Direction-of-arrival estimation for co-located multiple-input multiple-output radar using structural sparsity Bayesian learning*

Wen Fang-Qing(文方青)[a)b)], Zhang Gong(张 弓)[a)b)†], and Ben De(贲 德)[a)b)c)]

[a)] College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b)] Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing 210016, China
[c)] Nanjing Research Institute of Electronics Technology, Nanjing 210039, China

(Received 27 March 2015; revised manuscript received 18 June 2015; published online 10 October 2015)

This paper addresses the direction of arrival (DOA) estimation problem for the co-located multiple-input multiple-output (MIMO) radar with random arrays. The spatially distributed sparsity of the targets in the background makes compressive sensing (CS) desirable for DOA estimation. A spatial CS framework is presented, which links the DOA estimation problem to support recovery from a known over-complete dictionary. A modified statistical model is developed to accurately represent the intra-block correlation of the received signal. A structural sparsity Bayesian learning algorithm is proposed for the sparse recovery problem. The proposed algorithm, which exploits intra-signal correlation, is capable being applied to limited data support and low signal-to-noise ratio (SNR) scene. Furthermore, the proposed algorithm has less computation load compared to the classical Bayesian algorithm. Simulation results show that the proposed algorithm has a more accurate DOA estimation than the traditional multiple signal classification (MUSIC) algorithm and other CS recovery algorithms.

Keywords: multiple-input multiple-output radar, random arrays, direction of arrival estimation, sparse Bayesian learning

PACS: 02.30.Zz, 02.50.–r, 87.16.dt **DOI:** 10.1088/1674-1056/24/11/110201

1. Introduction

The multiple-input multiple-output (MIMO) radar exploits multiple transmitting antennas to simultaneously transmit orthogonal waveforms, and utilizes multiple receiving antennas to receive the reflected signal. The information of an individual transmit path is extracted by a set of matched filters in the receiver, and the information from all the transmit paths enables the MIMO radar to achieve better performance than the traditional radars.[1] Theoretical research indicates that the MIMO radar has several built-in advantages in noise suppression, overcoming the fading effect, improving spatial resolution, enhancing parameter identifiability, etc.[2,3]

Direction of arrival (DOA) estimation is a fundamental task in the MIMO radar that has been widely investigated in the past decades. However, most of the previous studies focused on the uniform array geometry, such as uniform linear arrays (ULA), uniform circular-shape, and L-shape arrays. To avoid phase ambiguity, elements in the uniform array geometry are spaced at intervals no larger than half wavelength of the transmit signal carrier. The uniform array setup, which performs spatial sampling at the Nyquist rate, requires the product of transmit and receive elements scale linearly with the array aperture. In addition, this kind of configuration leads to high volume of data, which may be more serious in a massive MIMO radar system.[4] Generally speaking, angle resolution improves with the increase of array aperture. For a given number of antenna elements, the minimum redundancy linear array (MRLA) can achieve the maximum resolution by reducing the redundancy of spatial correlation lags.[5] An MRLA configuration was introduced to the MIMO radar in Ref. [6] and achieved similar resolution with the ULA architecture. Compared with the MRLA configuration, the random array pattern provides a more flexible form for the MIMO radar, where transmit and receive elements are randomly placed over a large aperture and spatial sampling is applied at sub-Nyquist rate. The random array setup would achieve similar resolution with significantly fewer elements. Hence, the random array configuration has attracted much attention in recent years.[7]

DOA estimation for the random array architecture MIMO radar can be accomplished with a variety of algorithms. Classical angle estimation algorithms such as multiple signal classification (MUSIC)[8] estimate parameters via the spectral peak searching. The estimation method of signal parameters via rotational invariance techniques (ESPRIT)[9] utilizes the invariance property of both transmit array and receive array for angle estimation. The array in ESPRIT algorithm must be composed of two identical, translationally invariant sub-

*Project supported by the National Natural Science Foundation of China (Grant Nos. 61071163, 61271327, and 61471191), the Funding for Outstanding Doctoral Dissertation in Nanjing University of Aeronautics and Astronautics, China (Grant No. BCXJ14-08), the Funding of Innovation Program for Graduate Education of Jiangsu Province, China (Grant No. KYLX 0277), the Fundamental Research Funds for the Central Universities, China (Grant No. 3082015NP2015504), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PADA), China.

†Corresponding author. E-mail: gzhang@nuaa.edu.cn

© 2015 Chinese Physical Society and IOP Publishing Ltd

arrays. One main drawback of these algorithms is the necessity of processing singular value decomposition (SVD) or eigenvalue decomposition (EVD) of the received data, which requires the prior information of the incident signal number. Besides, large numbers of snapshots are necessary to ensure the stability of SVD or EVD, which is impractical for the fast-moving targets, as the direction of target is continuously changing over a period of time. As a result, only a few snapshots are available for DOA estimation. From a sparsity point of view, the number of targets in the detection background is much smaller than the total number of the interesting cells. Therefore, the angle estimation from the subsampled array data can be linked to a sparse inverse problem from multiple measurement vectors (MMV) in compressive sensing (CS),[10,11] which can be solved by technologies such as basis pursuit (BP) algorithm,[12] orthogonal matching pursuit (OMP) algorithm,[13] FOCal underdetermined system solver (FOCUSS) algorithm,[14] and sparse bayesian learning (SBL) algorithm.[15] The motivation for sparse solution is to estimate the parameters more accurately compared to the traditional methods. Unfortunately, the complexity of BP algorithm is too high to implement. OMP and FOCUSS are sensitive to noise and have the problem of choosing a proper regularization parameter. As one important family of Bayesian algorithms, SBL has received much attention recently since it always achieves the sparsest global minima.[15] Besides, the statistical model in SBL provides a flexible framework to exploit special structures in the sparse signal,[16–23] which may significantly improve the recovery performance.

