

# A new direction-of-arrival estimation method exploiting signal structure information

Bo Lin<sup>1</sup> · Jiying Liu<sup>1</sup> · Meihua Xie<sup>1</sup> · Jubo Zhu<sup>1</sup> ·  
Fengxia Yan<sup>1</sup>

Received: 27 October 2014 / Revised: 15 June 2015 / Accepted: 20 June 2015 /  
Published online: 2 July 2015  
© Springer Science+Business Media New York 2015

**Abstract** A new method is proposed to estimate the direction-of-arrival (DOA) based on uniform linear array sampling and named as sparsity and temporal correlation exploiting (SaTC-E). By exploiting the structure information of source signals, including spatial sparsity and temporal correlation of sources, SaTC-E accomplishes DOA estimation with superior performance via block sparse bayesian learning methodology and grid refined strategy. SaTC-E is applicable into time-varying manifold scenario, such as wideband sources, time-varying array, provided that the array manifold matrix is determinable. It has improved performance with some other merits, including superior resolution, requirement for a few snapshots, no knowledge of source number, and applicability to spatially and temporally corrected sources. Real data tests and numerical simulations are carried out to demonstrate the advantages of SaTC-E.

**Keywords** Direction-of-arrival (DOA) estimation · Block sparse bayesian learning (BSBL) · Temporal correlation · Spatial joint sparsity · Grid refined strategy

## 1 Introduction

Direction-of-arrival (DOA) estimation based on antenna array is an attractive issue in array signal processing and plays an important role in the application of radar, sonar, wireless communication, radio-astronomy. A number of high resolution methods have sprung up during the past years. As a representative of the classical super-resolution algorithms, the multiple signal classification (MUSIC; Schmidt 1986) has been used very widely. However, the performance of MUSIC deteriorates significantly with only a few snapshots or in the

---

This research was supported by the National Natural Science Foundation of China under Grants (Nos. 61205190, 61201332 and 61471369) and Basic Research Plan of NUDT (No. JC13-02-03).

---

✉ Bo Lin  
linbo@nudt.edu.cn

<sup>1</sup> College of Science, National University of Defense Technology, Changsha 410073, China

presence of correlated sources. In addition, it always requires the prior information about the number of sources, which is rarely practicable in the application.

Recently, thanks to the development of sparse signal recovery (Candes et al. 2006; Donoho 2006; Baraniuk 2007), many sparsity-based methods have been proposed to estimate the DOA (Malioutov et al. 2005; Hyder and Mahata 2010; Guo et al. 2010; Kim et al. 2012; Zhu and Chen 2015; Northardt et al. 2013; Du and Cheng 2014; Liu et al. 2012; Carlin et al. 2013), including  $\ell_1$  mixed-norm minimization approach (Malioutov et al. 2005; Hyder and Mahata 2010; Guo et al. 2010; Kim et al. 2012; Zhu and Chen 2015; Northardt et al. 2013; Du and Cheng 2014), and sparse bayesian learning (SBL) (Tipping (2001); Wipf et al. (2011)) technique. It has been demonstrated that most of the sparsity-based methods are adaptive to many scenarios, including closely spatial sources, a few snapshots, spatially correlated sources, and no requirement of source number (Malioutov et al. 2005). However, these methods mainly focus on the spatial sparsity but rarely take the temporal information of sources into account, whereas the temporal correlation exists universally in the practical applications. In this paper, we propose a new method called sparsity and temporal correlation exploiting (SaTC-E) based on the block sparse bayesian learning (BSBL) framework (Zhang and Rao 2011, 2013), which can exploit and self-adaptively learn the temporal correlation among the snapshots of each source to improve the performance of DOA estimation.

The original BSBL algorithm (Zhang and Rao 2011) has been successfully applied for multi-task learning (Wan et al. 2012) and wireless telemonitoring of fetal ECG (Zhang et al. 2013a,b). However, the performance of BSBL is very poor for DOA estimation, because the sensing matrix is determined by the specific array system and the discretization of bearing-space. BSBL has to discretize the bearing-space and assumes the sources lie on the bearing-grid, and yet the true sources are usually not located in this discretized grid. In particular, a large grid-interval leads to serious error of DOA estimation because of off-grid effect, while a small grid-interval results in strong column-coherence of the sensing matrix. To deal with this issue, we introduce a grid refined strategy in this paper. This strategy is made up of coarse and refined estimation stages. At the coarse stage, the grid-interval is relatively large and a local scope covering the DOA of sources is estimated by using a modified BSBL algorithm. Then at the refined stage, 1-D (one-dimensional) searching with fine grid-interval is implemented in the scope estimated previously to find the exact DOA. The processing technique at the refined stage keeps away from sparse signal recovery, so that the strong coherence of the refined sensing matrix does not destroy the performance of DOA estimation. Our criterion of grid refined strategy is to minimize the identical cost function derived from the BSBL methodology both at the coarse and refine stages (by using different mathematical tools), which is quite different from Malioutov et al. (2005), Northardt et al. (2013). In Liu et al. (2012), independent relevance vector machine-DOA (iRVM-DOA) algorithm employs SBL perspective and grid refined strategy as well, but it neglects the temporal information of sources.

The proposed method has many advantages, such as improved resolution, requirement of only a few snapshots, and no requirement of source number. Furthermore, it is adaptive to temporally and spatially correlated sources, as well as being applicable to any time-varying manifold scenario, provided that the manifold matrix is determinable, such as wideband sources and time-varying array.

The rest of this paper is organized as follows. In Sect. 2, we formulate the problem of DOA estimation and describe the main contributions of this paper briefly. In Sect. 3, we propose a new method called SaTC-E for narrowband DOA estimation, which can exploit the spatial sparsity and temporal correlation of sources. In Sect. 4, we generalize SaTC-E for

time-varying manifold scenario as long as the manifold matrix is determinable. In Sect. 5, real data tests and numerical simulations are presented to verify the performance of SaTC-E.

The notations in this paper are introduced as follows:  $\mathbf{I}_L$  denotes the identity matrix with size  $L \times L$ , and  $\mathbf{I}$  stands for the identity matrix, of which the dimension can be determined from the context.  $\mathbf{A}_{i \cdot}$  and  $\mathbf{A}_{\cdot j}$  are the  $i$ th row and the  $j$ th column of the matrix  $\mathbf{A}$ , respectively.  $\mathbf{A}_{ij}$  is the element of  $\mathbf{A}$  in the  $i$ th row and  $j$ th column.  $\mathbf{A} \otimes \mathbf{B}$  represents the Kronecker product of matrices  $\mathbf{A}$  and  $\mathbf{B}$ .  $\text{vec}(\mathbf{A})$  denotes the vectorization of  $\mathbf{A}$  by stacking its columns one by one into a column-vector.  $\mathbf{A}^T$  and  $\mathbf{A}^H$  are the transpose and the conjugate transpose of  $\mathbf{A}$ , respectively.  $\bar{\mathbf{a}}$  means the conjugate of the complex vector  $\mathbf{a}$ .  $\text{Toep}(\mathbf{a})$  denotes the  $M \times M$  Hermitian Toeplitz matrix determined by its first column  $\mathbf{a}$ .

## 2 Problem formulation

In this section, we formulate the problem of DOA estimation and present the main contributions of our work. The narrowband and stationary sources are taken into account.

Assume that  $K$  far-field narrowband sources impinge on an  $M$ -sensor array from directions  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]^T$ , and the inter-sensor spacing is half the wavelength of the incident signals. The array output at snapshot  $t$  is represented as

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}(t) + \mathbf{n}(t), \quad t = 1, 2, \dots, L, \tag{1}$$

where  $L$  denotes the number of snapshots,  $\mathbf{s}(t) = [s_1(t), \dots, s_K(t)]^T$  is the signal vector of sources, and  $\mathbf{n}(t)$  is the zero-mean additive complex white Gaussian noise with covariance matrix  $\mathbf{R}_{\mathbf{n}} = \sigma^2 \mathbf{I}$  and  $\sigma^2$  unknown. The noise is uncorrelated with sources. The array manifold matrix  $\mathbf{A}(\boldsymbol{\theta})$  is represented as,

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)], \tag{2}$$

where  $\mathbf{a}(\theta_k) = [e^{-j2\pi f_0 \tau_{1,k}}, \dots, e^{-j2\pi f_0 \tau_{M,k}}]^T$  is the  $k$ th steering vector, and  $f_0$  denotes the carrier frequency.  $\tau_{m,k}$  related to  $\theta_k$  is the propagational time-delay of the  $k$ th source between the reference and the  $m$ th sensor.

The purpose of DOA estimation is to estimate the directions  $\boldsymbol{\theta}$  using the given observation  $\{\mathbf{y}(t)\}_{t=1}^L$  and the formulation of matrix  $\mathbf{A}(\boldsymbol{\theta})$ . The main contributions of this paper are listed as follows:

- We exploit the spatial joint sparsity of all sources and the temporal correlation in each source simultaneously to estimate DOA without any parameter tuning and knowledge of source number.
- After coarse estimation, we implement a local 1-D searching at the refined stage to minimize the cost function derived consistently from BSBL methodology. This procedure reduces the error of DOA estimation caused by off-grid effect and strong coherence of the manifold matrix.
- We generalize the proposed method for any time-varying manifold scenario, provided that the array manifold matrix has a determinable formulation.

### 3 Narrowband DOA estimation based on spatial sparsity and temporal correlation

To cast the addressed DOA estimation as a problem of sparse signal recovery, we present the sparse representation of (1) firstly, and then propose an efficient method for DOA estimation. This method can capture the structure information of source signals and improve the accuracy of DOA estimation using a grid refined strategy.

#### 3.1 Sparse representation and temporal correlation

Since the sources are spatially sparse, the potential bearing-space can be discretized to form a bearing-grid  $\Theta = [\vartheta_1, \dots, \vartheta_N]^T (N \gg K)$ , which covers the true directions  $\theta$ . The array output (1) can be rewritten as a sparse linear combination of the steering vectors,

$$\mathbf{y}(t) = \Phi(\Theta) \mathbf{x}(t) + \mathbf{n}(t), \quad t = 1, 2, \dots, L, \tag{3}$$

where  $\mathbf{x}(t) \in \mathbb{C}^{N \times 1}$  is a zero-padded version of  $\mathbf{s}(t)$  from  $\theta$  to  $\Theta$ , that is, if  $\vartheta_i = \theta_k \in \theta$ , then  $x_i(t) = s_k(t) \neq 0$ ; otherwise,  $x_i(t) = 0$ . Therefore,  $\mathbf{x}(t)$  has only  $K (\ll N)$  nonzero elements.  $\Phi(\Theta) \in \mathbb{C}^{M \times N}$  is a corresponding extension of the manifold matrix  $\mathbf{A}(\theta)$ , that is

$$\Phi(\Theta) = [\mathbf{a}(\vartheta_1), \dots, \mathbf{a}(\vartheta_N)], \tag{4}$$

where  $\mathbf{a}(\vartheta_k) = [e^{-j2\pi f_0 \tilde{\tau}_{1,k}}, \dots, e^{-j2\pi f_0 \tilde{\tau}_{M,k}}]^T$  with  $\tilde{\tau}_{1,k}$  related to  $\vartheta_k$ . Since the cardinality of  $\Theta$  is usually much larger than the number of sensors, i.e.,  $N \gg M$ ,  $\Phi(\Theta)$  is an overcomplete dictionary.

