Journal of Systems Engineering and Electronics
Vol. 37, No. 3, June 2026, pp.826 – 835

DOA estimation via a Newton-like method under unknown mutual coupling

LYU Jihui[1], LIU Shuai[2,*], JIN Ming[2], YAN Fenggang[2], and SONG Lizhong[2]

1. School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China;
2. School of Information Science and Engineering, Harbin Institute of Technology (Weihai), Weihai 264200, China

Abstract: As is well known, mutual coupling between array elements has a significant negative impact on direction of arrival (DOA) estimation. To achieve DOA estimation under unknown mutual coupling, this paper proposes a low computational complexity Newton-like method. Firstly, a block sparse model based on the signal subspace is established, and the Lagrangian function is established according to the block sparse model. Secondly, since the Hessian matrix of the Lagrangian function cannot always ensure positive definiteness and the computational complexity of the inverse matrix of the Hessian matrix is enormous, the Newton method is no longer applicable. Therefore, this paper proposes a Newton-like method to achieve DOA estimation under mutual coupling and reduce the computational complexity by matrix inversion lemma. Finally, compared with existing methods of DOA estimation under array mutual coupling, the simulation results validate the effectiveness of the proposed method.

Keywords: direction of arrival (DOA), mutual coupling, block sparse, Newton-like method.

DOI: 10.23919/JSEE.2026.000110

1. Introduction

The research of direction of arrival (DOA) estimation plays a crucial role in the field of array signal processing [1−3], which is widely used in military and economic fields, such as radar [4,5], communication [6,7], and sonar [8]. In the past few decades, subspace decomposition methods have developed rapidly and achieved great success. This method's core process is obtaining the signal subspace from the sampling covariance matrix of the incident signals. A representative method of this type is the multiple signal classification (MUSIC) [9] method. In recent years, signal sparsity theory has developed rapidly and has been well applied in the research of array signal processing, such as DOA estimation [10−12]. Compared with subspace decomposition techniques, sparse DOA estimation methods can provide better accuracy in more harsh environments, such as low signal-to-noise ratio (SNR) or small snapshot situations [13,14]. However, the above methods did not consider non-ideal factors, such as the mutual coupling between array elements. Weiss et al. [15] demonstrated that mutual coupling significantly impacts the accuracy of the incident signal's DOA when the array manifold matrix is affected by it.

Recently, several papers have emerged on DOA estimation under array mutual coupling. The solution methods for this problem mainly consist of three categories. The first category is preprocessing methods. This category of methods construct a preprocessing operator based on the number of mutual coupling coefficients, which can effectively reduce the negative impact of array mutual coupling. After preprocessing, the non-ideal array manifold matrix affect by mutual coupling can be restored to the ideal Vandermonde matrix [16]. However, the construction of the preprocessing operator comes at the cost of sacrificing the array aperture [17,18]. The second category involves the unique deformation of the production of the mutual coupling matrix (MCM) and the array manifold matrix. Since the MCM is a Toeplitz matrix and the array manifold matrix is a Vandermonde matrix, their product can be represented as a new transformation. Additionally, based on the orthogonality between the noise subspaces and the steering vector of the incident signal [14], Liao et al. [19] and Ge et al. [20] proposed a rank deficiency method, respectively. Since the noise subspace is indispensable in the methods above, these methods are not applicable in scenarios with coherent signals. The third category of methods is the block sparse method. For example, according to the preprocessing method proposed by [17], Dai et al. [21] restored the

Manuscript received September 10, 2024.
*Corresponding author.
This work was supported by the National Natural Science Foundation of China (62071144; 62171150), and Taishan Scholar Special Funding Project of Shandong Province (tsqn202211087).

array manifold matrix affected by mutual coupling to the ideal Vandermonde matrix and established a $l_{2,0}$ mixed norm to solve the model, where $l_{2,0}$ is the mixed norms of l_2 and l_0; Wang et al. [22] used the unique deformation of the product of the MCM and the array manifold matrix to establish a new block sparse model about mutual coupling. Unlike [21], Wang et al. [22] solved this problem using the $l_{F,0}$ norm, which provided a novel perspective for us, where $l_{F,0}$ is the mixed norms of l_F and l_0. Since singular value decomposition (SVD) is used for preprocessing, the above methods still apply to coherent signals. However, the DOA estimation accuracy of the above two methods is not ideal; in order to further improve the DOA estimation performance of the method proposed by [22], Wang et al. [23] used a set of weights that can enhance the sparsity of the $l_{F,1}$ norm, which improved the DOA estimation accuracy, where $l_{F,1}$ is the mixed norms of l_F and l_1. Meng et al. [24,25] weighted the signal subspace affected by mutual coupling appropriately, which also improves the accuracy of DOA estimation. However, the above block sparse methods have high computational complexity, so they cannot achieve DOA estimation quickly.