Generally speaking, the special structures of the source matrix in the MMV model can be divided into two categories.[16] The first category is inter-correlation that describes the distribution of the nonzero entities in each column vector. The second category is intra-correlation that indicates the correlation between the entries in each of the non-zero rows. In Ref. [24], the intra-correlation was modeled with a zero-mean Gaussian process, and a maximum-likelihood estimator is derived for DOA estimation. To reduce the computational load in Ref. [24], a block Gaussian *priori* was present to describe the inter-cluster of the spatial power spectrum in array antenna DOA estimation.[25] The coarse locations of the sources were obtained via SBL, and a peak searching procedure was introduced for refined DOA estimation. A block Gaussian process was modeled in Ref. [26] to capture the intra-correlation of the incident sources, and the block sparse Bayesian learning (BSBL) was derived for the off-grid DOA estimation. The sparse representation coefficients in distributed MIMO radar were assumed to satisfy a parameterized multivariate Gaussian distribution in Ref. [27], and BSBL was exploited for multi-targets localization. These structures have been shown to be very effective in robust parameter estimation with limited snapshots. However, intra-block correlation was ignored by the previous literature, which may result in performance loss in the reconstruction.

This paper links the DOA estimation problem in random manifold configured co-located MIMO radar to support recovery in the CS framework. The intra-signal correlation of the received snapshots is modeled with a modified statistical model, which represents the correlation of the snapshots more accurately. The BSBL framework is introduced to exploit the special correlation. Angle estimation is then linked to hyperparameters learning in the Bayesian framework, and a low complexity Bayesian learning algorithm is developed to speed up the learning process. The proposed algorithm does not require the prior information of the target number, which means that our algorithm has blind and robust characteristic, therefore the proposed algorithm has higher working efficiency. The proposed algorithm has better DOA estimation accuracy than MUSIC method[8] and conventional CS methods. Besides, our approach reduces the computational cost of the conventional SBL algorithm.[15] Extensive experiments show the superiority of the proposed algorithm.

The rest of the paper is organized as follows. Section 2 presents the signal model for the spatial CS-MIMO radar and provides the proposed Bayesian algorithm. Simulation results are given in Section 3. Finally, conclusion remarks are drawn in Section 4.

Notation Lower case and capital letters in bold denote, respectively, vectors and matrices. The superscripts $(\bullet)^{\mathrm{T}}$, $(\bullet)^{\mathrm{H}}$, $(\bullet)^{-1}$, and $(\bullet)^{\dagger}$ represent the operations of transpose, Hermitian transpose, inverse, and pseudo-inverse, respectively; the subscripts $\|\bullet\|_0$, $\|\bullet\|_1$, $\|\bullet\|_2$, and $\|\bullet\|_F$ separately denote the l_0-norm, l_1-norm, l_2-norm, and Frobenius norm, respectively; the vec$(\bullet)$ operator stacks the columns of its matrix argument in one tall column; $\otimes$ and $\odot$ represent the Kronecker product and Khatri–Rao product (column-wise Kronecker product).

2. Signal model and the proposed algorithm

2.1. Signal model

Consider the co-located MIMO radar model in Ref. [7], which is equipped with an M-element transmitter and an N-element receiver (as shown in Fig. 1). The transmitter and receiver are assumed to be linearly distributed in the x axis (possibly overlapping), with the total aperture Z_{TX} and Z_{RX} (normalized in wavelength units), respectively. The physical locations of the m-th transmitter and the n-th receiver are denoted by $Z\zeta_m/2$ and $Z\xi_n/2$, respectively. Here, $Z = Z_{\mathrm{TX}} + Z_{\mathrm{RX}}$, ζ_m lies in the interval $[-Z_{\mathrm{TX}}/Z,\ Z_{\mathrm{TX}}/Z]$, and ξ_n lies in the interval $\left[-Z_{\mathrm{RX}}Z,\ Z_{\mathrm{RX}}/Z\right]$. We assume that the transmitter antennas' locations ζ are independent and identically distributed (i. i. d.)

random variables governed by a distribution $p(\zeta)$. Similarly, the positions of the receivers ξ are drawn i. i. d. from a distribution $p(\xi)$. It is also assumed that there are K non-coherent targets appearing in the far-field of the radar system, the DOA of the k-th target is denoted by θ_k. Additional assumptions are that each transmit antenna emits a pulse train with L orthogonal waveforms, during which θ_k is a constant and the effect of jammers and interference is compensated. The output of the matched filters at the n-th receiver can be expressed as

$$\boldsymbol{y}_n^{(l)} = \sum_{k=1}^{K} \exp(\mathrm{j}\pi Z\theta_k\xi_n)\,\boldsymbol{c}(\theta_k)\,s_k^{(l)} + \boldsymbol{e}_n^{(l)}, \quad l = 1,2,\ldots,L, \tag{1}$$

where $\boldsymbol{y}_n^{(l)} \in \mathrm{C}^{M\times 1}$ is the matched filtering results for the l-th pulse, $\boldsymbol{c}(\theta_k) = [\exp(\mathrm{j}\pi Z\theta_k\zeta_1), \exp(\mathrm{j}\pi Z\theta_k\zeta_2),\ldots, \exp(\mathrm{j}\pi Z\theta_k\zeta_M)]^{\mathrm{T}} \in \mathrm{C}^{M\times 1}$ is the steering vector of the transmit array, $s_k^{(l)}$ is the radar cross section (RCS) fading coefficient of target k during the l-th pulse, and $\boldsymbol{e}_n^{(l)} \in \mathrm{C}^{M\times 1}$ is an additive Gaussian white noise vector.

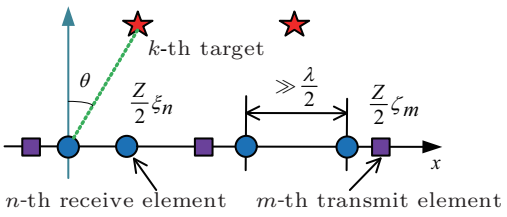


Fig. 1. (color online) MIMO radar system model.