Since the true directions are rarely located in the spatial grid  $\Theta$  exactly, (3) is not an accurate extension of (1). However, the processing of sparse signal recovery based on (3) will be always feasible with some recovery error, while how to reduce the error of DOA estimation caused by off-grid effect is another assignment we will study in the following section. Therefore, the model error in (3) is skipped over.

In the following we present the BSBL model (Zhang and Rao 2011) for DOA estimation. The array output (3) with  $L$  snapshots is represented as a multiple measurement vectors (MMV) model (Eldar and Mishali 2009)

$$\mathbf{Y} = \Phi \mathbf{X} + \mathbf{N}, \tag{5}$$

where  $\mathbf{Y} = [\mathbf{y}(t_1), \dots, \mathbf{y}(t_L)] \in \mathbb{C}^{M \times L}$ ,  $\mathbf{X} = [\mathbf{x}(t_1), \dots, \mathbf{x}(t_L)] \in \mathbb{C}^{N \times L}$ ,  $\Phi = \Phi(\Theta)$ , and  $\mathbf{N} = [\mathbf{n}(t_1), \dots, \mathbf{n}(t_L)] \in \mathbb{C}^{M \times L}$ . For stationary sources, the supports of all columns in  $\mathbf{X}$  are identical, so that  $\mathbf{X}$  has only  $K$  nonzero rows and each row  $\mathbf{X}_i$  denotes a potential source from direction  $\vartheta_i$ . As the sources are stochastic and temporally correlated, that is, the source signal at the time  $t$  is related to the signal at the last time  $t - 1$ , we assume that each row  $\mathbf{X}_i$  has a parameterized Gaussian distribution

$$p(\mathbf{X}_i; \gamma_i, \mathbf{B}_i) \sim \mathcal{CN}(\mathbf{0}, \gamma_i \mathbf{B}_i), \quad i = 1, 2, \dots, N \tag{6}$$

with parameters  $\gamma_i$  and  $\mathbf{B}_i$  to be estimated. Obviously,  $\gamma_i$  is a parameter controlling the row-support of  $\mathbf{X}$ , that is, if  $\gamma_i = 0$ , then  $\mathbf{X}_i$  is a zero vector and there is no source from direction  $\vartheta_i$ ; otherwise,  $\mathbf{X}_i \neq \mathbf{0}$  and there is a source from  $\vartheta_i$ .  $\mathbf{B}_i \in \mathbb{R}^{L \times L}$  is a Hermitian matrix that captures the correlation among the elements in  $\mathbf{X}_i$ , and can be interpreted as a covariance matrix of  $\mathbf{X}_i$ , up to a constant factor  $\gamma_i$ . Obviously, the temporal correlation of source signals is characterized by matrix  $\mathbf{B}_i$ . The estimate of  $\{\gamma_i\}$  is what we need for DOA estimation, that

is, by finding the peaks of  $\{\gamma_i\}$  we can get the DOA estimation according to the indices of those peaks.

For convenience of computing the temporal correlation of sources, (5) is transformed to a single measurement vector (SMV) model as follows,

$$\mathbf{z} = \Psi \mathbf{w} + \boldsymbol{\varepsilon}, \tag{7}$$

where  $\mathbf{z} = \text{vec}(\mathbf{Y}^T) \in \mathbb{C}^{ML \times 1}$ ,  $\Psi = \Phi \otimes \mathbf{I}_L \in \mathbb{C}^{ML \times NL}$ ,  $\mathbf{w} = \text{vec}(\mathbf{X}^T) \in \mathbb{C}^{NL \times 1}$ , and  $\boldsymbol{\varepsilon} = \text{vec}(\mathbf{N}^T) \in \mathbb{C}^{ML \times 1}$ . Consequently,  $\mathbf{w} = [\mathbf{w}_1^T, \dots, \mathbf{w}_N^T]^T$  is a block-sparse vector with  $\mathbf{w}_i = \mathbf{X}_i^T$ .

Under the assumption that the sources are spatially uncorrelated, the prior distribution of  $\mathbf{w}$  in (7) is also a Gaussian distribution,

$$p(\mathbf{w}; \{\gamma_i, \mathbf{B}_i\}) \sim \mathcal{CN}(0, \Sigma_0) \tag{8}$$

with the covariance matrix  $\Sigma_0 = \text{diag}[\gamma_1 \mathbf{B}_1, \dots, \gamma_N \mathbf{B}_N]$ .

Assume that the noises are mutually independent and  $\mathbf{N}_{ij} \sim \mathcal{CN}(0, \sigma^2)$ , then the conditional distribution of  $\mathbf{z}$  in (7) is

$$p(\mathbf{z} | \mathbf{w}; \sigma^2) \sim \mathcal{CN}(\Psi \mathbf{w}, \sigma^2 \mathbf{I}). \tag{9}$$

According to the Bayes rule, the posterior PDF of  $\mathbf{w}$  is expressed as

$$\begin{aligned} p(\mathbf{w} | \mathbf{z}; \{\gamma_i, \mathbf{B}_i\}, \sigma^2) \\ = (2\pi)^{-\frac{NL}{2}} |\Sigma_{\mathbf{w}}|^{-\frac{1}{2}} e^{\left\{-\frac{1}{2}(\mathbf{w}-\mu_{\mathbf{w}})^H \Sigma_{\mathbf{w}}^{-1}(\mathbf{w}-\mu_{\mathbf{w}})\right\}}, \end{aligned} \tag{10}$$

then obviously,  $p(\mathbf{w} | \mathbf{z}; \{\gamma_i, \mathbf{B}_i\}, \sigma^2) \sim \mathcal{CN}(\mu_{\mathbf{w}}, \Sigma_{\mathbf{w}})$  with

$$\begin{aligned} \Sigma_{\mathbf{w}} &= \left(\sigma^{-2} \Psi^H \Psi + \Sigma_0^{-1}\right)^{-1}, \\ \mu_{\mathbf{w}} &= \Sigma_0 \Psi^H \left(\sigma^2 \mathbf{I} + \Psi \Sigma_0 \Psi^H\right)^{-1} \mathbf{z}. \end{aligned} \tag{11}$$

As we mentioned above, the final task remains on the estimate of parameters  $\{\gamma_i\}$ . We achieve the estimate of  $\{\gamma_i\}$  by maximizing  $p(\mathbf{z}; \{\gamma_i, \mathbf{B}_i\}, \sigma^2)$ , and find the remarkable peaks in the estimate  $\hat{\boldsymbol{\gamma}}$  of  $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_N]^T$  which indicates the DOA of the incident sources. In Zhang and Rao (2011), the authors employed expectation–maximization (EM) technique (Dempster et al. 1977) to learn  $\{\gamma_i, \mathbf{B}_i\}$  and  $\sigma^2$ . Before estimating these parameters, they supposed  $\mathbf{B}_i = \mathbf{B}(i = 1, 2, \dots, N)$  to avoid the overfitting problem. The learning rules for  $\gamma_i, \mathbf{B}$  and  $\sigma^2$  are presented as (Zhang and Rao 2011)

$$\begin{aligned}
 \mathbf{\Omega}^{(k+1)} &= \left( \left( \mathbf{\Gamma}^{(k)} \right)^{-1} + \left[ \left( \sigma^2 \right)^{(k)} \right]^{-1} \mathbf{\Phi}^H \mathbf{\Phi} \right)^{-1}, \\
 \mathbf{X}^{(k+1)} &= \mathbf{\Gamma}^{(k)} \mathbf{\Phi}^H \left( \left( \sigma^2 \right)^{(k)} \mathbf{I} + \mathbf{\Phi} \mathbf{\Gamma}^{(k)} \mathbf{\Phi}^H \right)^{-1} \mathbf{Y}, \\
 \gamma_i^{(k+1)} &= \frac{1}{L} \left( \mathbf{X}^{(k+1)} \right)_i \left( \mathbf{B}^{(k)} \right)^{-1} \left( \mathbf{X}^{(k+1)} \right)_i^H + \left( \mathbf{\Omega}^{(k+1)} \right)_{ii}, \\
 \mathbf{B}^{(k+1)} &= \frac{1}{N} \sum_{i=1}^N \frac{\left( \boldsymbol{\Sigma}_{\mathbf{w}}^{(k)} \right)^{ii} + \left( \boldsymbol{\mu}_{\mathbf{w}}^{(k)} \right)^i \left( \left( \boldsymbol{\mu}_{\mathbf{w}}^{(k)} \right)^i \right)^H}{\gamma_i^{(k+1)}}, \\
 \left( \sigma^2 \right)^{(k+1)} &= \frac{\left\| \mathbf{z} - \boldsymbol{\Psi} \boldsymbol{\mu}_{\mathbf{w}}^{(k)} \right\|_2^2 + \left( \sigma^2 \right)^{(k)} \left[ NL - \text{tr} \left( \boldsymbol{\Sigma}_{\mathbf{w}}^{(k)} \boldsymbol{\Sigma}_0^{-1} \right) \right]}{ML},
 \end{aligned} \tag{12}$$

where  $\mathbf{\Gamma} = \text{diag} (\gamma_1, \dots, \gamma_N)$ ,  $\left( \boldsymbol{\Sigma}_{\mathbf{w}}^{(k)} \right)^{ii}$  denotes the  $i$ th principal diagonal block with size  $L \times L$  in the previous estimate of  $\boldsymbol{\Sigma}_{\mathbf{w}}$ , and  $\left( \boldsymbol{\mu}_{\mathbf{w}}^{(k)} \right)^i$  is the  $i$ th block-subvector with size  $L \times 1$  in the previous estimate of  $\boldsymbol{\mu}_{\mathbf{w}}$ . The iteration procedure for updating  $\gamma_i$ ,  $\mathbf{B}$  and  $\sigma^2$  will continue until a predefined criterion is satisfied. For example, if the  $k + 1$ th iteration satisfies  $\left\| \mathbf{y}^{(k+1)} - \mathbf{y}^{(k)} \right\|_2 / \left\| \mathbf{y}^{(k)} \right\|_2 \leq \delta$  where  $\delta$  is a small enough positive threshold, the iteration is over. The details of the derivation for the above learning rules refer to [Zhang and Rao \(2011\)](#).