In this paper, a low computational complexity Newton-like method is designed to solve the $l_{F,0}$ norm, achieving DOA estimation under unknown mutual coupling. Firstly, a block sparse model under array mutual coupling is established via signal subspace. Then, we use exponential to represent the $l_{F,0}$ norm and establish the Lagrangian function $L_\sigma(\bar{\boldsymbol{B}})$. However, the Lagrangian function's Hessian matrix cannot always guarantee positive definiteness, and the dimension of the Hessian matrix of $L_\sigma(\bar{\boldsymbol{B}})$ is very large. The dimensionality of the Hessian matrix is too large, which means that the calculation of its inverse matrix will be a difficult task. To maximize computational efficiency as much as possible, we design a Newton-like method based on mapping $\zeta(\bar{\boldsymbol{B}})$ to achieve DOA estimation and reduce its computational complexity by matrix inversion lemma. Finally, simulations demonstrate the effectiveness of the proposed method in DOA estimation.

2. Signal model

Assuming the number of signal sources is Q, and the incident DOA of the signals are $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_Q]$. The array consists of M elements with a uniform spacing of half signal wavelength, i.e., $d = \bar{\lambda}/2$, $\bar{\lambda}$ is the signal's wavelength. As shown in Fig. 1, the received signal of the array at time t under array mutual coupling is represented as

$$\boldsymbol{X}(t) = \sum_{q=1}^{Q} \boldsymbol{C}\boldsymbol{a}(\theta_q)\boldsymbol{s}_q(t) + \boldsymbol{n}(t) = \boldsymbol{C}\boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{1}$$

where $\boldsymbol{X}(t)$ is the received signal from far-field at time t under mutual coupling, $\boldsymbol{a}(\theta_q) = [1, \beta(\theta_q), \cdots, \beta(\theta_q)^{M-1}]^{\mathrm{T}}$ is the steering vector unaffected by mutual coupling of the qth far-field signal, $\beta(\theta_q) = \exp\left(-\mathrm{j}2\pi d \sin\theta_q/\bar{\lambda}\right)$, and $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \cdots, \boldsymbol{a}(\theta_Q)]$ is the array manifold matrix unaffected by mutual coupling, $\boldsymbol{s}(t) = [\boldsymbol{s}_1(t), \boldsymbol{s}_2(t), \cdots, \boldsymbol{s}_Q(t)]^{\mathrm{T}}$ is the source signal, $\boldsymbol{n}(t)$ represents Gaussian additive white noise $\sim \mathrm{CN}(0, \sigma_n^2 \boldsymbol{I}_M)$, σ_n^2 is the noise power, $\boldsymbol{C} = \mathrm{Toeplitz}(\boldsymbol{c}, \boldsymbol{0}_{M-P}^{\mathrm{T}})$ is the MCM, $\boldsymbol{C} \in \mathbf{C}^{M\times M}$ is a Toeplitz matrix, where $\boldsymbol{c}$ is the row vector composed of mutual coupling coefficients, $\boldsymbol{c} = [1, c_1, \cdots, c_{P-1}]$ and satisfies $1 > |c_1| > |c_2| > \cdots > |c_{P-1}|$, P is the length of $\boldsymbol{c}$.

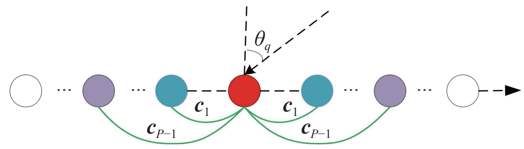

Fig. 1 Array with mutual coupling

Therefore, the received signal under mutual coupling can be represented as

$$\boldsymbol{X}(t) = \tilde{\boldsymbol{A}}\boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{2}$$

where $\tilde{\boldsymbol{A}}$ is the array manifold matrix affected by mutual coupling, $\tilde{\boldsymbol{A}} = [\boldsymbol{C}\boldsymbol{a}(\theta_1), \boldsymbol{C}\boldsymbol{a}(\theta_2), \cdots, \boldsymbol{C}\boldsymbol{a}(\theta_Q)]$. The ideal covariance matrix can be represented as

$$\boldsymbol{R} = \mathrm{E}\left\{\boldsymbol{X}(t)\boldsymbol{X}^{\mathrm{H}}(t)\right\} \tag{3}$$

where $\mathrm{E}\{\cdot\}$ is the mathematical expectation.

When the number of snapshots is limited, the covariance matrix can be expressed as

$$\hat{\boldsymbol{R}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{X}(t)\boldsymbol{X}^{\mathrm{H}}(t) = \boldsymbol{U}_s\boldsymbol{\Sigma}_s\boldsymbol{U}_s^{\mathrm{H}} + \boldsymbol{U}_n\boldsymbol{\Sigma}_n\boldsymbol{U}_n^{\mathrm{H}} \tag{4}$$

where T represents the number of snapshots, $(\cdot)^{\mathrm{H}}$ is the conjugate transpose, $\boldsymbol{\Sigma}_s$ is a diagonal matrix composed of the first Q large eigenvalues, $\boldsymbol{\Sigma}_s = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_Q\}$, and $\boldsymbol{\Sigma}_n$ is a diagonal matrix composed of the last Q+1 to M small eigenvalues, $\boldsymbol{\Sigma}_n = \mathrm{diag}\{\lambda_{Q+1}, \lambda_{Q+2}, \cdots, \lambda_M\}$; $\boldsymbol{U}_s$ represents the signal subspace, $\boldsymbol{U}_s$ is composed of eigenvectors corresponding to the first Q large eigenvalues; $\boldsymbol{U}_n$ represents the noise subspace, $\boldsymbol{U}_n$ is composed of eigenvectors corresponding to the last Q+1 to M small eigenvalues.