Define $\boldsymbol{b}(\theta_k) = [\exp(\mathrm{j}\pi Z\theta_k\xi_1), \exp(\mathrm{j}\pi Z\theta_k\xi_2),\ldots, \exp(\mathrm{j}\pi Z\theta_k\xi_N)]^{\mathrm{T}} \in \mathrm{C}^{N\times 1}$ as the steering vector of the receive array. Let the receive direction matrix $\boldsymbol{B}(\boldsymbol{\theta}) = [\boldsymbol{b}(\theta_1), \boldsymbol{b}(\theta_2),\ldots,\boldsymbol{b}(\theta_K)]$, and the transmit direction matrix $\boldsymbol{C}(\boldsymbol{\theta}) = [\boldsymbol{c}(\theta_1), \boldsymbol{c}(\theta_2),\ldots,\boldsymbol{c}(\theta_K)]$. Arrange the matched filtering results as $\boldsymbol{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2,\ldots,\boldsymbol{Y}_N]^{\mathrm{T}} \in \mathrm{C}^{MN\times L}$ with $\boldsymbol{Y}_n = \left[\boldsymbol{y}_n^{(1)}, \boldsymbol{y}_n^{(2)},\ldots,\boldsymbol{y}_n^{(L)}\right] \in \mathrm{C}^{M\times L}$, $n = 1,2,\ldots,N$, then the received signal matrix $\boldsymbol{Y}$ can be expressed as

$$\boldsymbol{Y} = \tilde{\boldsymbol{A}}(\boldsymbol{\theta})\,\boldsymbol{S} + \boldsymbol{E}, \tag{2}$$

where $\tilde{\boldsymbol{A}}(\boldsymbol{\theta}) = \boldsymbol{B}(\boldsymbol{\theta}),\ldots,\boldsymbol{C}(\boldsymbol{\theta}) \in \mathrm{C}^{MN\times K}$ is the virtual array direction matrix, $S = [\boldsymbol{S}_1, \boldsymbol{S}_2,\ldots,\boldsymbol{S}_L] \in \mathrm{C}^{K\times L}$ with $\boldsymbol{S}_l = \left[s_1^l, s_2^l,\ldots,s_K^l\right]^{\mathrm{T}} \in \mathrm{C}^{K\times 1}$, $\boldsymbol{E} \in \mathrm{C}^{MN\times L}$ denotes the received noise matrix.

It is assumed that the targets are located in the maximum unambiguous angles. By discretizing the possible target DOA on a fine uniform grid, i.e., $[\varphi_1, \varphi_2,\ldots,\varphi_G]$ ($K \ll G$), we can obtain an over-complete dictionary $\boldsymbol{A}(\boldsymbol{\varphi}) = [\boldsymbol{a}(\varphi_1), \boldsymbol{a}(\varphi_2),\ldots,\boldsymbol{a}(\varphi_G)] \in \mathrm{C}^{MN\times G}$ for DOA estimation, where $\boldsymbol{a}(\varphi_g) = \boldsymbol{b}(\varphi_g) \otimes \boldsymbol{c}(\varphi_g)$, $g = 1,2,\ldots,G$. The received signal in Eq. (2) can be written as

$$\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{\varphi})\,\boldsymbol{X} + \boldsymbol{E}, \quad \boldsymbol{X}_{g\bullet} = \begin{cases} S_{k\bullet}, & \varphi_g = \theta_k, \\ 0, & \text{others}, \end{cases} \tag{3}$$

where $\boldsymbol{X}_{g\bullet}$ denotes the g-th row vector of matrix $\boldsymbol{X}$. Our aim is to recover θ from the given measurements $\boldsymbol{Y}$ and the known matrix $\boldsymbol{A}(\boldsymbol{\varphi})$, which is equal to the recovery of the support $\mathrm{supp}(\boldsymbol{X})$ (non-zero rows of X). Note that $K \ll MN < G$, equation (3) is a sparse inverse problem. The research in Ref. [7] has shown that once MN satisfies certain boundary conditions, $\boldsymbol{A}(\boldsymbol{\varphi})$ would provide a uniform recovery guarantee for Eq. (3). By relying on a random array geometry, the model in Eq. (3) is linked to the CS framework. In the case of $L > 1$, the sparse inverse problem in Eq. (3) can be extended to the noisy MMV model,[10] where common sparsity assumption on $\boldsymbol{X}$ is valid.

2.2. Problem statement

DOA estimation problem from Eq. (3) is equivalent to supporting recovery from the received array data, which can be accomplished by solving the following non-convex l_0-norm problem:

$$\mathrm{supp}(\boldsymbol{X}) = \arg\min \|\boldsymbol{X}\|_0, \ \text{s.t.} \ \|\boldsymbol{Y} - \boldsymbol{A}(\boldsymbol{\varphi})\,\boldsymbol{X}\|_{\mathrm{F}} \le \varepsilon, \tag{4}$$

where $\|\boldsymbol{X}\|_0$ is a row-diversity measure which counts the number of nonzero rows in $\boldsymbol{X}$, and ε represents the noise level. The cost function (4) can be globally minimized using a variety of optimization algorithms. However, as $MN \ll G$, the linear system in Eq. (3) is highly underdetermined and equation (4) is usually NP-hard. CS theory has pointed out that under specific conditions,[11] the l_0-norm optimization problem can be approximated by its l_1-norm relaxation with a bounded error, which is expressed as

$$\mathrm{supp}(\boldsymbol{X}) = \arg\min \|X\|_1, \ \text{s.t.} \ \|\boldsymbol{Y} - \boldsymbol{A}(\boldsymbol{\varphi})\,\boldsymbol{X}\|_{\mathrm{F}} \le \varepsilon. \tag{5}$$

Since the noise is present, the solution is decidedly nebulous. In practice, equation (5) is always transformed into an unconstrained regularized formulation and accomplished with an alternative strategy, suggesting the relaxed optimization problem

$$\mathrm{supp}(\boldsymbol{X}) = \arg\min_{\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{A}(\boldsymbol{\varphi})\,\boldsymbol{X}\|_{\mathrm{F}}^2 + \eta\,\Im(\boldsymbol{X}), \tag{6}$$

where η is a tradeoff parameter balancing the estimation quality. The regularization term $\Im(\boldsymbol{X})$ denotes the essential diversity measure, and it varies with the algorithm. When $\Im(\boldsymbol{X}) = \sum_{g=1}^{G} \|\boldsymbol{X}_{g\bullet}\|_2$, the algorithm in this category is a natural extension of the BP method,[12] while $\Im(\boldsymbol{X}) = \sum_{g=1}^{G} \log\|\boldsymbol{X}_{g\bullet}\|_2$ corresponds to the Jeffreys algorithm[13] and $\Im(\boldsymbol{X}) = \sum_{g=1}^{G} (\|X_{g\bullet}\|_2)^p$ to the FOCUSS algorithm.[14] By using an $\exp[-(\bullet)]$ transformation, the algorithm can be recast in Bayesian terms.