We observe that the quadratic Mahalanobis distance (MD) measure of  $\mathbf{X}_i, \mathbf{X}_i \mathbf{B}^{-1} \mathbf{X}_i^H$ , is applied for the learning rule of  $\gamma_i$ , while the Euler distance (ED) measure is used for parameter estimation in iRVM-DOA algorithm ([Liu et al. 2012](#)). The MD measure can capture the correlations among the elements of  $\mathbf{X}_i$ , while the ED measure cannot. Theorem 1 in [Zhang and Rao \(2011\)](#) has verified that regardless of the choice of  $\mathbf{B}_i$ , the global minimum of the cost function  $L (\{\gamma_i\}, \mathbf{B}_i, \sigma)$  (see (15)) always leads to the true sparse solution in noiseless case. Supported by this fact, assuming  $\mathbf{B}_i = \mathbf{B}, i = 1, 2, \dots, N$  is reasonable. In the rest of this paper,  $\mathbf{B}$  is parameterized as ([Zhang and Rao 2013](#))

$$\mathbf{B} = \text{Toep} \left( \left[ 1, b, b^2, \dots, b^{L-1} \right] \right). \tag{13}$$

As we know, the accuracy of  $\hat{\boldsymbol{\gamma}}$  is the core of DOA estimation. However, the accuracy of noise variance estimation  $\hat{\sigma}^2$  will affect the accuracy of  $\hat{\boldsymbol{\gamma}}$  significantly. Unfortunately, the estimate method of  $\sigma^2$  based on learning rule (12) is not effective enough. Instead of (12), we use the specific maximum likelihood estimator (MLE) of  $\sigma^2$  given as ([Stoica and Nehorai 1989](#))

$$\hat{\sigma}^2 = \frac{1}{ML} \text{tr} \left[ \mathbf{Y}^H \left( \mathbf{I} - \hat{\mathbf{A}} \left( \hat{\mathbf{A}}^H \hat{\mathbf{A}} \right)^{-1} \hat{\mathbf{A}} \right) \mathbf{Y} \right], \tag{14}$$

where  $\hat{\mathbf{A}} = \left[ \mathbf{a} \left( \hat{\vartheta}_{k_1} \right) \mathbf{a} \left( \hat{\vartheta}_{k_2} \right) \dots \mathbf{a} \left( \hat{\vartheta}_{k_h} \right) \right]$ , and  $\hat{\vartheta}_{k_i}$  is the estimate of DOA corresponding to the  $i$ th peak of  $\hat{\boldsymbol{\gamma}}$ .

### 3.2 Grid refined strategy associated with cost function

In this subsection, a grid refined strategy is presented to mitigate the error of DOA estimation caused by off-grid effect. This strategy includes the coarse and refined estimation stages. At the coarse stage with relatively large grid-interval, we learn  $\hat{\boldsymbol{\gamma}}$  by EM iteration (12), followed by finding the peaks of  $\hat{\boldsymbol{\gamma}}$ . Then at the refined stage with fine grid-interval, the

DOA is estimated via local 1-D searching around the bearing-scopes corresponding to the peak-indices of  $\hat{\boldsymbol{\gamma}}$  obtained at the coarse stage.

The EM technique at the coarse stage is essentially equivalent to minimizing the cost function (Zhang and Rao 2011)

$$L(\{\gamma_i\}, \mathbf{B}, \sigma) = \log |\boldsymbol{\Sigma}| + \mathbf{z}^H \boldsymbol{\Sigma}^{-1} \mathbf{z} \tag{15}$$

with  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I} + \boldsymbol{\Psi} \boldsymbol{\Sigma}_0 \boldsymbol{\Psi}^H$ . At the refined stage, we aim to minimize the cost function (15) consistently with a fine grid-interval as well. The procedure of the refined stage is derived as follows.

Let  $\boldsymbol{\Psi} = \boldsymbol{\Phi} \otimes \mathbf{I}_L$ ,  $\boldsymbol{\Phi} = [\mathbf{a}(\vartheta_1), \dots, \mathbf{a}(\vartheta_N)]$ , and  $\boldsymbol{\Sigma}_0 = \boldsymbol{\Gamma} \otimes \mathbf{B}$ . The matrix  $\boldsymbol{\Sigma}$  can be represented as

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I} + \sum_{i=1}^N \gamma_i \mathbf{a}(\vartheta_i) \mathbf{a}^H(\vartheta_i) \otimes \mathbf{B}. \tag{16}$$

In order to facilitate the DOA of the  $k$ th source,  $\boldsymbol{\Sigma}$  is approximated by

$$\boldsymbol{\Sigma} \approx \boldsymbol{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \otimes \mathbf{B}, \tag{17}$$

where  $\boldsymbol{\Sigma}_{-k} = \sigma^2 \mathbf{I} + \sum_{i=1, i \neq k, k'}^N \gamma_i \mathbf{a}(\vartheta_i) \mathbf{a}^H(\vartheta_i) \otimes \mathbf{B}$ .  $\gamma_i (i \neq k, k', k' = k - 1 \text{ or } k' = k + 1)$ ,

$\mathbf{B}$  and  $\sigma^2$  have been learned previously at the coarse stage.  $\tilde{\gamma}_k$  and  $\tilde{\vartheta}_k$  are the parameters to be estimated.

As we mentioned, we calculate  $\tilde{\gamma}_k$  and  $\tilde{\vartheta}_k$  by minimizing the cost function  $L(\{\gamma_i\}, \mathbf{B}, \sigma)$  approximately, that is,

$$\left\{ \hat{\tilde{\gamma}}_k, \hat{\tilde{\vartheta}}_k \right\} = \arg \min_{\tilde{\gamma}_k, \tilde{\vartheta}_k \in \Theta_k} \left\{ L(\tilde{\gamma}_k, \tilde{\vartheta}_k) \right\}, \tag{18}$$

where

$$L(\tilde{\gamma}_k, \tilde{\vartheta}_k) = \log \left| \boldsymbol{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \otimes \mathbf{B} \right| + \mathbf{z}^H \left( \boldsymbol{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \otimes \mathbf{B} \right)^{-1} \mathbf{z}, \tag{19}$$

and  $\Theta_k$  denotes the searching scope for the  $k$ th source. In this paper, we set  $\Theta_k = [\underline{\vartheta}_k, \hat{\vartheta}_k]$  or  $\Theta_k = [\hat{\vartheta}_k, \bar{\vartheta}_k]$ , where  $\underline{\vartheta}_k$  and  $\bar{\vartheta}_k$  are the adjacent directions of DOA estimation  $\hat{\vartheta}_k$  obtained previously according to the peaks of  $\hat{\boldsymbol{\gamma}}$  learned at the coarse stage. We repeat (18) for all  $\hat{\vartheta}_k, k = 1, 2, \dots, K$ .

To solve (18), the alternating minimization strategy (AMS) on  $\tilde{\gamma}_k, \tilde{\vartheta}_k$  is utilized. Firstly, the partial derivatives of  $L(\tilde{\gamma}_k, \tilde{\vartheta}_k)$  with respect to (w.r.t.)  $\tilde{\gamma}_k$  and  $\tilde{\vartheta}_k$  are calculated as

$$\begin{aligned} \frac{\partial L(\tilde{\gamma}_k, \tilde{\vartheta}_k)}{\partial \tilde{\gamma}_k} &= \sum_{i=1}^{r_B} \left[ \mathbf{a}(\tilde{\vartheta}_k) \otimes \mathbf{u}_i^B \right]^H \mathbf{V} \left[ \mathbf{a}(\tilde{\vartheta}_k) \otimes \mathbf{u}_i^B \right], \\ \frac{\partial L(\tilde{\gamma}_k, \tilde{\vartheta}_k)}{\partial \tilde{\vartheta}_k} &= \tilde{\gamma}_k \text{tr} \left\{ \mathbf{V} \left[ \mathbf{a}(\tilde{\vartheta}_k) \mathbf{d}^H(\tilde{\vartheta}_k) \otimes \mathbf{B} \right] \right\} \\ &\quad + \tilde{\gamma}_k \text{tr} \left\{ \mathbf{V} \left[ \mathbf{d}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \otimes \mathbf{B} \right] \right\}, \end{aligned} \tag{20}$$

---



---

**Algorithm 1: Refined DOA Estimation**

---

**Step 1:** Give desired precision  $\Delta\vartheta_r$ ; set  $k = 1$ .

**Step 2:** Give initial values  $\tilde{\gamma}_k^{(0)}$ ,  $\tilde{\vartheta}_k^{(0)}$ , and  $\Theta_k$  according to the coarse estimation.

**Step 3:** Compute  $\tilde{\gamma}_k^{(i+1)}$  by solving

$$\sum_{i=1}^{r_{\mathbf{B}}} \left[ \mathbf{a} \left( \tilde{\vartheta}_k^{(i)} \right) \otimes \mathbf{u}_i^{\mathbf{B}} \right]^H \mathbf{V}_i \left[ \mathbf{a} \left( \tilde{\vartheta}_k^{(i)} \right) \otimes \mathbf{u}_i^{\mathbf{B}} \right] = 0$$

via the nonlinear least squares method (NonLSM), where

$$\mathbf{V}_i = \mathbf{W}_i^{-1} - \mathbf{W}_i^{-1} \mathbf{z} \mathbf{z}^H \mathbf{W}_i^{-1},$$

$$\mathbf{W}_i = \Sigma_{-k} + \tilde{\gamma}_k^{(i)} \mathbf{a} \left( \tilde{\vartheta}_k^{(i)} \right) \mathbf{a}^H \left( \tilde{\vartheta}_k^{(i)} \right) \otimes \mathbf{B}.$$

**Step 4:** Compute  $\tilde{\vartheta}_k^{(i+1)}$  by local searching in  $\Theta_k$  with resolution  $\Delta\vartheta_r$ , that is,

$$\tilde{\vartheta}_k^{(i+1)} = \arg \min_{\tilde{\vartheta}_k \in \Theta_k} \left| \frac{\partial L \left( \tilde{\gamma}_k^{(i+1)}, \tilde{\vartheta}_k \right)}{\partial \tilde{\vartheta}_k} \right|.$$

**Step 5:** If  $\tilde{\vartheta}_k^{(i+1)} = \tilde{\vartheta}_k^{(i)}$ , we get  $\hat{\vartheta}_k = \tilde{\vartheta}_k^{(i+1)}$ ,  $k = k + 1$ , go to **Step 6**;  
otherwise,  $i = i + 1$ , return to **Step 3**.

**Step 6:** If  $k < K$ , return to **Step 2**; otherwise, the algorithm ends.

---



---

where  $\mathbf{V} = \mathbf{W}^{-1} - \mathbf{W}^{-1} \mathbf{z} \mathbf{z}^H \mathbf{W}^{-1}$ ,  $\mathbf{W} = \Sigma_{-k} + \tilde{\gamma}_k \mathbf{a} \left( \tilde{\vartheta}_k \right) \mathbf{a}^H \left( \tilde{\vartheta}_k \right) \otimes \mathbf{B}$ ,  $\mathbf{B} = \sum_{i=1}^{r_{\mathbf{B}}} \mathbf{u}_i^{\mathbf{B}} \left( \mathbf{u}_i^{\mathbf{B}} \right)^H$ , and  $\mathbf{u}_i^{\mathbf{B}}$  is the product of the  $i$ th singular value and the corresponding eigenvector of  $\mathbf{B}$ .  $r_{\mathbf{B}} = \text{rank} \left( \mathbf{B} \right)$ , and  $\mathbf{d} \left( \tilde{\vartheta}_k \right) = \partial \mathbf{a} \left( \tilde{\vartheta}_k \right) / \partial \tilde{\vartheta}_k$ . The detail for these derivations is presented in the Appendix 1. Based on (20), the procedure for the refined DOAs estimation is shown as in Algorithm 1.