When the signal is incoherent, there exists a nonsingular matrix $\boldsymbol{B} \in \mathbf{C}^{Q\times Q}$, and the signal subspace can be represented as

$$\boldsymbol{U}_s = \boldsymbol{CAB} = \tilde{\boldsymbol{A}}\boldsymbol{B}. \tag{5}$$

According to [26], the product of MCM and the incident signal's steering vector can be represented as

$$\boldsymbol{Ca}(\theta_q) = \boldsymbol{\varLambda}(\theta_q)\varOmega(\theta_q)\boldsymbol{\varGamma}(\theta_q) \tag{6}$$

where

$$\boldsymbol{\varLambda}(\theta_q) = \begin{bmatrix} 1 & & & & & & 0 \\ & \beta(\theta_q) & & & & & \\ & & \ddots & & & & \\ & & & \beta(\theta_q)^{P-1} & & & \\ & & & \vdots & & & \\ & & & \beta(\theta_q)^{M-P} & & & \\ & & & & & \ddots & \\ 0 & & & & & & \beta(\theta_q)^{M-1} \end{bmatrix}, \tag{7}$$

$\boldsymbol{\varLambda}(\theta_q) \in \mathbf{C}^{M\times(2P-1)}$, $\boldsymbol{\varLambda}(\theta_q)$ is only related to the DOA of the incident signal and is independent of the mutual coupling coefficients, and

$$\varOmega(\theta_q) = 1 + \sum_{i=1}^{P-1}\left(c_i\beta(\theta_q)^i + c_i\beta(\theta_q)^{-i}\right) \tag{8}$$

where $\varOmega(\theta_q)$ is a constant. In this paper, it is assumed that $\varOmega(\theta_q) \neq 0$, and the expression of $\boldsymbol{\varGamma}(\theta_q)$ is

$$\boldsymbol{\varGamma}(\theta_q) = [\mu_1, \cdots, \mu_{P-1}, 1, \tau_1, \cdots, \tau_{P-1}]^{\mathrm{T}} \tag{9}$$

where $\boldsymbol{\varGamma}(\theta_q) \in \mathbf{C}^{2P-1}$ and

$$\mu_l = \frac{\beta(\theta_q)^{P-1} + \sum_{i=1}^{l-1} c_i\beta(\theta_q)^{P-1-i} + \sum_{i=1}^{P-1} c_i\beta(\theta_q)^{P-1+i}}{\beta(\theta_q)^{P-1} + \sum_{i=1}^{P-1} c_i\beta(\theta_q)^{P-1-i} + \sum_{i=1}^{P-1} c_i\beta(\theta_q)^{P-1+i}}, \tag{10}$$

$$\tau_l = \frac{\beta(\theta_q)^{P-1} + \sum_{i=1}^{P-1} c_i\beta(\theta_q)^{P-1-i} + \sum_{i=1}^{P-1-l} c_i\beta(\theta_q)^{P-1+i}}{\beta(\theta_q)^{P-1} + \sum_{i=1}^{P-1} c_i\beta(\theta_q)^{P-1-i} + \sum_{i=1}^{P-1} c_i\beta(\theta_q)^{P-1+i}}, \tag{11}$$

where $l = 1, 2, \cdots, P-1$.

By substituting the above results into (5), we can obtain that

$$\boldsymbol{U}_s = \tilde{\boldsymbol{A}}\boldsymbol{B} = \hat{\boldsymbol{A}}\boldsymbol{\varDelta}\boldsymbol{B} \tag{12}$$

where $\hat{\boldsymbol{A}} = [\boldsymbol{\varLambda}(\theta_1), \boldsymbol{\varLambda}(\theta_2), \cdots, \boldsymbol{\varLambda}(\theta_Q)] \in \mathbf{C}^{M\times Q(2P-1)}$ can be regarded as a block array manifold matrix, $\boldsymbol{\varLambda}(\theta_q) \in \mathbf{C}^{M\times(2P-1)}$ can be regarded as a block steering matrix that is only related to the DOA and independent of the mutual coupling coefficients, and assuming $\tilde{Q} = 2P-1$, $\boldsymbol{\varLambda}(\theta_q) \in \mathbf{C}^{M\times\tilde{Q}}$, and

$$\boldsymbol{\varDelta} = \begin{bmatrix} \varOmega(\theta_1)\boldsymbol{\varGamma}(\theta_1) & & & 0 \\ & \varOmega(\theta_2)\boldsymbol{\varGamma}(\theta_2) & & \\ & & \ddots & \\ 0 & & & \varOmega(\theta_Q)\boldsymbol{\varGamma}(\theta_Q) \end{bmatrix} \tag{13}$$