2.3. Block sparse Bayesian learning

The key idea of applying Bayesian framework to recover the row sparse signal $\boldsymbol{X}$ is to establish a hierarchical prior in $\boldsymbol{X}$. We assume that all sources $\boldsymbol{X}_{g\bullet}$ $(g=1,2,\ldots,G)$ are mutually independent, thus the inter-signal correlation in of the source is assumed to be an independent Gaussian process. We assign a structural prior on $\boldsymbol{X}_{g\bullet}$, the density of $\boldsymbol{X}_{g\bullet}$ is therefore given by

$$p(\boldsymbol{X}_{g\bullet};\gamma_g,\boldsymbol{R}_g)\sim N(0,\gamma_g\boldsymbol{R}_g),\quad g=1,2,\ldots,G, \tag{7}$$

where $N(0,\gamma_g\boldsymbol{R}_g)$ represents the Gaussian distribution with zero mean and covariance $\gamma_g R_g$. In this prior model, γ_g is a hyperparameter, $\boldsymbol{R}_g$ is the covariance matrix of $\boldsymbol{X}_{g\bullet}$. Hence, the inter-signal correlation of the source is modeled with a parameterized multivariate Gaussian process. The non-negative hyperparameter γ_g controls the row sparsity of $\boldsymbol{X}_{g\bullet}$, i.e., $\gamma_g=0$, the associated $\boldsymbol{X}_{g\bullet}$ becomes zeroes. Unlike the block sparse model in Ref. [21], the added positive matrix $\boldsymbol{R}_g$ captures the intra-block correlation of $\boldsymbol{X}_{g\bullet}$ and needs to be estimated.

We construct the following vectors: $\boldsymbol{y}=\mathrm{vec}\left(\boldsymbol{Y}^{\mathrm{T}}\right)$, $\boldsymbol{x}=\mathrm{vec}\left(\boldsymbol{X}^{\mathrm{T}}\right)$, and $\boldsymbol{n}=\mathrm{vec}\left(\boldsymbol{E}^{\mathrm{T}}\right)$. Let $\boldsymbol{\Theta}=\boldsymbol{A}(\varphi)\otimes\boldsymbol{I}_L$ with $\boldsymbol{I}_L$ denoting the identity matrix with the size of $L\times L$, the aforementioned MMV model is therefore transformed into a block sparse single measurement vector (SMV) model

$$\begin{aligned}\boldsymbol{y} &= [\boldsymbol{a}(\varphi_1)\otimes\boldsymbol{I}_L,\boldsymbol{a}(\varphi_2)\otimes\boldsymbol{I}_L,\ldots,\boldsymbol{a}(\varphi_G)\otimes\boldsymbol{I}_L]\\ &\quad\times[\boldsymbol{X}_{1\bullet},\boldsymbol{X}_{2\bullet},\ldots,\boldsymbol{X}_{G\bullet}]^{\mathrm{T}}+\boldsymbol{n}\\ &= \textstyle\sum_{g=1}^{G}\{\boldsymbol{a}(\varphi_1)\otimes\boldsymbol{I}_L\}\boldsymbol{X}_{g\bullet}+\boldsymbol{n}\\ &= \boldsymbol{\Theta}\boldsymbol{x}+\boldsymbol{n}.\end{aligned} \tag{8}$$

We postulate $\boldsymbol{n}$ to be i.i.d. zero mean Gaussian variable with variance λ, i.e., $p(\boldsymbol{n})\sim N(0,\lambda)$. As a result, the likelihood function for $\boldsymbol{y}$ in Eq. (8) could be formulated as

$$p(\boldsymbol{y}|\boldsymbol{x},\lambda)=(2\pi\lambda)^{-MNL/2}\exp\left(-\frac{1}{2\lambda}\|\boldsymbol{y}-\boldsymbol{\Theta}\boldsymbol{x}\|_2^2\right), \tag{9}$$

which is consistent with the likelihood model implied by Eq. (6) and previous Bayesian methods. To simplify the model, we just assume that all the sources have the same correlation structure, i.e., $\boldsymbol{R}_1=\boldsymbol{R}_2=\cdots=\boldsymbol{R}_G=\boldsymbol{R}$. We define $\gamma=[\gamma_1,\gamma_2,\ldots,\gamma_G]$, let $\boldsymbol{\Lambda}=\mathrm{diag}(\gamma)$ and $\boldsymbol{\Gamma}=\boldsymbol{\Lambda}\otimes R$. According to Eq. (7), the prior of $\boldsymbol{x}$ is

$$p(\boldsymbol{x};\gamma,\boldsymbol{R})=(2\pi)^{-\frac{GL}{2}}\det{}^{-\frac{1}{2}}(\boldsymbol{\Gamma})\exp\left\{-\frac{1}{2}x^{\mathrm{T}}(\boldsymbol{\Gamma})^{-1}x\right\}, \tag{10}$$

where $\det(\boldsymbol{\Gamma})$ denotes the determinant of matrix $\boldsymbol{\Gamma}$. By using the Bayes rule, the posterior probability function for $\boldsymbol{x}$ can be derived from Eqs. (9) and (10), which is also Gaussian and given by

$$\begin{aligned}&p(\boldsymbol{x}|\boldsymbol{y};\lambda,\gamma,\boldsymbol{R})\\ &= \frac{p(\boldsymbol{y}|\boldsymbol{x};\lambda)\,p(\boldsymbol{x}|\gamma;\boldsymbol{R})}{p(\boldsymbol{y}|\lambda,\gamma,\boldsymbol{R})}\\ &= (2\pi)^{-\frac{GL+1}{2}}\det{}^{-\frac{1}{2}}(\boldsymbol{\Sigma})\exp\left\{-\frac{1}{2}[x-\mu]^{\mathrm{T}}(\boldsymbol{\Sigma})^{-1}[\boldsymbol{x}-\boldsymbol{\mu}]\right\},\end{aligned} \tag{11}$$