It is obvious that if the spatial grid-interval  $\Delta\vartheta_c$  at the coarse stage is small enough, e.g.,  $\Delta\vartheta_c = 1^\circ$ , then the scope we search is only  $K \Delta\vartheta_c$  and much smaller than the whole space  $[-90^\circ, 90^\circ]$ . Therefore, when the proposed method can mitigate the error of DOA estimation resulted by off-grid effect, it reduces the computational cost of the original BSBL algorithm as well. Meanwhile, since the refined estimation keeps away from the processing of sparse signal recovery, its performance has not been destroyed by the strong coherence among the refined steering vectors.

*Remark 1* The iRVM-DOA algorithm in Liu et al. (2012) also employs the ideal of SBL (but not BSBL) and grid-refinement to estimate DOA and has an excellent performance. However, there exists a fact that if we set  $\mathbf{B} = \mathbf{I}_L$ , the derived algorithm in this work is exactly equivalent to the iRVM-DOA.

**Proposition 1** With  $\sigma^2$  fixed and  $\mathbf{B} = \mathbf{I}_L$ , the updating formula of  $\gamma_i$  at the coarse stage is simplified into

$$\gamma_i^{(k+1)} = \frac{1}{L} \left( \mathbf{X}^{(k+1)} \right)_i \cdot \left( \mathbf{X}^{(k+1)} \right)_i^H + \left( \boldsymbol{\Omega}^{(k+1)} \right)_{ii}, \tag{21}$$

which is exactly identical to the one used in the iRVM-DOA.

Besides the equivalence at the coarse stage, the equivalence at the refined stage should also be proved. In fact, we have



**Proposition 2** When  $\mathbf{B} = \mathbf{I}_L$ , the objective function at the refined stage (19) is equal to

$$\begin{aligned} \check{L}(\tilde{\gamma}_k, \tilde{\vartheta}_k) &= L \log \left| \check{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right| \\ &\quad + \text{tr} \left[ \mathbf{Y}^H \left( \check{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right)^{-1} \mathbf{Y} \right], \end{aligned} \tag{22}$$

with  $\check{\Sigma}_{-k} = \sigma^2 \mathbf{I}_M + \sum_{i=1, i \neq k, k'}^N \gamma_i \mathbf{a}(\vartheta_i) \mathbf{a}^H(\vartheta_i)$ , which is the objective function at the refined stage in the iRVM-DOA.

The proof of Proposition 2 is given in the Appendix 2.

These results demonstrate that the proposed method generalizes the iRVM-DOA algorithm, and the latter cannot exploit the temporal correlation in each source while the former can.

### 4 Extension to time-varying manifold

In this section, we extend the proposed method to time-varying manifold scenario, such as wideband sources and time-varying array. In time-varying manifold scenario, the array manifold matrix is associated with time-variable so that the existing sparsity-based methods are no longer in force for DOA estimation. Fortunately, the array output can be represented as a block-sparsity model like (7), provided that the manifold matrix at each snapshot is determinable. Therefore, the proposed method is applicable to time-varying manifold scenario. In this section, wideband sources are considered.

Consider the sources are linear modulated frequency (LFM) signals,

$$s_k(t) = u_k(t) e^{j2\pi\left(f_0 t + \frac{1}{2} K_c t^2\right)}, \quad k = 1, 2, \dots, K, \tag{23}$$

where  $u_k(t)$  is the complex amplitude,  $f_0$  is the initial frequency, and  $K_c$  is the chip rate. Then the output of the  $m$ th sensor is represented as

$$y_m(t) = \sum_{k=1}^K s_k(t) e^{j2\pi\left(\frac{1}{2} K_c \tau_{mk}^2 - t K_c \tau_{mk} - f_0 \tau_{mk}\right)} + n_m(t), \quad t = 1, 2, \dots, L, \tag{24}$$

with noise  $n_m(t)$ .

Based on the spatial discretization described in Sect. 3.1, the array output takes a matrix-vector form as

$$\mathbf{y}(t) = \mathbf{A}(\Theta, t) \mathbf{s}(t) + \mathbf{n}(t), \quad t = 1, 2, \dots, L, \tag{25}$$

where  $\mathbf{y}(t) = [y_1(t), \dots, y_M(t)]^T$  is the array observation at snapshot  $t$ ,  $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$  is the source signal-vector, and  $\mathbf{n}(t) = [n_1(t), \dots, n_M(t)]^T$  denotes the receiving noise.  $\mathbf{A}(\Theta, t)$  is the array manifold matrix at snapshot  $t$ , which is denoted as

$$\mathbf{A}(\Theta, t) = [\mathbf{a}(\vartheta_1, t), \dots, \mathbf{a}(\vartheta_N, t)], \tag{26}$$

with bearing-grid  $\Theta = [\vartheta_1, \dots, \vartheta_N]^T$  and steering vector

$$\mathbf{a}(\vartheta_n, t) = \left[ 1, \dots, e^{-j2\pi(f_0 + K_c t)\tau_{Mn} + j\pi K_c \tau_{Mn}^2} \right]^T. \tag{27}$$

Note that  $\mathbf{A}(\Theta, t)$  changes as snapshot  $t$ , so (25) is a time-varying manifold model.

In order to exploit the advantages of the spatial joint sparsity and the temporal correlation, we reformulate (25) as a block-sparsity model in the following.

Let  $\mathbf{w}$  denote a stack of vector-blocks  $\mathbf{s}_n(\mathbf{t}) = [s_n(1), \dots, s_n(L)]^T$  ( $n = 1, 2, \dots, N$ ) as

$$\mathbf{w} = \left[ \underbrace{s_1(1), \dots, s_1(L)}_{\mathbf{s}_1^T(\mathbf{t})}, \dots, \underbrace{s_N(1), \dots, s_N(L)}_{\mathbf{s}_N^T(\mathbf{t})} \right]^T, \tag{28}$$

Let  $\mathbf{D}_{[ij]} = \text{diag}[a_{ij}(1), \dots, a_{ij}(L)]$ ,  $i = 1, 2, \dots, M$ ,  $j = 1, 2, \dots, N$ , where  $a_{ij}(t)$  stands for the element lying at the  $i$ th row and the  $j$ th column of  $\mathbf{A}(\Theta, t)$ , then an overcomplete dictionary is constructed as

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_{[11]} & \mathbf{D}_{[12]} & \cdots & \mathbf{D}_{[1N]} \\ \mathbf{D}_{[21]} & \mathbf{D}_{[22]} & \cdots & \mathbf{D}_{[2N]} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}_{[M1]} & \mathbf{D}_{[M2]} & \cdots & \mathbf{D}_{[MN]} \end{bmatrix} \in \mathbb{C}^{ML \times NL} \tag{29}$$

Now the array output (25) with multiple snapshots can be reformulated as a block-sparsity model

$$\mathbf{z} = \mathbf{D}\mathbf{w} + \boldsymbol{\varepsilon}, \tag{30}$$

where  $\mathbf{z} = \text{vec}(\mathbf{Y}^T)$  with  $\mathbf{Y} = [\mathbf{y}(t_1) \ \dots \ \mathbf{y}(t_L)]$ , and  $\boldsymbol{\varepsilon} = \text{vec}(\mathbf{N}^T)$  with  $\mathbf{N} = [\mathbf{n}(t_1) \ \dots \ \mathbf{n}(t_L)]$ .

Under the assumption that each source  $\mathbf{s}_n(\mathbf{t})$  is Gaussian distributed, i.e.,  $p(\mathbf{s}_n(\mathbf{t}); \gamma_i, \mathbf{B}) \sim \mathcal{CN}(\mathbf{0}, \gamma_i \mathbf{B})$ ,  $i = 1, 2, \dots, N$ , we have the learning rules (Zhang and Rao 2011) by EM technique

$$\begin{aligned} \gamma_i^{(k+1)} &= \frac{\text{tr} \left[ \left( \mathbf{B}^{(k)} \right)^{-1} \left( \left( \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}^{(k)} \right)^{ii} + \left( \tilde{\boldsymbol{\mu}}_{\mathbf{w}}^{(k)} \right)^i \left( \tilde{\boldsymbol{\mu}}_{\mathbf{w}}^{(k)} \right)^H \right) \right]}{L}, \\ \mathbf{B}^{(k+1)} &= \frac{1}{N} \sum_{i=1}^N \frac{\left( \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}^{(k)} \right)^{ii} + \left( \tilde{\boldsymbol{\mu}}_{\mathbf{w}}^{(k)} \right)^i \left( \tilde{\boldsymbol{\mu}}_{\mathbf{w}}^{(k)} \right)^H}{\gamma_i^{(k+1)}}, \\ \left( \sigma^2 \right)^{(k+1)} &= \frac{\left\| \mathbf{z} - \mathbf{D} \tilde{\boldsymbol{\mu}}_{\mathbf{w}}^{(k)} \right\|_2^2 + \left( \sigma^2 \right)^{(k)} \left[ NL - \text{tr} \left( \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}}^{(k)} \boldsymbol{\Sigma}_0^{-1} \right) \right]}{ML}, \end{aligned} \tag{31}$$

with all the notations similar to the mentioned in Sect. 3. Matrix  $\Psi$  in Sect. 3 is replaced by  $\mathbf{D}$  here, so that the covariance and mean in (11) become

$$\begin{aligned} \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}} &= \left( \sigma^{-2} \mathbf{D}^H \mathbf{D} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1}, \\ \tilde{\boldsymbol{\mu}}_{\mathbf{w}} &= \boldsymbol{\Sigma}_0 \mathbf{D}^H \left( \sigma^2 \mathbf{I} + \mathbf{D} \boldsymbol{\Sigma}_0 \mathbf{D}^H \right)^{-1} \mathbf{z}. \end{aligned} \tag{32}$$

*Remark 2* By setting  $\mathbf{A}(\Theta, t) = \Phi(\Theta)$ ,  $t = 1, 2, \dots, L$ , it is easy to verify that the learning rules in time-varying manifold scenario are the generalization of the counterpart for narrowband scenario, while the latter ones are only used for the fixed manifold model.