where $\boldsymbol{\varDelta} \in \mathbf{C}^{Q\tilde{Q}\times Q}$. Since $\boldsymbol{\varDelta}$ is a block diagonal matrix, therefore

$$\boldsymbol{U}_s = \hat{\boldsymbol{A}}\boldsymbol{\varDelta}\boldsymbol{B} = \hat{\boldsymbol{A}}\hat{\boldsymbol{B}} \tag{14}$$

where $\hat{\boldsymbol{B}} = \boldsymbol{\varDelta}\boldsymbol{B} = [\hat{\boldsymbol{B}}_1^{\mathrm{T}}, \hat{\boldsymbol{B}}_2^{\mathrm{T}}, \cdots, \hat{\boldsymbol{B}}_Q^{\mathrm{T}}]^{\mathrm{T}} \in \mathbf{C}^{Q\tilde{Q}\times Q}$ is the block signal after deformation, $\hat{\boldsymbol{B}}_q = \varOmega(\theta_q)\boldsymbol{\varGamma}(\theta_q) \otimes \boldsymbol{B}_{(q,:)} \in \mathbf{C}^{\tilde{Q}\times Q}$, $q = 1, 2, \cdots, Q$, $\hat{\boldsymbol{B}}_q$ represents the qth block matrix of $\hat{\boldsymbol{B}}$, $\boldsymbol{B}_{(q,:)}$ represents the qth row's elements of the matrix $\boldsymbol{B}$, $\otimes$ is the Kronecker product. Since $\hat{\boldsymbol{A}}$ is only related to the angle, and the DOA of the incident signal in the spatial domain is sparse, a block sparse model about mutual coupling can be constructed according to (14).

For the purpose of DOA estimation, the spatial region is evenly partitioned into N grids, i.e., the spatial domain is $\boldsymbol{\varTheta} = [\theta_1, \theta_2, \cdots, \theta_N]$, the block sparse model about the mutual coupling of $\boldsymbol{U}_s$ can be represented as

$$\boldsymbol{U}_s = \bar{\boldsymbol{A}}\bar{\boldsymbol{\varDelta}}\bar{\boldsymbol{B}}_s = \bar{\boldsymbol{A}}\bar{\boldsymbol{B}} \tag{15}$$

where $\bar{\boldsymbol{A}} = [\boldsymbol{\varLambda}(\theta_1), \boldsymbol{\varLambda}(\theta_2), \cdots, \boldsymbol{\varLambda}(\theta_N)] \in \mathbf{C}^{M\times N(2P-1)}$ is the overcomplete dictionary for the block sparse model, and $\boldsymbol{\varDelta}$ is represented by $\bar{\boldsymbol{\varDelta}}$ in the sparse model, $\bar{\boldsymbol{\varDelta}} \in \mathbf{C}^{N\tilde{Q}\times N}$, and

$$\bar{\boldsymbol{\varDelta}} = \begin{bmatrix} \varOmega(\theta_1)\boldsymbol{\varGamma}(\theta_1) & & & 0 \\ & \varOmega(\theta_2)\boldsymbol{\varGamma}(\theta_2) & & \\ & & \ddots & \\ 0 & & & \varOmega(\theta_N)\boldsymbol{\varGamma}(\theta_N) \end{bmatrix}, \tag{16}$$

$\bar{\boldsymbol{B}}_s$ is the representation of $\boldsymbol{B}$ in the sparse model, and whether the elements contained in the nth row of $\bar{\boldsymbol{B}}_s$ are all 0, which is contingent on the presence of a signal incident from the DOA corresponding to the grid point, $n = 1, 2, \cdots, N$. In addition, $\bar{\boldsymbol{B}} = \bar{\boldsymbol{\varDelta}}\bar{\boldsymbol{B}}_s = [\bar{\boldsymbol{B}}_1^{\mathrm{T}}, \bar{\boldsymbol{B}}_2^{\mathrm{T}}, \cdots, \bar{\boldsymbol{B}}_N^{\mathrm{T}}]^{\mathrm{T}} \in \mathbf{C}^{N\tilde{Q}\times Q}$ is the block sparse model, $\bar{\boldsymbol{B}}_n = \varOmega(\theta_n)\boldsymbol{\varGamma}(\theta_n) \otimes \bar{\boldsymbol{B}}_{s(n,:)} \in \mathbf{C}^{\tilde{Q}\times Q}$ is the nth block matrix of $\bar{\boldsymbol{B}}$, $n = 1, 2, \cdots, N$. Therefore, whether the elements of $\bar{\boldsymbol{B}}_n = \bar{\boldsymbol{B}}((n-1)\tilde{Q}+1, n\tilde{Q}, :)$ are all 0 depends on whether the element of $\bar{\boldsymbol{B}}_{s(n,:)}$ are all 0, where $\bar{\boldsymbol{B}}((n-1)\tilde{Q}+1, n\tilde{Q}, :)$ are the elements from row $(n-1)\tilde{Q}+1$ to row $n\tilde{Q}$ of $\bar{\boldsymbol{B}}$, $\bar{\boldsymbol{B}}_{s(n,:)}$ represents the elements of the nth row of the matrix $\bar{\boldsymbol{B}}_s$.