with mean and covariance given by

$$\begin{cases}\boldsymbol{\mu}=\left(\lambda\boldsymbol{\Gamma}^{-1}+\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta}\right)^{-1}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{y}\\ \quad=\boldsymbol{\Gamma}\boldsymbol{\Theta}^{\mathrm{T}}\left(\lambda\boldsymbol{I}+\boldsymbol{\Theta}\boldsymbol{\Gamma}\boldsymbol{\Theta}^{\mathrm{T}}\right)^{-1}\boldsymbol{y},\\ \boldsymbol{\Sigma}=\left(\boldsymbol{\Gamma}^{-1}+\frac{1}{\lambda}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta}\right)^{-1}\\ \quad=\boldsymbol{\Gamma}-\boldsymbol{\Gamma}\boldsymbol{\Theta}^{\mathrm{T}}\left(\lambda\boldsymbol{I}+\boldsymbol{\Theta}\boldsymbol{\Gamma}\boldsymbol{\Theta}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Theta}\boldsymbol{\Gamma}.\end{cases} \tag{12}$$

Subsequently, to obtain the posterior probability function in Eq. (11), we have to estimate the hyperparameters λ, γ, and $\boldsymbol{R}$. Therefore, estimating the sparsity profile of $\boldsymbol{X}$ is conveniently shifted to hyperparameters learning. Usually, the expectation maximization (EM) method is employed for deriving these hyperparameters. The logarithm marginal likelihood function is constructed as

$$\begin{aligned}L(\lambda,\gamma,\boldsymbol{R}) &= \log p(\boldsymbol{y}|\lambda,\gamma,\boldsymbol{R})\\ &= \log p(\boldsymbol{y}|\boldsymbol{x},\lambda)+\log p(\boldsymbol{x}|\gamma,\boldsymbol{R})\\ &\propto -\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{y}-\log\{\det(\boldsymbol{\Phi})\},\end{aligned} \tag{13}$$

where $\boldsymbol{\Phi}=\lambda\boldsymbol{I}+\boldsymbol{\Theta}\boldsymbol{\Gamma}\boldsymbol{\Theta}^{\mathrm{T}}$. To obtain the hyperparameter λ, the EM method proceeds by treating γ and $\boldsymbol{R}$ as hidden variables and then maximizing the formulation in Eq. (13), we can obtain

$$\begin{aligned}L(\lambda) &= \log p(\boldsymbol{y}|\boldsymbol{x},\lambda)\\ &\propto -\frac{MNL}{2}\log\lambda-\frac{1}{2\lambda}\left\{\|\boldsymbol{y}-\boldsymbol{\Theta}\boldsymbol{x}\|_2^2+\mathrm{Tr}\left(\boldsymbol{\Sigma}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta}\right)\right\}\\ &= -\frac{MNL}{2}\log\lambda-\frac{1}{2\lambda}\left\{\|\boldsymbol{y}-\boldsymbol{\Theta}\boldsymbol{x}\|_2^2\right.\\ &\quad\left.-\hat{\lambda}\left[GL-\mathrm{Tr}\left(\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1}\right)\right]\right\},\end{aligned} \tag{14}$$

where $\hat{\lambda}$ represents the estimated λ in the previous iteration. Taking the derivative of Eq. (14) over λ to zero $\partial L(\lambda)/\partial\lambda=0$ yielding the learning rule for λ in the $(d+1)$-th iteration as

$$\lambda_{d+1}=\frac{\|\boldsymbol{y}-\boldsymbol{\Theta}\boldsymbol{x}_d\|_2^2+\lambda_d\left[GL-\mathrm{Tr}\left(\boldsymbol{\Sigma}_d\boldsymbol{\Gamma}_d^{-1}\right)\right]}{MNL}, \tag{15}$$

where λ_d, $\boldsymbol{x}_d$, $\boldsymbol{\Sigma}_d$, and $\boldsymbol{\Gamma}_d$ are previously obtained estimates of λ, $\boldsymbol{x}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Gamma}$, respectively. Similarly, to obtain the hyperparameters γ and $\boldsymbol{R}$, the EM method proceeds by treating λ as a hidden variable and then maximizing the formulation in Eq. (13), we can obtain

$$\begin{aligned}L(\gamma,\boldsymbol{R}) &= \log p(\boldsymbol{x}|\gamma,\boldsymbol{R})\\ &= -\frac{L}{2}\log\{\det(\boldsymbol{\Lambda})\}-\frac{G}{2}\log\{\det(R)\}\\ &\quad-\frac{1}{2}\mathrm{Tr}\left\{\left(\boldsymbol{\Lambda}^{-1}\otimes R^{-1}\right)\left(\boldsymbol{\Sigma}+\mu\mu^{\mathrm{T}}\right)\right\}.\end{aligned} \tag{16}$$

Taking the gradient of Eq. (16) with respect to γ_g $(g=1,2,\ldots,G)$ equal to zero, i.e., $\partial L(\gamma,\boldsymbol{R})/\partial\gamma_g=0$, and treating $\boldsymbol{R}$ as a known matrix, we obtain the learning rule for γ_g in the $(d+1)$-th iteration as

$$(\gamma_g)_{d+1}=\frac{1}{L}\mathrm{Tr}\left\{\boldsymbol{R}_d^{-1}\left[\boldsymbol{\Sigma}_d+\boldsymbol{\mu}_d\boldsymbol{\mu}_d^{\mathrm{T}}\right]\right\}, \tag{17}$$

where $\boldsymbol{\mu}_d$, $\boldsymbol{\Sigma}_d$, and $\boldsymbol{R}_d$ are previously obtained estimates of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{R}$, respectively. By treating γ and λ as hidden variables, taking the derivative of Eq. (16) over $\boldsymbol{R}$ to zero, i.e., $\partial L(\gamma,\boldsymbol{R})/\partial\boldsymbol{R}=0$, while treating γ as a known vector, we obtain the learning rule for $\boldsymbol{R}$ in the $(d+1)$-th iteration as

$$\boldsymbol{R}_{d+1}=\frac{1}{Q}\sum_{q=1}^{Q}\frac{\boldsymbol{\Sigma}_d+\boldsymbol{\mu}_d\boldsymbol{\mu}_d^{\mathrm{T}}}{(\gamma_q)_d}, \tag{18}$$

where $\boldsymbol{\mu}_d$, $\boldsymbol{\Sigma}_d$, and $(\gamma_q)_d$ are previously obtained estimates of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and γ_q, respectively. The algorithm using iterative learning rules (12), (15), (17), and (18) is denoted by block SBL.