Similarly to narrowband scenario, by following the coarse estimation via EM iteration (31), we turn into the refined estimation stage. Based on the cost function

$$\begin{aligned} \tilde{L}(\{\gamma_i\}, \mathbf{B}, \sigma) &= \log \left| \sigma^2 \mathbf{I} + \mathbf{D} \Sigma_0 \mathbf{D}^H \right| \\ &\quad + \mathbf{z}^H \left( \sigma^2 \mathbf{I} + \mathbf{D} \Sigma_0 \mathbf{D}^H \right)^{-1} \mathbf{z} \end{aligned} \tag{33}$$

and the approximation

$$\tilde{\Sigma} = \sigma^2 \mathbf{I} + \mathbf{D} \Sigma_0 \mathbf{D}^H \approx \tilde{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{D} \left[ \tilde{\vartheta}_k \right] \mathbf{B} \mathbf{D} \left[ \tilde{\vartheta}_k \right]^H \tag{34}$$

we have the cost function reformulated as

$$\begin{aligned} \tilde{L}(\tilde{\gamma}_k, \tilde{\vartheta}_k) &= \log \left| \tilde{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{D} \left[ \tilde{\vartheta}_k \right] \mathbf{B} \mathbf{D} \left[ \tilde{\vartheta}_k \right]^H \right| \\ &\quad + \mathbf{z}^H \left( \tilde{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{D} \left[ \tilde{\vartheta}_k \right] \mathbf{B} \mathbf{D} \left[ \tilde{\vartheta}_k \right]^H \right)^{-1} \mathbf{z}. \end{aligned} \tag{35}$$

where  $\tilde{\Sigma}_{-k} = \sigma^2 \mathbf{I} + \sum_{n=1, n \neq k, k'}^N \gamma_n \mathbf{D} [\vartheta_n] \mathbf{B} \mathbf{D} [\vartheta_n]^H$  and

$$\mathbf{D} [\vartheta_n] = \begin{bmatrix} \mathbf{D}_{[1n]} \\ \mathbf{D}_{[2n]} \\ \vdots \\ \mathbf{D}_{[Mn]} \end{bmatrix}, n = 1, 2, \dots, N. \tag{36}$$

The partial derivatives of  $\tilde{L}(\tilde{\gamma}_k, \tilde{\vartheta}_k)$  w.r.t.  $\tilde{\gamma}_k$  and  $\tilde{\vartheta}_k$  are calculated as

$$\begin{aligned} \frac{\partial \tilde{L}(\tilde{\gamma}_k, \tilde{\vartheta}_k)}{\partial \tilde{\gamma}_k} &= \sum_{i=1}^{r_B} \left( \mathbf{D} \left[ \tilde{\vartheta}_k \right] \mathbf{u}_i^B \right)^H \tilde{\mathbf{V}} \left( \mathbf{D} \left[ \tilde{\vartheta}_k \right] \mathbf{u}_i^B \right), \\ \frac{\partial \tilde{L}(\tilde{\gamma}_k, \tilde{\vartheta}_k)}{\partial \tilde{\vartheta}_k} &= \tilde{\gamma}_k \text{tr} \left\{ \tilde{\mathbf{V}} \mathbf{D} \left[ \tilde{\vartheta}_k \right] \mathbf{B} \left( \mathbf{F} \left[ \tilde{\vartheta}_k \right] \right)^H \right\} \\ &\quad + \tilde{\gamma}_k \text{tr} \left\{ \tilde{\mathbf{V}} \mathbf{F} \left[ \tilde{\vartheta}_k \right] \mathbf{B} \left( \mathbf{D} \left[ \tilde{\vartheta}_k \right] \right)^H \right\}, \end{aligned} \tag{37}$$

where  $\tilde{\mathbf{V}} = \tilde{\mathbf{W}}^{-1} - \tilde{\mathbf{W}}^{-1} \mathbf{z} \mathbf{z}^H \tilde{\mathbf{W}}^{-1}$ ,  $\tilde{\mathbf{W}} = \tilde{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{D} \left[ \tilde{\vartheta}_k \right] \mathbf{B} \mathbf{D} \left[ \tilde{\vartheta}_k \right]^H$  and  $\mathbf{F} \left[ \tilde{\vartheta}_k \right] = \partial \mathbf{D} \left[ \tilde{\vartheta}_k \right] / \partial \tilde{\vartheta}_k$ . The procedure for the refined estimation in time-varying manifold scenario is similar to Algorithm 1 for narrowband scenario, so we do not show it again.

*Remark 3* Similarly as the discussion in Sect. 3, in time-varying manifold scenario we replace the learning rule of noise variance  $\sigma^2$  with the MLE

$$\hat{\sigma}^2 = \frac{1}{M} \left[ \mathbf{z}^H \left( \mathbf{I} - \hat{\mathbf{D}} \left( \hat{\mathbf{D}}^H \hat{\mathbf{D}} \right)^{-1} \hat{\mathbf{D}} \right) \mathbf{z} \right], \tag{38}$$

where  $\hat{\mathbf{D}} = \left[ \mathbf{D} \left[ \hat{\vartheta}_{k_1} \right] \mathbf{D} \left[ \hat{\vartheta}_{k_2} \right] \dots \mathbf{D} \left[ \hat{\vartheta}_{k_h} \right] \right]$ ,  $\hat{\vartheta}_{k_i}$  is the DOA estimation corresponding to the  $i$ th peak of  $\hat{\boldsymbol{\gamma}}$ , and  $\mathbf{D} \left[ \hat{\vartheta}_{k_1} \right]$  has an identical form as  $\mathbf{D} [\vartheta_n]$ .

*Remark 4* While the learning rules (31) at the coarse stage are proved to be the extension of that in narrowband scenario, the generalization at the refined stage is obvious in terms of Proposition 3.

**Proposition 3** *Let  $\mathbf{A}(\Theta, t) = \Phi(\Theta)$ , we have that the cost function (35) in time-varying scenario equals to the cost function (19) in narrowband scenario.*

In summary, we propose a unified method, which utilizes the spatial sparsity and exploits the temporal correlation of sources simultaneously, to accomplish the DOA estimation for any time-varying manifold scenario in this section. This method is called as SaTC-E, which means exploiting the sparsity and temporal correlation simultaneously.

## 5 Experiment results

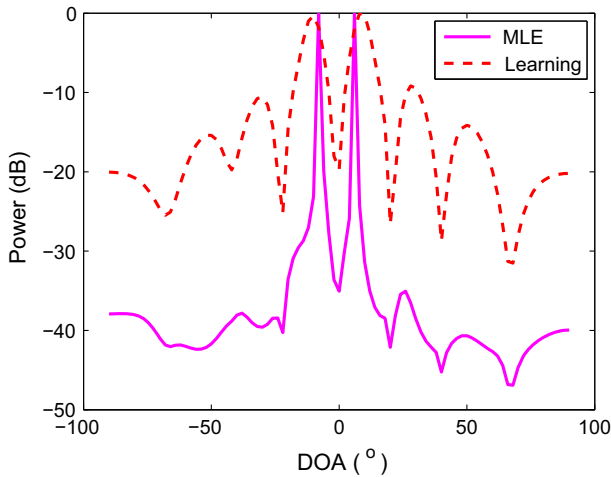
In this section, several experiments are carried out to compare SaTC-E with some classical algorithms, including MUSIC Schmidt (1986), L1-SVD (Malioutov et al. 2005), iRVM-DOA (Liu et al. 2012) for narrowband scenario, and two-side correlation transformation (TCT; Valaee and Kabal 1995), coherent signal-subspace method (CSSM; Wang and Kaveh 1985), wideband iRVM-DOA (WiRVM-DOA; Liu et al. 2012) for wideband scenario. L1-SVD algorithm is taken as one of the representatives of the sparsity-based methods.

The setups are as follows. For narrowband scenario, the incident sources are two complex sinusoid signals with carrier frequency  $f_0 = 2.143$  GHz, while for wideband scenario the sources are two LFM signals. The center frequency, the bandwidth, and the chip rate of LFMs are  $f_c = 2.118$  GHz,  $B = 50$  MHz, and  $K_c = 10^3$ , respectively. The directions of the incident sources are  $-6.54^\circ$  and  $6.54^\circ$ . The uniform linear array (ULA) is considered. The number of sensors is  $M = 8$ , and the inter-sensor spacing is half the wavelength. The number of snapshots is  $L$ . The number of sources is assumed to be known for MUSIC, TCT, and CSSM. The uniform spatial grids are equipped, and the grid-intervals are set to  $\Delta\vartheta_c$  for SaTC-E, iRVM-DOA and WiRVM-DOA at the coarse stage and  $\Delta\vartheta_r$  for MUSIC, TCT and CSSM, respectively. The searching precision at the refined stage is also set to  $\Delta\vartheta_r$  for iRVM-DOA, WiRVM-DOA and SaTC-E. The stopping criterion of EM iteration for SaTC-E is  $\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|_2 / \|\mathbf{y}^{(k)}\|_2 \leq \delta$  with  $\delta = 10^{-4}$ , or that EM iteration achieves 3000 times. If no special mention, the covariance matrix for SaTC-E is calculated as  $\mathbf{B} = \text{Toep}([1, b, b^2, \dots, b^{L-1}])$  with  $b = -0.99$ . L1-SVD is implemented as in Malioutov et al. (2005).

### 5.1 Real data tests

In this subsection, we utilize the real data to present the spectra of SaTC-E and the compared methods and illustrate the performance of SaTC-E. The real data was collected by the antenna ULA in a microwave black-room. The sampling frequency was 500 Msp. Before DOA estimation, the array output has already been pre-treated with phase correction and amplitude correction.

Firstly, we compare the performance of SaTC-E using the learning rule of  $\sigma^2$  with that using the MLE of  $\sigma^2$ . The sources are wideband LFM signals. Ten snapshots are used. The spectra are shown in Fig. 1. The spectrum-peaks using the MLE of  $\sigma^2$  are much sharper than those using the learning rule of  $\sigma^2$ . The reason is that  $\hat{\sigma}^2$  learned from the cost function (33) is not accurate enough for the specific application of DOA estimation, while the MLE (38)



**Fig. 1** Spatial spectra of SaTC-E using the learning rule and the MLE of  $\sigma^2$ , respectively. The sources are two LFM signals. The number of snapshots is  $L = 10$

is derived from the spectral analysis of DOA estimation. According to this result, we employ MLE to estimate  $\sigma^2$  for SaTC-E in the following experiments.

For narrowband scenario, Fig. 2a plots the spatial spectrum of MUSIC and spectra of L1-SVD, iRVM-DOA and SaTC-E at the coarse stage. The spectra at the refined stage of SaTC-E and iRVM-DOA within the local scope are displayed in Fig. 2b, c. The number of snapshots is  $L = 20$ , the grid-intervals are  $\Delta\vartheta_c = 1^\circ$  and  $\Delta\vartheta_r = 0.01^\circ$ . The peaks of SaTC-E are much shaper than those of MUSIC and L1-SVD, and thus SaTC-E surpasses MUSIC and L1-SVD. Meanwhile, SaTC-E has similar performance to iRVM-DOA. The counterparts for wideband scenario are shown in Fig. 3. The grid-intervals are set to  $\Delta\vartheta_c = 2^\circ$  and  $\Delta\vartheta_r = 0.01^\circ$ . Although the number of snapshots used by TCT and CSSM ( $L = 320$ ) is much larger than that used by SaTC-E ( $L = 10$ ), the spectrum of SaTC-E is more remarkable than those of TCT and CSSM. On the other hand, the spectrum of WiRVM-DOA suffers several false peaks. The comparison results illustrate that SaTC-E is applicable to DOA estimation both for narrowband and wideband, even it uses only a few snapshots.