3. Proposed method

According to (15), $\bar{\boldsymbol{B}}$ is the block sparse signal, and the index n of $\bar{\boldsymbol{B}}$'s block matrix $\bar{\boldsymbol{B}}_n$ corresponds to θ_n. Therefore, after recovering the block sparse signal $\bar{\boldsymbol{B}}$, solve for all $\|\bar{\boldsymbol{B}}_n\|_F$ and find the first Q maximum values to obtain the true DOA estimation. From [10], we know that the most direct sparse recovery method for $\bar{\boldsymbol{B}}$ is to use l_0 norm penalty, i.e.

$$\min \| \bar{\boldsymbol{B}}^{l_F} \|_0 \quad \text{s.t.} \ \| \boldsymbol{U}_s - \bar{\boldsymbol{A}}\bar{\boldsymbol{B}} \|_F^2 \leqslant \xi \tag{17}$$

where $\bar{\boldsymbol{B}}^{l_F} = \left[\|\bar{\boldsymbol{B}}_1\|_F, \|\bar{\boldsymbol{B}}_2\|_F, \cdots, \|\bar{\boldsymbol{B}}_N\|_F\right]^{\mathrm{T}}$, ξ is the threshold parameter, which determines the upper bound of the fitting error. However, the l_0 norm question is an NP-hard problem, which is a challenging and intractable optimization problem. Therefore, the following work focuses on solving the problem with greater efficiency. $\| \bar{\boldsymbol{B}}^{l_F} \|_0$ can be represented as

$$\left\| \bar{\boldsymbol{B}}^{l_F} \right\|_0 = \sum_{n=1}^{N} \mathcal{I}\left(\|\bar{\boldsymbol{B}}_n\|_F\right) \tag{18}$$

where the function expression for $\mathcal{I}(\alpha)$ is

$$\mathcal{I}(\alpha) = \begin{cases} 0, & \alpha = 0 \\ 1, & \text{otherwise} \end{cases} \tag{19}$$

considering the non-smooth function of $\| \bar{\boldsymbol{B}}^{l_F} \|_0$, this paper employs an approximation technique for $\| \bar{\boldsymbol{B}}^{l_F} \|_0$ by utilizing a series of exponential functions. The expression of the exponential function is as follows:

$$f_\sigma(\alpha) = \mathrm{e}^{-\alpha^2/2\sigma^2}, \tag{20}$$

so we can get $\lim_{\sigma\to 0} f_\sigma(\alpha) = 1 - \mathcal{I}(\alpha)$, when $\alpha \neq 0$, and

$$F_\sigma(\bar{B}) = \sum_{n=1}^{N} f_\sigma\left(\|\bar{\boldsymbol{B}}_n\|_F\right) = \sum_{n=1}^{N} \exp\left(-\|\bar{\boldsymbol{B}}_n\|_F^2 / 2\sigma^2\right). \tag{21}$$

So

$$\lim_{\sigma\to 0} f_\sigma(\| \bar{\boldsymbol{B}}_n \|_F) = \begin{cases} 0, & \| \bar{\boldsymbol{B}}_n \|_F \neq 0 \\ 1, & \| \bar{\boldsymbol{B}}_n \|_F = 0 \end{cases}. \tag{22}$$

When $\sigma \to 0$, then there is

$$\lim_{\sigma\to 0} \left\| \bar{\boldsymbol{B}}^{l_F} \right\|_{\mathrm{b}} \approx N - F_\sigma(\bar{\boldsymbol{B}}). \tag{23}$$

According to (21), (17) can be represented as

$$L_\sigma(\bar{\boldsymbol{B}}) = -F_\sigma(\bar{\boldsymbol{B}}) + \lambda \left\| \boldsymbol{U}_s - \bar{\boldsymbol{A}}\bar{\boldsymbol{B}} \right\|_{\mathrm{F}}^2. \tag{24}$$

Our upcoming research will focus on solving (24). By solving this Lagrangian function, we can obtain

$$\bar{\boldsymbol{B}}_*(\sigma) = \arg\min_{\bar{\boldsymbol{B}}} L_\sigma(\bar{\boldsymbol{B}}) \tag{25}$$

where $\bar{\boldsymbol{B}}_*(\sigma)$ is the optimal solution that satisfies $\bar{\boldsymbol{B}}_*(\sigma) = \arg\min_{\bar{B}} L_\sigma(\bar{\boldsymbol{B}})$. The parameter λ regulates the balance between the signal's sparsity and the residual energy [27], which is a constant. The choice of the value of σ will have different effects: the smaller σ, the better approximation of $N - F_\sigma(\bar{\boldsymbol{B}})$ to $\| \bar{\boldsymbol{B}}^{l_F} \|_0$, but there is negative effect of many local minima; in addition, the larger σ, the smoother $N - F_\sigma(\bar{\boldsymbol{B}})$. To regulate this process, a set of gradually decreasing σ is selected until the iteration termination condition is satisfied [28], i.e., the value of σ gradually decreases from large to small. Additionally, Mohimani et al. [28] demonstrated that

