]}{\text{Tr} [\mathbf{F}^2]} &\geq \frac{\text{Tr} [\mathbf{F}^2 \mathbf{R}^{-1} (\tilde{\boldsymbol{\theta}}_s)]}{\text{Tr} [\mathbf{F}^2]} - \eta \\ &= \frac{\sum_{j \in J \setminus J_0} \boldsymbol{\psi}_j^* \mathbf{F}^2 \boldsymbol{\psi}_j / \tilde{\theta}_j}{\text{Tr} [\mathbf{F}^2]} - \eta \\ &\geq \frac{1}{\|\mathbf{F}\|_2^2} - \eta \end{aligned} \quad (\text{B.5})$$

for all $j \in J \setminus J_0$ and hence describe a lower bound on the right-hand side of (B.1) as

$$\min_{j \in J \setminus J_0} \frac{\text{Tr} \left[\left(\mathbf{F} \mathbf{R}^{-1/2} (\hat{\boldsymbol{\theta}}) \right)^2 \right]}{\text{Tr} [\mathbf{F}^2]} \geq \frac{1}{\|\mathbf{F}\|_2^2} - \eta. \quad (\text{B.6})$$

Combining (B.4) with (B.6), the support J_0 can be identified if $\mathbf{R}$ satisfies

$$\alpha_{L+1}^w (\mathbf{R}; J_0) > \kappa \quad (\text{B.7})$$

for

$$\kappa \geq \frac{L \|\mathbf{F}\|_2^2 \sum_{j \in J_0} \boldsymbol{\psi}_j^* \boldsymbol{\psi}_j}{1 - 2\eta \|\mathbf{F}\|_2^2}. \quad (\text{B.8})$$

Conflict of Interests

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

Acknowledgment

This work was supported in part by the National Science Foundation of China under Grant 61201410 and in part by Fundamental Research Funds for the Central Universities under Grand no. HEUCF130804.

References

- [1] J. D. Dai, W. C. Xu, and D. Zhao, "Real-valued DOA estimation for uniform linear array with unknown mutual coupling," *Signal Processing*, vol. 92, no. 9, pp. 2056–2065, 2012.
- [2] L. Bai, C.-Y. Peng, and S. Biswas, "Association of DOA estimation from two ULAs," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 6, pp. 1094–1101, 2008.
- [3] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of IEEE*, vol. 57, pp. 1408–1418, 1969.
- [4] J. P. Burg, "Maximum entropy spectral analysis," in *Proceedings of the 37th Annual International SEG meeting Oklahoma City*, 1967.
- [5] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [6] 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.
- [7] 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.
- [8] R. Bachl, "Forward-backward averaging technique applied to TLS-ESPRIT processing," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2691–2699, 1995.
- [9] Z. Sun and Z. Yang, "Study of nonlinear parameter identification using UKF and Maximum Likelihood method," in *Proceedings of the IEEE International Conference on Control Applications (CCA '10)*, pp. 671–676, September 2010.
- [10] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2436–2449, 1991.
- [11] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [12] E. T. Northardt, I. Bilik, and Y. I. Abramovich, "Spatial compressive sensing for direction-of-arrival estimation with bias mitigation via expected likelihood," *IEEE Transactions on Signal Processing*, vol. 61, no. 5, pp. 1183–1195, 2013.
- [13] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A. Massa, "Direction-of-arrival estimation through bayesian compressive sensing strategies," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 7, pp. 3828–3838, 2013.
- [14] 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.
- [15] A. C. Gürbüz, J. H. McClellan, and V. Cevher, "A compressive beamforming method," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 2617–2620, April 2008.
- [16] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *Proceedings of the IEEE/SP 15th Workshop on Statistical Signal Processing (SSP '09)*, pp. 626–629, September 2009.
- [17] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 35–47, 2010.

- [18] P. Stoica and P. Babu, "SPICE and LIKES: two hyperparameter-free methods for sparse-parameter estimation," *Signal Processing*, vol. 92, no. 7, pp. 1580–1590, 2012.
- [19] D. Malioutov, M. Çetin, 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.
- [20] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive MUSIC: revisiting the link between compressive sensing and array signal processing," *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 278–301, 2012.
- [21] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for Joint sparse recovery," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3613–3640, 2012.
- [22] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4395–4401, 2010.
- [23] 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.
- [24] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.
- [25] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '96)*, vol. 3, pp. 1688–1691, May 1996.
- [26] 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.
- [27] S. Yin, H. Luo, and S. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 2402–2411, 2014.
- [28] S. Yin, G. Wang, and H. Karimi, "Data-driven design of robust fault detection system for wind turbines," *Mechatronics*, 2013.
- [29] S. Yin, S. Ding, A. Haghani, and H. Hao, "Data-driven monitoring for stochastic systems and its application on batch process," *International Journal of Systems Science*, vol. 44, no. 7, pp. 1366–1376, 2013.
- [30] S. Yin, S. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic datadriven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.
- [31] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.
- [32] 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.
- [33] D. I. Svergun, "Determination of the regularization parameter in indirect-transform methods using perceptual criteria," *Journal of Applied Crystallography*, vol. 25, no. 4, pp. 495–503, 1992.
- [34] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, Society for Industrial and Applied Mathematics, 1994.
- [35] A. Leshem and A.-J. D. van Veen, "Direction-of-arrival estimation for constant modulus signals," *IEEE Transactions on Signal Processing*, vol. 47, no. 11, pp. 3125–3129, 1999.
- [36] 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.
- [37] 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.
- [38] E. J. Candès and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.