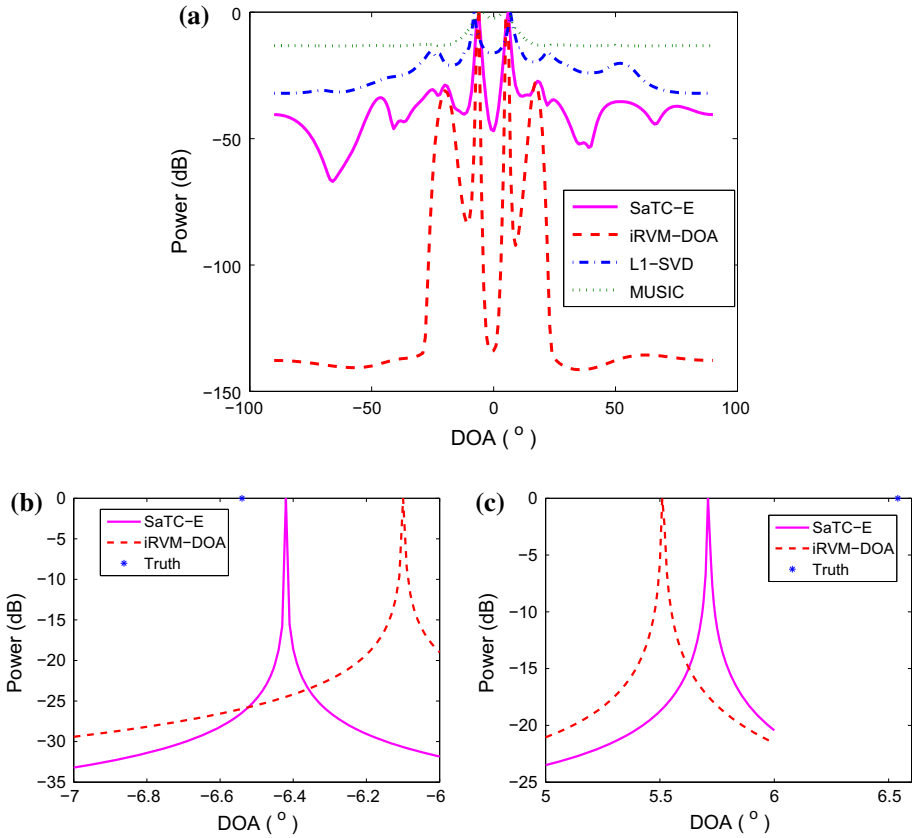
### 5.2 Numerical simulations

In this subsection, we analyze the statistical performance of SaTC-E. Monte Carlo trials are carried out to evaluate the statistical performance of the various methods. The statistical performance is measured by root mean square error (RMSE) of DOA estimation, which is calculated as

$$RMSE = \sqrt{\frac{1}{K \cdot MC} \sum_{mc=1}^{MC} \sum_{k=1}^K \left( \hat{\theta}_k^{(mc)} - \theta_k^{(mc)} \right)^2}, \tag{39}$$

where  $MC$  is the total number of Monte Carlo trials, and  $\hat{\theta}_k^{(mc)}$  is the estimate of the true direction  $\theta_k^{(mc)}$  of the  $k$ th source in the  $m$ cth trial. 200 independent trials are carried out to calculate the RMSE.

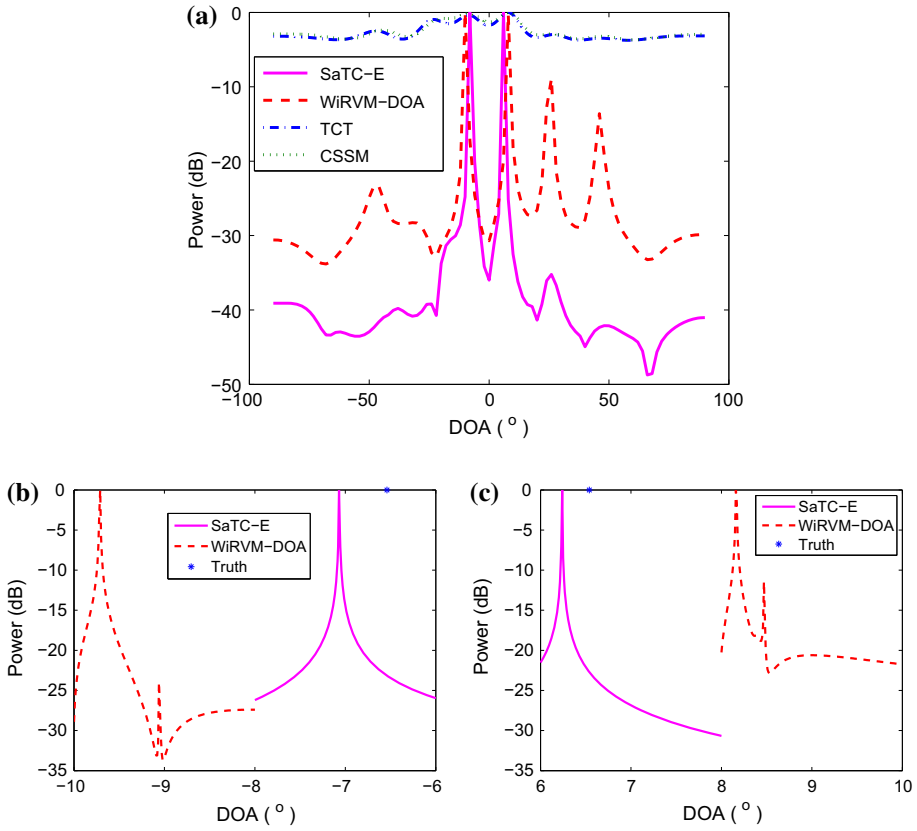
Before comparing the RMSE, we show the iteration convergence property of  $\boldsymbol{\gamma}$  updated as (12). Two narrowband sources are considered. The SNR is set to 20 dB, and the number



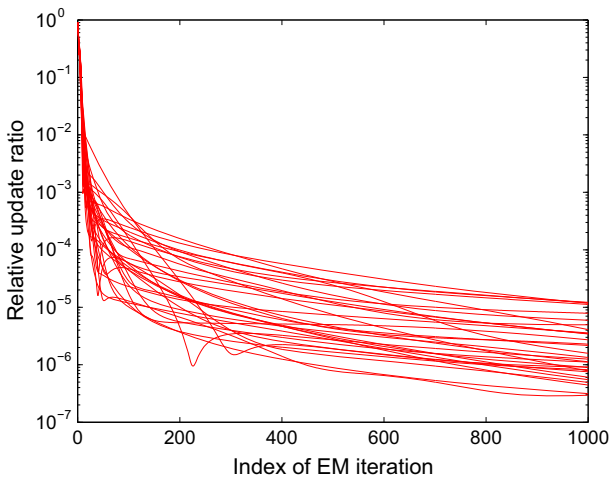
**Fig. 2** a Spatial spectrum of MUSIC and spatial spectra of L1-SVD, iRVM-DOA and SaTC-E at the coarse stage, and b, c spatial spectra of SaTC-E and iRVM-DOA at the refined stage. The two sources are narrowband. The number of snapshots is  $L = 20$

of snapshots is  $L = 20$ . We repeat 30 trials with 1000 iterations in each trial. The relative update ratio of  $\boldsymbol{\gamma}$ ,  $\|\boldsymbol{\gamma}^{(k+1)} - \boldsymbol{\gamma}^{(k)}\|_2 / \|\boldsymbol{\gamma}^{(k)}\|_2$ , is presented in Fig. 4. The curves indicate that in each trial, the estimate of  $\boldsymbol{\gamma}$  is approximately convergent with some small perturbations.

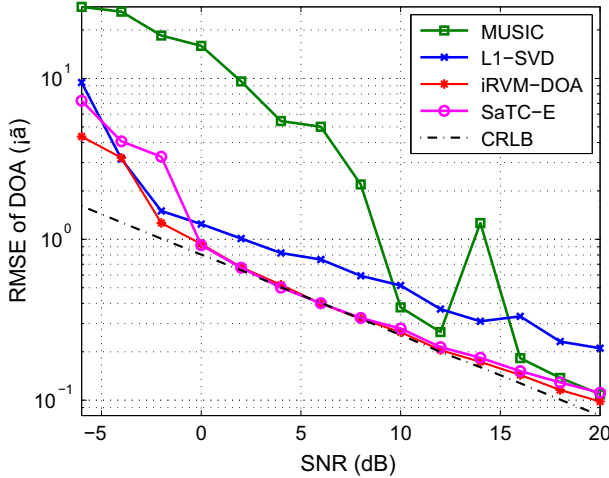
Assume that two temporally correlated sources imping on the 8-sensor ULA from directions  $-5^\circ + \Delta\theta$  and  $5^\circ + \Delta\theta$ , respectively, with  $\Delta\theta$  produced uniformly and randomly within the spatial scope  $[-1^\circ, 1^\circ]$ . The grid-intervals are  $\Delta\theta_c = 1^\circ$  and  $\Delta\theta_r = 0.01^\circ$ , and thus the source directions are off-grid usually. Each source is generated as AR(1) process, and thus the AR coefficient indicates the temporal correlation of this source. In this simulation, the AR coefficients of two sources in each trial are randomly chosen from  $[0.8, 1)$ . The RMSEs of DOA estimation with respect to SNR are presented in Fig. 5. The number of snapshots is  $L = 20$ . The SNR varies from  $-6$  to  $20$  dB with step  $2$  dB. The result shows that the RMSE of SaTC-E is very close to Cramer–Rao lower bound (CRLB) and performs as well as iRVM-DOA algorithm when SNR is larger than  $0$  dB. It significantly outperforms L1-SVD and MUSIC. Both for Fig. 5 and Fig. 6, when SNR is low, e.g.,  $\text{SNR} < 0$  dB, the RMSE of SaTC-E is larger than iRVM-DOA. This is because when the noise dominates the model error, the temporal correlation modeled by  $\mathbf{B}$  maybe not accurate enough.



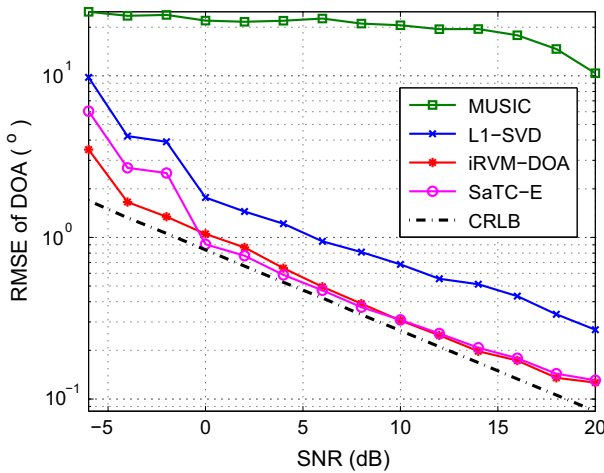
**Fig. 3** **a** Spatial spectra of TCT and CSSM, and spatial spectra of WiRVM-DOA and SaTC-E at the coarse stage, and **b**, **c** spatial spectra of SaTC-E and WiRVM-DOA at the refined stage. The sources are two LFM signals. The number of snapshots for TCT and CSSM is  $L = 320$ , while the number of snapshots for WiRVM-DOA and SaTC-E is  $L = 10$



**Fig. 4** The relative update ratio of  $\gamma$  in narrowband scenario. SNR=20dB and  $L = 20$



**Fig. 5** The RMSEs versus the SNR for two temporally correlated sources produced by AR(1) process. The number of snapshots is  $L = 20$



**Fig. 6** The RMSEs versus the SNR for two complex sinusoid sources. The number of snapshots is  $L = 20$

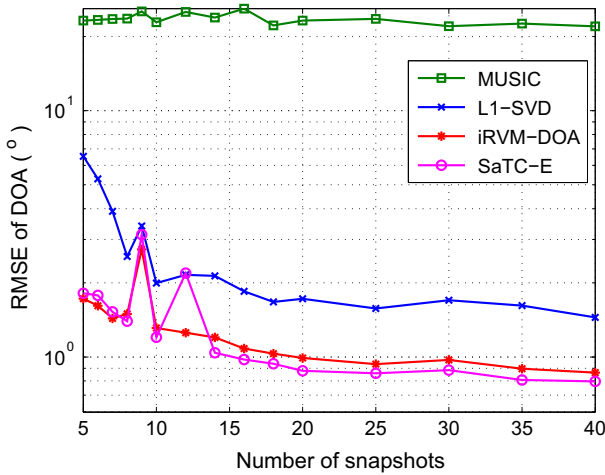
In the following simulations, we consider a common scenario in array signal processing that two complex sinusoids sources imping on the 8-sensor ULA. Other parameters are set as the simulation in Fig. 5. The complex sinusoid signal at time  $t$  is represented as

$$s_k(t) = u_k(t) \exp(2\pi f_0 t), \quad k = 1, 2, \tag{40}$$

where the amplitude  $u_k(t)$  obeys a Gaussian distribution with mean 100 and variance 5.