$$\lim_{\sigma\to\infty} \bar{\boldsymbol{B}}_*^{(0)}(\sigma) = \bar{\boldsymbol{A}}^{\mathrm{H}}\left(\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^{\mathrm{H}}\right)^{-1} \boldsymbol{U}_s \tag{26}$$

where $\bar{\boldsymbol{B}}_*^{(0)}(\sigma)$ is the optimal initial value in the iterative solution (24).

It should be emphasized that (24) cannot use the Newton method because the Hessian matrix of $L_\sigma(\bar{\boldsymbol{B}})$ cannot always guarantee to be positive definite [29]. In addition, the calculation of $L_\sigma(\bar{\boldsymbol{B}})$'s Hessian matrix is very complex when the block sparse signal $\bar{\boldsymbol{B}}$ is no longer a real-valued column vector. For example, when $\bar{\boldsymbol{B}} \in \mathbf{C}^{N\tilde{Q}\times Q}$, the Hessian matrix dimension of the real-valued scalar $L_\sigma(\bar{\boldsymbol{B}})$ is $2N\tilde{Q}Q \times 2N\tilde{Q}Q$, the computational complexity of the inverse matrix of this Hessian matrix is extremely large. Therefore, this paper develops a Newton-like method based on mapping $\xi(\bar{\boldsymbol{B}})$ to solve (24), which can minimize computational complexity as much as possible.

Lemma 1 Define the mapping $\zeta : \mathbf{C}^{N\tilde{Q}\times Q} \to \mathbf{C}^{N\tilde{Q}\times Q}$, i.e.,

$$\zeta(\bar{\boldsymbol{B}}) = 2\lambda \left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda \bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}} \right]^{-1} \bar{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{U}_s. \tag{27}$$

In reality, $\zeta(\bar{\boldsymbol{B}})$ is the expression for $\bar{\boldsymbol{B}}$ when $\boldsymbol{G}(\bar{\boldsymbol{B}}) = 0$, where $\boldsymbol{G}(\bar{\boldsymbol{B}})$ is the gradient of $L_\sigma(\bar{\boldsymbol{B}})$ at $\bar{\boldsymbol{B}}$:

$$\boldsymbol{G}(\bar{\boldsymbol{B}}) = \frac{\partial L_\sigma(\bar{\boldsymbol{B}})}{\partial \bar{\boldsymbol{B}}} = \frac{1}{\sigma^2} W(\bar{\boldsymbol{B}})\bar{\boldsymbol{B}} - 2\lambda \bar{\boldsymbol{A}}^{\mathrm{H}}(\boldsymbol{U}_s - \bar{\boldsymbol{A}}\bar{\boldsymbol{B}}), \tag{28}$$

the specific derivation process of $\boldsymbol{G}(\bar{\boldsymbol{B}})$ can be found in Proof 1. $\boldsymbol{W}(\bar{\boldsymbol{B}})$ is a diagonal matrix:

$$\boldsymbol{W}(\bar{\boldsymbol{B}}) = \begin{bmatrix} \boldsymbol{W}_1(\bar{\boldsymbol{B}}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{W}_N(\bar{\boldsymbol{B}}) \end{bmatrix} \in \mathbf{R}^{N\tilde{Q}\times N\tilde{Q}} \tag{29}$$

and $\boldsymbol{W}_n(\bar{\boldsymbol{B}}) = f_\sigma(\| \bar{\boldsymbol{B}}_n \|_F)\boldsymbol{I}_{\tilde{Q}}, n = 1, 2, \cdots, N$.

In addition, for any $\bar{\boldsymbol{B}}$, there exists a scalar real number $\kappa \geqslant 0$ that satisfies

$$L_\sigma(\bar{\boldsymbol{B}} + \kappa(\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}})) \leqslant L_\sigma(\bar{\boldsymbol{B}}). \tag{30}$$

In reality, (30) reveals that the value of $L_\sigma(\bar{\boldsymbol{B}})$ along the $\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}}$ direction gradually decreases until the iteration termination condition is satisfied. The rationality of direction $\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}}$ is given in as follows.