2.4. A faster version for parameters learning

BSBL was proved to provide highly accurate solutions. However, it suffers from high computational complexity, as the vec$(\bullet)$ operator maps the inverse problem to a higher-dimensional space. To speed up the learning process of block SBL, we map the learning of the parameters in a higher-dimensional space to the original problem. The mean and covariance in Eq. (12) are updated using the rules in MSBL method[15]

$$\begin{cases}(\boldsymbol{\Sigma}_{\boldsymbol{X}})_{d+1}=\left(\boldsymbol{\Lambda}_d^{-1}+\dfrac{1}{\lambda_d}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{\varphi})\boldsymbol{\Lambda}_d\right),\\ \boldsymbol{X}_{d+1}=\boldsymbol{H}_d\boldsymbol{Y},\end{cases} \tag{19}$$

where $\boldsymbol{H}_d=\boldsymbol{\Lambda}_d\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{\varphi})\left(\lambda_d\boldsymbol{I}+\boldsymbol{A}(\boldsymbol{\varphi})\boldsymbol{\Lambda}_d\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{\varphi})\right)^{-1}$. A series of reasonable approximations are proposed in Ref. [16] to simplify the learning rules of Eqs. (15), (17), and (18), which are summarized as the iteration of the following steps:

$$(\gamma_g)_{d+1}=\frac{1}{L}(\boldsymbol{X}_{g\bullet})_d\boldsymbol{R}_d^{-1}\left(\boldsymbol{X}_{g\bullet}^{\mathrm{T}}\right)_d+((\boldsymbol{\Sigma}_{\boldsymbol{X}})_d)_{gg},\quad g=1,2,\ldots,G; \tag{20}$$

$$\boldsymbol{R}_{d+1}=\tilde{\boldsymbol{R}}_{d+1}/\left\|\tilde{\boldsymbol{R}}_{d+1}\right\|_{\mathrm{F}},$$

$$\tilde{\boldsymbol{R}}_{d+1}=\sum_{g=1}^{G}\left(\boldsymbol{X}_{g\bullet}^{\mathrm{T}}\right)_d(\boldsymbol{X}_{g\bullet})_d+\eta\boldsymbol{I}; \tag{21}$$

$$\lambda_{d+1}=\frac{1}{MNL}\left\|\boldsymbol{Y}-\boldsymbol{A}(\boldsymbol{\varphi})\boldsymbol{X}_d\right\|_{\mathrm{F}}^2+\frac{\lambda_d}{G}\mathrm{Tr}\left[\boldsymbol{A}(\boldsymbol{\varphi})\boldsymbol{H}_d\right], \tag{22}$$

where η is a positive scalar. The faster version is very similar to that of MSBL method. The only difference is the added $\boldsymbol{R}$ in γ_g learning. Matrix $\boldsymbol{R}$ incorporates the structural correlation of the sources, which is positive definite and symmetric while with only a small size, therefore, the extra computational load is low. It has been proved that the faster version has the very desirable property that the global minimum is at the sparsest solution to the MMV problem.[16]

Till now, we have achieved the proposal for the structural sparsity Bayesian learning based on the DOA estimation algorithm in the MIMO radar. The major steps of the proposed algorithm are shown as follows.

1) Initialize $\boldsymbol{X}$ to a matrix that elements are all 1, initialize $\boldsymbol{R}$ to an identity matrix, set $\lambda=10^{-6}$ and set the iteration count $d=0$.

2) Subsequently, solve the MMV minimization problem by alternatively iterating among Eqs. (20)–(22).

3) Afterwards, compute the posterior moments $\boldsymbol{X}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ using Eq. (19).

4) Repeat Eqs. (2) and (3) until convergence or the iteration count d attains a specified maximum number.

3. Simulation results and analysis

In this section, Monte Carlo simulations are used to assess the DOA estimation performance of the proposed algorithm. Success-rate and average running time are used for performance evaluation. A successful trial is recognized if the indexes of estimated sources with the K largest l_2-norms are the support of $\boldsymbol{X}$. In our simulation, the total aperture of the elements is $Z=250$. It is assumed that targets are located in the far field, and the RCS coefficients of each target satisfy the first-order autoregressive process with the correlation coefficient $\beta=0.9$. In our experiments, we compare our algorithm with the traditional MUSIC algorithm and the following CS recovery algorithms for the MMV model: the MOMP method proposed in Ref. [13], the MFOCUSS algorithm proposed in Ref. [14], and the MSBL method proposed in Ref. [15], which could be regarded as a faster version of the BSBL method. The computer used to run the simulation is a HP Z400 workstation, which is configured with four Intel Xeon W3550 3.07 GHz processors, and 12 GB of RAM.

Figure 2 depicts the DOA estimation performance comparison of all the algorithms with $L=5$, $M=N=6$, $\boldsymbol{\theta}=[15,40,65]$ and different values of SNR in the numerical simulations. According to Fig. 2, all algorithms would achieve better performance with the increase of the SNR. It is obvious that the proposed algorithm could provide almost precise results. Notably, compared with the MSBL method, our algorithm has better DOA estimation accuracy at low SNR, and the proposed algorithm is appealing from the running time point of view, since more precision of the estimated temporary variables would result in much fewer learning steps. As shown in this figure, the MOMP method possesses poor success rate. This is caused by the severe mutual coherence between the atoms of the dictionary $A(\varphi)$. The greedy MOMP algorithm ensures that the residual is orthogonal to the atoms chosen in the last iteration, which may destroy the information hiding in

the residual when updating the new residual. Meanwhile, the traditional MUSIC algorithm achieves a relatively poor estimation accuracy with a limited number of snapshots.