The RMSEs of DOA estimation versus the SNR are shown in Fig. 6. The SNR varies from  $-6$  to  $20$  dB with step 2 dB. The number of snapshots is  $L = 20$ . In this simulation, when SNR is high enough (e.g., SNR exceeds 0 dB), the RMSE of SaTC-E is very close to CRLB, while MUSIC fails in each SNR. In particular, the result in Fig. 6 indicates that SaTC-E surpasses the other methods including L1-SVD, MUSIC, and iRVM-DOA in middle SNR.





**Fig. 7** The RMSEs versus the number of snapshots. SNR= 0 dB

**Table 1** The average running time for DOA estimation (in seconds)

SNR (dB)	-4	0	4	8	12	16	20
TMSBL	747.94	735.37	704.42	617.93	514.02	468.69	447.76
SaTC-E	7.6137	1.7606	1.7749	1.8829	1.7962	1.6118	1.6270

When SNR is high, e.g.,  $SNR \geq 10$  dB in this simulation, SaTC-E significantly outperforms L1-SVD and MUSIC, and meanwhile it has quite similar performance to iRVM-DOA.

Figure 7 exhibits the curves of RMSE versus the number of snapshots. The number of snapshots  $L$  varies from 5 to 40. SNR is set to 0 dB. When  $L$  is large, e.g.,  $L > 10$  in this simulation, the RMSE of SaTC-E is smaller than that of iRMV-DOA. Furthermore, the RMSE of SaTC-E is much smaller than those of L1-SVD and MUSIC all along. The bad performances at SNR = 9 and 12 dB are caused by few outliers among the 200 Monte Carlo trials. This result demonstrates that SaTC-E has a superior performance of DOA estimation even uses a few snapshots.

Finally, we illustrate numerically that our method SaTC-E can reduce the computational complexity of BSBL proposed in Zhang and Rao (2011) for the problem of DOA estimation. TMSBL (Zhang and Rao 2011) is utilized as the original BSBL algorithm. The sources are produced as (40), the number of snapshots are  $L = 20$ , and the SNR varies from -4 to 20 dB with step 4 dB. The coarse grid-interval for SaTC-E is  $\Delta\vartheta_c = 2^\circ$  and the refined one is  $\Delta\vartheta_r = 0.1^\circ$ . The grid-interval for TMSBL is  $\Delta\vartheta = 0.1^\circ$ , too. This experiment is implemented in Matlab v.7.8.0 on a PC with a Window XP system and a 4 GHz CPU. The average running times of SaTC-E and TMSBL over 200 trials are presented in Table 1. The data in Table 1 indicates that SaTC-E is much faster than TMSBL as our expectation. In addition, we find that TMSBL almost can not recover the power  $\gamma$ , especially in low SNR, because of the serious correlation between the adjacent steering vectors. Consequently, TMSBL does no longer work for DOA estimation, while SaTC-E has improved performance with much higher efficiency.

## 6 Conclusion

A new method called SaTC-E has been proposed to estimate DOA for time-varying manifold scenario, which treats the scenario of narrowband stationary sources as a special case. SaTC-E exploits the spatial joint sparsity and temporal correlation of sources simultaneously via the BSBL methodology. In order to enhance the adaptation of BSBL algorithm to DOA estimation, SaTC-E utilizes the MLE of noise variance to take the place of the learning rule used in original BSBL algorithm. Meanwhile, it establishes a grid refined strategy to mitigate the estimate error caused by off-grid effect and reduce the computational complex of BSBL with improved performance. SaTC-E has some merits, including improved resolution, requirement of a few snapshots, no prior knowledge of source number, and applicability to spatially and temporally correlated sources. Real data tests and numerical simulations have been implemented to demonstrate the superiority of SaTC-E.

## Appendix 1

The detail for the derivatives (20) is presented as follows.

Let  $\mathbf{W} = \Sigma_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \otimes \mathbf{B}$ , and the derivatives are expressed as

$$\begin{aligned} \frac{\partial L(\tilde{\gamma}_k, \tilde{\vartheta}_k)}{\partial \tilde{\gamma}_k} &= \frac{\partial \ln |\mathbf{W}|}{\partial \tilde{\gamma}_k} + \frac{\partial \mathbf{z}^H \mathbf{W}^{-1} \mathbf{z}}{\partial \tilde{\gamma}_k} \\ &= \left\langle \frac{\partial \ln |\mathbf{W}|}{\partial \mathbf{W}}, \frac{\partial \mathbf{W}}{\partial \tilde{\gamma}_k} \right\rangle + \left\langle \frac{\partial \mathbf{z}^H \mathbf{W}^{-1} \mathbf{z}}{\partial \mathbf{W}}, \frac{\partial \mathbf{W}}{\partial \tilde{\gamma}_k} \right\rangle, \\ \frac{\partial L(\tilde{\gamma}_k, \tilde{\vartheta}_k)}{\partial \tilde{\vartheta}_k} &= \frac{\partial \ln |\mathbf{W}|}{\partial \tilde{\vartheta}_k} + \frac{\partial \mathbf{z}^H \mathbf{W}^{-1} \mathbf{z}}{\partial \tilde{\vartheta}_k} \\ &= \left\langle \frac{\partial \ln |\mathbf{W}|}{\partial \mathbf{W}}, \frac{\partial \mathbf{W}}{\partial \tilde{\vartheta}_k} \right\rangle + \left\langle \frac{\partial \mathbf{z}^H \mathbf{W}^{-1} \mathbf{z}}{\partial \mathbf{W}}, \frac{\partial \mathbf{W}}{\partial \tilde{\vartheta}_k} \right\rangle. \end{aligned}$$

Employing the matrix differential formulas that

$$\begin{aligned} \frac{\partial \ln |\mathbf{W}|}{\partial \mathbf{W}} &= \mathbf{W}^{-1}, \\ \frac{\partial \mathbf{W}}{\partial \tilde{\gamma}_k} &= \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \otimes \mathbf{B}, \\ \frac{\partial \mathbf{W}}{\partial \tilde{\vartheta}_k} &= \tilde{\gamma}_k \left( \mathbf{a}(\tilde{\vartheta}_k) \mathbf{d}^H(\tilde{\vartheta}_k) + \mathbf{d}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right) \otimes \mathbf{B}, \\ \frac{\partial \mathbf{z}^H \mathbf{W}^{-1} \mathbf{z}}{\partial \mathbf{W}} &= -\mathbf{W}^{-1} \mathbf{z} \mathbf{z}^H \mathbf{W}^{-1}, \end{aligned}$$

where  $\mathbf{W}$  is a conjugated symmetrical matrix, and  $\mathbf{d}(\tilde{\vartheta}_k) = \partial \mathbf{a}(\tilde{\vartheta}_k) / \partial \tilde{\vartheta}_k$ , and let  $\mathbf{V} = \mathbf{W}^{-1} - \mathbf{W}^{-1} \mathbf{z} \mathbf{z}^H \mathbf{W}^{-1}$ , we have

$$\begin{aligned} \frac{\partial L(\tilde{\gamma}_k, \tilde{\vartheta}_k)}{\partial \tilde{\gamma}_k} &= \left\langle \mathbf{V}, \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \otimes \mathbf{B} \right\rangle \\ &= \text{tr} \left\{ (\mathbf{V}) \left[ \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \otimes \mathbf{B} \right] \right\}, \\ \frac{\partial L(\tilde{\gamma}_k, \tilde{\vartheta}_k)}{\partial \tilde{\vartheta}_k} &= \tilde{\gamma}_k \left\langle \mathbf{V}, \left( \mathbf{a}(\tilde{\vartheta}_k) \mathbf{d}^H(\tilde{\vartheta}_k) + \mathbf{d}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right) \otimes \mathbf{B} \right\rangle \\ &= \tilde{\gamma}_k \text{tr} \left\{ \mathbf{V} \left[ \left( \mathbf{a}(\tilde{\vartheta}_k) \mathbf{d}^H(\tilde{\vartheta}_k) \right) \otimes \mathbf{B} \right] \right\} \\ &\quad + \tilde{\gamma}_k \text{tr} \left\{ \mathbf{V} \left[ \left( \mathbf{d}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right) \otimes \mathbf{B} \right] \right\}. \end{aligned}$$

In terms of the singular value decomposition (SVD),  $\mathbf{B}$  can be rewritten as

$$\mathbf{B} = \sum_{i=1}^{r_{\mathbf{B}}} \mathbf{u}_i^{\mathbf{B}} (\mathbf{u}_i^{\mathbf{B}})^H$$

with  $r_{\mathbf{B}} = \text{rank}(\mathbf{B})$  and  $\mathbf{u}_i^{\mathbf{B}}$  being the product of the  $i$ th singular value and the associated eigenvector of  $\mathbf{B}$ , So that

$$\begin{aligned} \frac{\partial L(\tilde{\gamma}_k, \tilde{\vartheta}_k)}{\partial \tilde{\gamma}_k} &= \text{tr} \left\{ \mathbf{V} \left[ \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \otimes \sum_{i=1}^{r_{\mathbf{B}}} \mathbf{u}_i^{\mathbf{B}} (\mathbf{u}_i^{\mathbf{B}})^H \right] \right\} \\ &= \sum_{i=1}^{r_{\mathbf{B}}} \text{tr} \left\{ \mathbf{V} \left[ \left( \mathbf{a}(\tilde{\vartheta}_k) \otimes \mathbf{u}_i^{\mathbf{B}} \right) \left( \mathbf{a}^H(\tilde{\vartheta}_k) \otimes (\mathbf{u}_i^{\mathbf{B}})^H \right) \right] \right\} \\ &= \sum_{i=1}^{r_{\mathbf{B}}} \left[ \mathbf{a}(\tilde{\vartheta}_k) \otimes \mathbf{u}_i^{\mathbf{B}} \right]^H \mathbf{V} \left[ \mathbf{a}(\tilde{\vartheta}_k) \otimes \mathbf{u}_i^{\mathbf{B}} \right]. \end{aligned}$$

The derivation is accomplished.