Proof 1 According to (24), the Lagrangian function can be obtained

$$L_\sigma(\bar{\boldsymbol{B}}) = -F_\sigma(\bar{\boldsymbol{B}}) + \lambda \| \boldsymbol{U}_s - \bar{\boldsymbol{A}}\bar{\boldsymbol{B}} \|_F^2 \tag{31}$$

and set $L_{1\sigma}(\bar{\boldsymbol{B}}) = F_\sigma(\bar{\boldsymbol{B}})$, $L_{2\sigma}(\bar{\boldsymbol{B}}) = \| \boldsymbol{U}_s - \bar{\boldsymbol{A}}\bar{\boldsymbol{B}} \|_F^2$. Taking the derivative of $\bar{\boldsymbol{B}}$:

$$\frac{\partial L_{1\sigma}(\bar{\boldsymbol{B}})}{\partial \bar{\boldsymbol{B}}[i,j]} = -\frac{f_\sigma(\|\bar{\boldsymbol{B}}((n-1)\tilde{\boldsymbol{Q}}+1, n\tilde{\boldsymbol{Q}}, :)\|_F)}{\sigma^2}\bar{\boldsymbol{B}}[i,j] \quad (32)$$

where $\bar{\boldsymbol{B}}[i,j]$ is the element of the jth column and ith row of $\bar{\boldsymbol{B}}$, where $i \in (n-1)\tilde{Q}+1, \cdots, n\tilde{Q}$, so

$$\frac{\partial L_{1\sigma}(\bar{\boldsymbol{B}})}{\partial \bar{\boldsymbol{B}}} = -\frac{1}{\sigma^2}\boldsymbol{W}(\bar{\boldsymbol{B}})\bar{\boldsymbol{B}}, \quad (33)$$

the specific expression of $\boldsymbol{W}(\bar{\boldsymbol{B}})$ can be found in (29).

$$\frac{\partial L_{2\sigma}(\bar{\boldsymbol{B}})}{\partial \bar{\boldsymbol{B}}} = -2\bar{\boldsymbol{A}}^{\mathrm{H}}(\boldsymbol{U}_s - \bar{\boldsymbol{A}}\bar{\boldsymbol{B}}) \quad (34)$$

Based on the above discussion,

$$\boldsymbol{G}(\bar{\boldsymbol{B}}) = \frac{\partial L_\sigma(\bar{\boldsymbol{B}})}{\partial \bar{\boldsymbol{B}}} = \frac{1}{\sigma^2}\boldsymbol{W}(\bar{\boldsymbol{B}})\bar{\boldsymbol{B}} - 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}(\boldsymbol{U}_s - \bar{\boldsymbol{A}}\bar{\boldsymbol{B}}) \quad (35)$$

where $\boldsymbol{G}(\bar{\boldsymbol{B}})$ is the gradient of $L(\bar{\boldsymbol{B}})$ at $\bar{\boldsymbol{B}}$, considering $\boldsymbol{G}(\bar{\boldsymbol{B}}) = 0$, so we can get

$$\bar{\boldsymbol{B}} = 2\lambda\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]^{-1}\bar{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{U}_s = \zeta(\bar{\boldsymbol{B}}), \quad (36)$$

$\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}}$ can be a descent direction for $L_\sigma(\bar{\boldsymbol{B}})$ to solve $\bar{\boldsymbol{B}}$, that is because

$$\operatorname{tr}\{[\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}}]^{\mathrm{H}}\boldsymbol{G}(\bar{\boldsymbol{B}})\} = \\ -\operatorname{tr}\left\{\boldsymbol{G}^{\mathrm{H}}(\bar{\boldsymbol{B}})\left[\frac{W(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]^{-1}\boldsymbol{G}(\bar{\boldsymbol{B}})\right\}, \quad (37)$$

i.e., $\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}} = -\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]^{-1} G(\bar{\boldsymbol{B}})$.

The specific proof process can be found in Proof 2.

In addition, define the matrix

$$\tilde{\boldsymbol{G}}(\bar{\boldsymbol{B}}) = \boldsymbol{G}^{\mathrm{H}}(\bar{\boldsymbol{B}})\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]^{-1}\boldsymbol{G}(\bar{\boldsymbol{B}}), \quad (38)$$

due to $\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}$ is a positive definite matrix [29], so

$$-\operatorname{tr}\{\tilde{\boldsymbol{G}}(\bar{\boldsymbol{B}})\} < 0. \quad (39)$$

This indicates that the inner product of $\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}}$ and $\boldsymbol{G}(\bar{\boldsymbol{B}})$ is less than 0, so the direction of $\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}}$ is opposite to the direction of the gradient $\boldsymbol{G}(\bar{\boldsymbol{B}})$. Therefore, as it changes along the direction $\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}}$, the value of function $L(\bar{\boldsymbol{B}})$ gradually decreases until it satisfies the threshold for terminating the iteration. □

Proof 2 According to (27), we can obtain

$$\zeta(\bar{\boldsymbol{B}}) = 2\lambda\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]^{-1}\bar{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{U}_s, \quad (40)$$

so that

$$\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]\zeta(\bar{\boldsymbol{B}}) = 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{U}_s. \quad (41)$$

Subtract $\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]\bar{\boldsymbol{B}}$ from both sides of the equation at the same time, then there is

$$\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]\zeta(\bar{\boldsymbol{B}}) - \left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]\bar{\boldsymbol{B}} = \\ 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}U_s - \left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]\bar{\boldsymbol{B}}. \quad (42)$$

According to the expression of $\boldsymbol{G}(\bar{\boldsymbol{B}})$, i.e., (42), then there is

$$\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]\left[\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}}\right] = -\boldsymbol{G}(\bar{\boldsymbol{B}}). \quad (43)$$

So

$$\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}} = -\left[\frac{W(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]^{-1}\boldsymbol{G}(\bar{\boldsymbol{B}}). \quad (44)$$

□

Therefore, by using a backtracking algorithm and designing corresponding step sizes and parameters, the required solution $\bar{\boldsymbol{B}}$ can be obtained. The process of Newton-like method is recorded in Method 1.