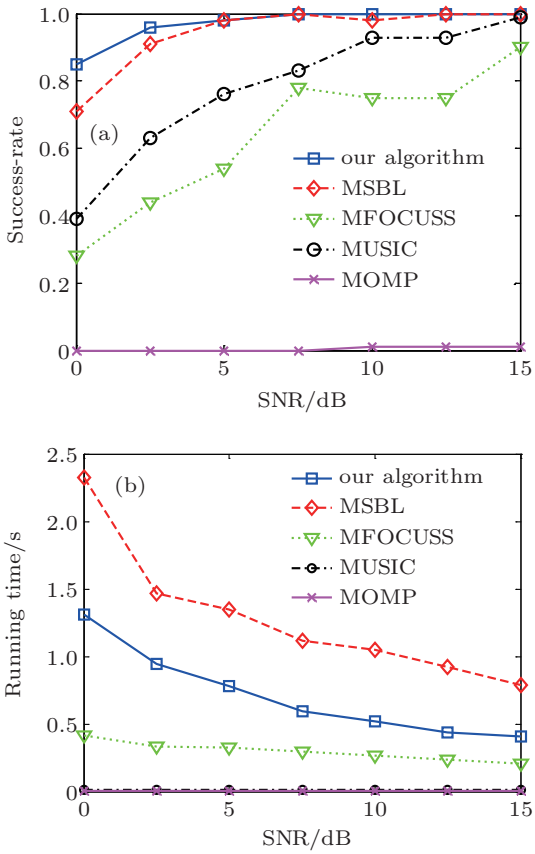


Fig. 2. (color online) Performance of all the algorithms with different SNR. (a) Success-rate comparison and (b) average running time comparison.

The DOA estimation performance comparison of all the algorithms with respect to different sparsity K was investigated in the simulation, as shown in Fig. 3. Therein, $M = N = 6$, $L = 5$, and SNR $= 10$ were considered in the simulation. In each trial, the locations of the targets were randomly uniformly chosen from the feasible regions 0°–90°. It is clearly shown that the Bayesian algorithm accurately recovers much more targets than other algorithms. Thanks to the learning process of the intra-block correlation parameter, the proposed algorithm performs better than the MSBL method.

Figure 4 shows the DOA estimation performance comparison of all the algorithms with different number of receive antenna N, where $M = 6$, $L = 5$, $\boldsymbol{\theta} = [15, 40, 65]$, and SNR $= 10$. It indicates that angle estimation accuracy of all the algorithms were improved with N increasing. The proposed algorithm requires significantly fewer numbers of antenna for the same successful enumeration, and only requires about 5 receive antennas to start stabilizing (success-rate more than 95%). Simulation results are consistent with CS theory, which indicates that recovery performance would rapidly be improved with the increase of the measurement length, as more measurements would provide more information for the sparse recovery.

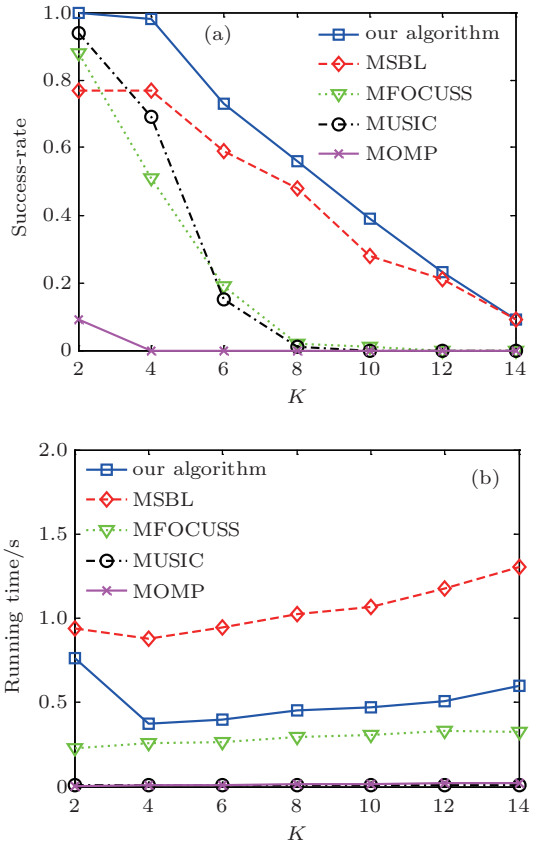


Fig. 3. (color online) Performance of all the algorithms with different sparsity K. (a) Success-rate comparison and (b) average running time comparison.

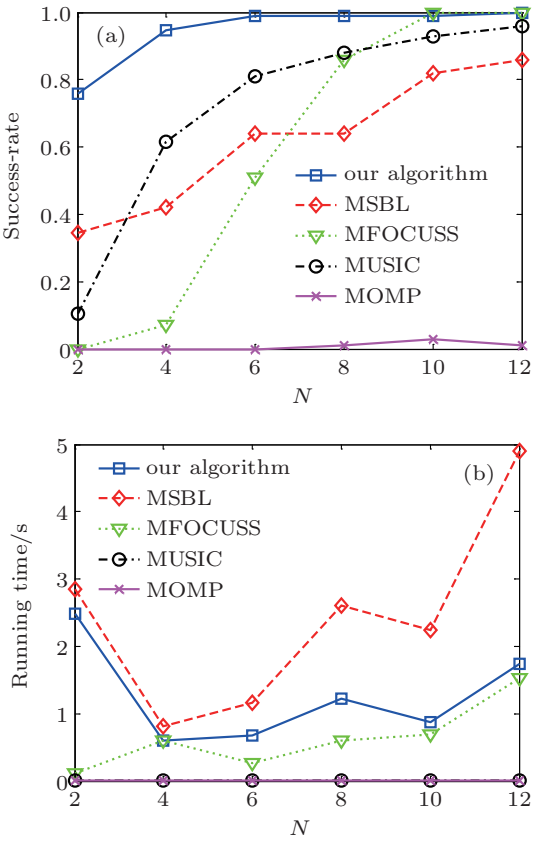


Fig. 4. (color online) Performance of all the algorithms with different antenna number N. (a) Success-rate comparison and (b) average running time comparison.