## Appendix 2

*Proof of Proposition 2* when  $\mathbf{B} = \mathbf{I}_L$ , we get  $\check{\Sigma}_{-k} = \check{\check{\Sigma}}_{-k} \otimes \mathbf{I}_L$ , and then  $L(\tilde{\gamma}_k, \tilde{\vartheta}_k)$  can be rewritten as

$$\begin{aligned} L(\tilde{\gamma}_k, \tilde{\vartheta}_k) &= \log \left| \left( \check{\check{\Sigma}}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right) \otimes \mathbf{I}_L \right| \\ &\quad + \mathbf{z}^H \left( \left( \check{\check{\Sigma}}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right) \otimes \mathbf{I}_L \right)^{-1} \mathbf{z}, \end{aligned}$$

and the first term of  $L(\tilde{\gamma}_k, \tilde{\vartheta}_k)$  equals to the first term of  $\check{L}(\tilde{\gamma}_k, \tilde{\vartheta}_k)$ , because

$$\begin{aligned} &\log \left| \left( \check{\check{\Sigma}}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right) \otimes \mathbf{I}_L \right| \\ &= \log \left( \left| \check{\check{\Sigma}}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right|^L |\mathbf{I}_L|^M \right) \\ &= \log \left| \check{\check{\Sigma}}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right|^L \\ &= L \log \left| \check{\check{\Sigma}}_{-k} + \tilde{\gamma}_k \mathbf{a}(\tilde{\vartheta}_k) \mathbf{a}^H(\tilde{\vartheta}_k) \right|. \end{aligned}$$

Let  $\mathbf{Q} = \left( \check{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a} \left( \tilde{\vartheta}_k \right) \mathbf{a}^H \left( \tilde{\vartheta}_k \right) \right)^{-1}$ , and write the array output matrix  $\mathbf{Y}$  as the form of

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} \in \mathbb{C}^{M \times L}, \mathbf{y}_i = [y_{i1} \ y_{i2} \ \dots \ y_{iL}] \in \mathbb{C}^{1 \times L},$$

so that  $\mathbf{z}^T = [\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_M] \in \mathbb{C}^{1 \times ML}$ . The second terms of  $L \left( \tilde{\gamma}_k, \tilde{\vartheta}_k \right)$  and  $\check{L} \left( \tilde{\gamma}_k, \tilde{\vartheta}_k \right)$  can be calculated, respectively, as

$$\begin{aligned} & \mathbf{z}^H \left( \left( \check{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a} \left( \tilde{\vartheta}_k \right) \mathbf{a}^H \left( \tilde{\vartheta}_k \right) \right) \otimes \mathbf{I}_L \right)^{-1} \mathbf{z} \\ &= \mathbf{z}^H \left( \mathbf{Q} \otimes \mathbf{I}_L \right) \mathbf{z} \\ &= \sum_{i=1}^M \sum_{j=1}^M q_{ij} \bar{\mathbf{y}}_i \mathbf{y}_j^T, \\ & \text{tr} \left( \mathbf{Y}^H \left( \check{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a} \left( \tilde{\vartheta}_k \right) \mathbf{a}^H \left( \tilde{\vartheta}_k \right) \right)^{-1} \mathbf{Y} \right) \\ &= \text{tr} \left( \mathbf{Q} \mathbf{Y} \mathbf{Y}^H \right) \\ &= \sum_{i=1}^M \sum_{j=1}^M q_{ij} \mathbf{y}_j \mathbf{y}_i^H. \end{aligned}$$

Because of  $\mathbf{y}_j \mathbf{y}_i^H = \sum_{k=1}^L y_{jk} \bar{y}_{ik} = \bar{\mathbf{y}}_i \mathbf{y}_j^T$ , we have

$$\begin{aligned} & \mathbf{z}^H \left( \left( \check{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a} \left( \tilde{\vartheta}_k \right) \mathbf{a}^H \left( \tilde{\vartheta}_k \right) \right) \otimes \mathbf{I}_L \right)^{-1} \mathbf{z} \\ &= \text{tr} \left( \mathbf{Y}^H \left( \check{\Sigma}_{-k} + \tilde{\gamma}_k \mathbf{a} \left( \tilde{\vartheta}_k \right) \mathbf{a}^H \left( \tilde{\vartheta}_k \right) \right)^{-1} \mathbf{Y} \right), \end{aligned}$$

and thus Proposition 2 is proved. □

### References

Baraniuk, R. (2007). Compressive sensing. *IEEE Signal Processing Magazine*, 24(4), 118–124.

Candes, E., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2), 489–509.

Carlin, M., Rocca, P., Oliveri, G., et al. (2013). Directions-of-arrival estimation through bayesian compressive sensing strategies. *IEEE Transactions on Antennas and Propagation*, 61(7), 3828–3838.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1), 1–38.

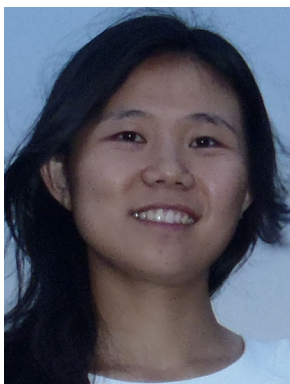
Donoho, D. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 1289–1306.

Du, X. P., & Cheng, L. Z. (2014). Three stochastic measurement schemes for direction-of-arrival estimation using compressed sensing method. *Multidimensional Systems and Signal Processing*, 25(4), 621–636.

Eldar, Y., & Mishali, M. (2009). Robust recovery of signals from a structured union of subspaces. *IEEE Transactions on Information Theory*, 55(11), 5302–5316.

Guo, X. S., Wan, Qun, Chang, C. Q., & Lam, E. Y. (2010). Source localization using a sparse representation framework to achieve superresolution. *Multidimensional Systems and Signal Processing*, 21(4), 391–402.

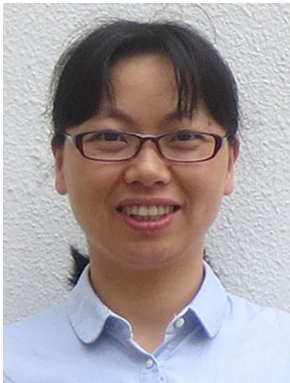
- Hyder, M., & Mahata, K. (2010). Direction-of-arrival estimation using a mixed  $l_{2,0}$  norm approximation. *IEEE Transactions on Signal Processing*, 58(9), 4646–4655.
- Kim, J., Lee, O., & Ye, J. (2012). Compressive MUSIC: Revisiting the link between compressive sensing and array signal processing. *IEEE Transactions on Information Theory*, 58(1), 278–301.
- Liu, Z., Huang, Z., & Zhou, Y. (2012). An efficient maximum likelihood method for direction-of-arrival estimation via sparse bayesian learning. *IEEE Transactions on Wireless Communications*, 11(10), 3607–3617.
- Malioutov, D., Cetin, M., & Willsky, A. (2005). A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Transactions on Signal Processing*, 53(8), 3010–3022.
- Northardt, E., Bilik, I., & Abramovich, Y. (2013). Spatial compressive sensing for direction-of-arrival estimation with bias mitigation via expected likelihood. *IEEE Transactions on Signal Processing*, 61(5), 1183–1195.
- Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3), 276–280.
- Stoica, P., & Nehorai, A. (1989). MUSIC, maximum likelihood and Cramer–Rao bound. *IEEE Transactions on Acoustics Speech and Signal Processing*, 37(5), 720–741.
- Tipping, M. (2001). Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1, 211–244.
- Valaee, S., & Kabal, P. (1995). Wideband array processing using a two-side correlation transformation. *IEEE Transactions on Signal Processing*, 43(1), 160–172.
- Wan, J., Zhang, Z., Yan, J., et al. (2012). Sparse bayesian multi-task learning for predicting cognitive outcomes from neuroimaging measures in Alzheimer’s disease. In *Proceedings of the IEEE international conference on computer vision and pattern recognition (CVPR)* (pp. 940–947), Providence, Rhode Island, June 2012.
- Wang, H., & Kaveh, M. (1985). Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources. *IEEE Transactions on Acoustics Speech and Signal Processing*, 33(4), 823–831.
- Wipf, D., Rao, B., & Nagarajan, S. (2011). Latent variable Bayesian models for promoting sparsity. *IEEE Transactions on Information Theory*, 57(9), 6236–6255.
- Zhang, Z., Jung, T., Makeig, S., et al. (2013). Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware. *IEEE Transactions on Biomedical Engineering*, 60(1), 221–224.
- Zhang, Z., Jung, T., Makeig, S., et al. (2013). Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse bayesian learning. *IEEE Transactions on Biomedical Engineering*, 60(2), 300–309.
- Zhang, Z., & Rao, B. (2011). Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning. *IEEE Journal of Selected Topics in Signal Processing*, 5(5), 912–926.
- Zhang, Z., & Rao, B. (2013). Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation. *IEEE Transactions on Signal Processing*, 61(8), 2009–2015.
- Zhu, W., & Chen, B. X. (2015). Novel methods of DOA estimation based on compressed sensing. *Multidimensional Systems and Signal Processing*, 26(1), 113–123.



**Bo Lin** received her B.S. degree in applied mathematics and M.S. degree in systems science both from the National University of Defense Technology (NUDT), Changsha, China, in 2008 and 2010, respectively. She is currently working towards the Ph.D degree in the Department of Mathematics and Systems Science, NUDT. Her research interests focus on array signal processing, signal sparse reconstruction, and compressive sensing.



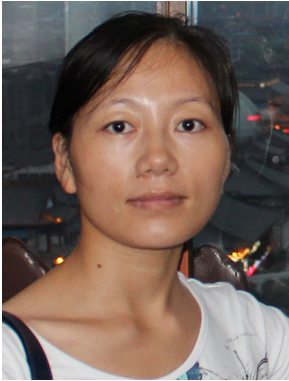
**Jiying Liu** received his B.S. degree in applied mathematics, M.S. degree in systems science and Ph.D degree in systems analysis and synthesis from NUDT, Changsha, China, in 2004, 2006, and 2010, respectively. He is currently a lecturer in the Department of Mathematics and Systems Science, NUDT. His research interests focus on signal processing, radar and communication applications, radar imaging and quantum imaging, and signal sparse reconstruction.



**Meihua Xie** received her B.S. degree and M.S. degree in applied mathematics, and Ph.D degree in systems engineering from NUDT, Changsha, China, in 1998, 2000, and 2005, respectively. She is currently an associate professor in the Department of Mathematics and Systems Science, NUDT. Her research interests focus on array signal processing, radar imaging, and image processing.



**Jubo Zhu** received his B.S. degree in applied mathematics from NUDT, Changsha, China, in 1988, M.S. degree in applied mathematics from Fudan University, Shanghai, China, in 1991, and Ph.D degree in information and communication engineering from NUDT in 2004, respectively. He is currently a professor in the Department of Mathematics and Systems Science, NUDT. His research interests focus on radar system design, mathematical imaging, signal sparse reconstruction, and compressive sensing.



**Fengxia Yan** received her M.S. degree and Ph.D. degree from National University of Defense Technology, Changsha, China, in 2003 and 2007. She is currently an associate professor with the Department of Mathematics and Systems Science, College of Science, NUDT. Her interests are in mathematical imaging, compressive sensing, and wavelet analysis with applications to image processing.