Figure 5 illustrates the angle estimation performance comparison of all the algorithms with different snapshots L, where $M = N = 6$, $\boldsymbol{\theta} = [15, 40, 65]$, and SNR $= 10$. Figure 5(a) shows clearly that the increasing number of snapshots improves the estimation resolution. However, once L reaches a value (approximately equal to 5 in our simulation), the benefit from MMV becomes indistinct. Therefore, an appropriate number of snapshots is important for the proposed algorithm, as a smaller L value may impair the angle estimation accuracy, and large L values may require more computational resources.

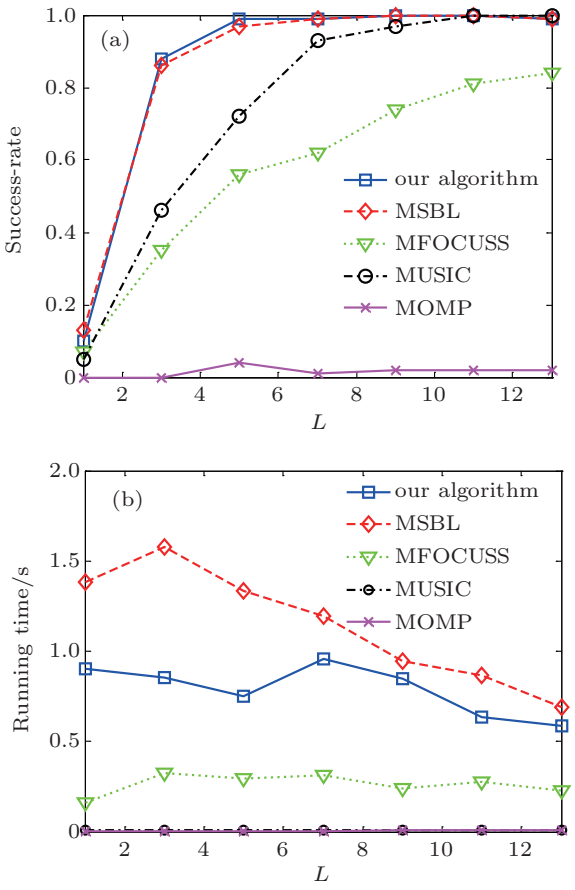


Fig. 5. (color online) Performance of all the algorithms with different snapshot L. (a) Success-rate comparison and (b) average running time comparison.

4. Conclusion

In this paper, we have proposed a structural sparsity Bayesian model for DOA estimation in the spatial CS MIMO radar. Our work links the DOA estimation problem to the parameters learning in the Bayesian framework, which significantly reduces the computation load of conventional Bayesian algorithm. The proposed algorithm requires no prior information of the number of targets, and achieves DOA estimation performance better than the traditional Bayesian algorithm. The proposed algorithm outperforms the traditional MUSIC algorithm and other CS recovery algorithms with fewer numbers of snapshots and low SNR scene.

References

[1] Fishler E, Haimovich A, Blum R, Chizhik D, Cimini L and Valenzuela R 2004 *IEEE Radar Conference*, April 26–29, 2004, Philadelphia, Pennsylvania, p. 71
[2] Xiao S, Cai J J, Wang R L, Liu M Z and Liu F 2009 *Chin. Phys. B* **18** 5103
[3] Huang C, Zhang D J, Zhang D L and Teng T T 2014 *Acta Phys. Sin.* **63** 188401 (in Chinese)
[4] Larsson E G, Edfors O, Tufvesson F and Marzetta T L 2014 *IEEE Commun. Mag.* **52** 186
[5] Duan H P, Ng B P, See C M S and Fang J 2008 *IEEE Trans. Signal Process.* **56** 2406
[6] Wen F Q and Zhang G 2013 *Math. Prob. Eng.* **2013** 427980
[7] Rossi M, Haimovich A M and Eldar Y C 2014 *IEEE Trans. Signal Process.* **62** 419
[8] Schmidt R O 1986 *IEEE Trans. Antennas Propag.* **34** 276
[9] Roy R and Kailath 1989 *IEEE Trans. Acous. Speech Signal Process.* **37** 984
[10] Cotter S F, Rao B D, Engan K and Kenneth K D 2005 *IEEE Trans. Signal Process.* **53** 2477
[11] Cand'es E J, Romberg J and Tao T 2006 *IEEE Trans. Inform. Theory* **52** 489
[12] Sun Y L and Tao J X 2014 *Chin. Phys. B* **23** 078703
[13] Tropp J A and Gilbert A C 2007 *IEEE Trans. Inform. Theory* **53** 4655
[14] Ning F L, He B J and Wei J 2013 *Acta Phys. Sin* **62** 174214 (in Chinese)
[15] Wipf D P and Rao B D 2007 *IEEE Trans. Signal Process.* **55** 3704
[16] Zhang Z L and Rao B D 2011 *IEEE J-STARS* **5** 912
[17] Eldar Y C and Mishali M 2009 *IEEE Trans. Inform. Theory* **55** 5302
[18] Huang S X, Zhao X F and Sheng Z 2009 *Chin. Phys. B* **18** 5084
[19] Sheng Z 2013 *Chin. Phys. B* **22** 029302
[20] Zhang Z L and Rao B D 2013 *IEEE Trans. Signal Process.* **61** 2009
[21] Negahban S and Wainwright M J 2011 *IEEE Trans. Inform. Theory* **57** 3841
[22] Zhao J, Yu L and Li J R 2013 *Acta Phys. Sin.* **62** 130201 (in Chinese)
[23] Wen F Q, Zhang G and Ben D 2015 *Acta Phys. Sin.* **64** 070201 (in Chinese)
[24] Chen C E, Lorenzelli F, Hudson R E and Yao K 2008 *IEEE Trans. Signal Process.* **56** 3038
[25] Liu Z M, Huang Z T and Zhou Y Y 2012 *IEEE Trans. Wirel. Commun.* **10** 1
[26] Zhang Y, Ye Z, Xu X and Hu N 2014 *Signal Process.* **98** 197
[27] Sun B, Chen H, Wei X and Li X 2014 *Int. J. Antenn. Propag.* **2014** 903902