Method 1 Iterative block sparse recovery via a Newton-like method

Input $\bar{\boldsymbol{B}}^{(0)} = \bar{\boldsymbol{A}}^{\mathrm{H}}(\bar{\boldsymbol{A}}\bar{\boldsymbol{A}}^{\mathrm{H}})^{-1}\boldsymbol{U}_s$

Output Sparse signal $\bar{\boldsymbol{B}}$

Setting $\sigma_0 = 5\max_n \left\|\bar{\boldsymbol{B}}_n^{(0)}\right\|_F; \lambda; \sigma_{\min}; \rho, \eta, \gamma \in (0,1)$

update

while $\sigma \geqslant \sigma_{\min}$

$\beta = 1$

while $L_\sigma(\bar{\boldsymbol{B}}^{(i)} + \beta(\zeta(\bar{\boldsymbol{B}}^{(i)}) - \bar{\boldsymbol{B}}^{(i)})) > L_\sigma(\bar{\boldsymbol{B}}^{(i)})$

$\beta = \gamma\beta$

end

$\bar{\boldsymbol{B}}^{(i+1)} = \bar{\boldsymbol{B}}^{(i)} + \beta(\zeta(\bar{\boldsymbol{B}}^{(i)}) - \bar{\boldsymbol{B}}^{(i)})$

If $\tau^{(i)} = \left\|\bar{\boldsymbol{B}}^{(i+1)} - \bar{\boldsymbol{B}}^{(i)}\right\|_F < \eta\sigma$ then $\sigma = \rho\sigma$

end

The descent direction $\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}}$ is very similar to the descent direction of the Newton method because

$$\zeta(\bar{\boldsymbol{B}}) - \bar{\boldsymbol{B}} = -\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2} + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]^{-1}\boldsymbol{G}(\bar{\boldsymbol{B}}) \quad (45)$$

where $\boldsymbol{W}(\bar{\boldsymbol{B}})/\sigma^2 + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}$ is a fully-rank positive definite matrix, which can be seen as a Hessian-like matrix. The specific derivation process of (45) can be found in Proof 2. In reality, when $\bar{\boldsymbol{B}}$ is a real-valued column vector, $\boldsymbol{W}(\bar{\boldsymbol{B}})/\sigma^2 + 2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}$ is the Hessian matrix of $L_\sigma(\bar{\boldsymbol{B}})$'s positive definite part [30].

Remark The computational burden of the proposed method comes from $\zeta(\bar{\boldsymbol{B}})$, and the computational complexity of $\zeta(\bar{\boldsymbol{B}})$ mainly comes from the inverse matrix

$\left[\dfrac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2}+2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]^{-1}$, whose computational complexity is $O((N\tilde{Q})^3)$, this computational complexity is enormous.

To reduce it, due to $\boldsymbol{U}_s=\bar{\boldsymbol{A}}\bar{\boldsymbol{B}}$, (45) can be represented as

$$\zeta(\bar{\boldsymbol{B}})-\bar{\boldsymbol{B}}=-\left[\frac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2}+2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]^{-1}\frac{1}{\sigma^2}\boldsymbol{W}(\bar{\boldsymbol{B}})\bar{\boldsymbol{B}}. \quad (46)$$

Define $\tilde{\boldsymbol{Z}}=\left[\dfrac{\boldsymbol{W}(\bar{\boldsymbol{B}})}{\sigma^2}+2\lambda\bar{\boldsymbol{A}}^{\mathrm{H}}\bar{\boldsymbol{A}}\right]$, according to the inverse matrix lemma, $\tilde{\boldsymbol{Z}}^{-1}$ can be represented as

$$\tilde{\boldsymbol{Z}}^{-1}=\sigma^2\boldsymbol{W}^{-1}(\bar{\boldsymbol{B}})-$$
$$\sigma^2\boldsymbol{W}^{-1}(\bar{\boldsymbol{B}})\bar{\boldsymbol{A}}^{\mathrm{H}}\left[\frac{\boldsymbol{I}_M}{2\lambda\sigma^2}+\bar{\boldsymbol{A}}\boldsymbol{W}^{-1}(\bar{\boldsymbol{B}})\bar{\boldsymbol{A}}^{\mathrm{H}}\right]^{-1}\bar{\boldsymbol{A}}\boldsymbol{W}^{-1}(\bar{\boldsymbol{B}}). \quad (47)$$

Therefore, $\zeta(\bar{\boldsymbol{B}})-\bar{\boldsymbol{B}}$ can be represented as

$$\zeta(\bar{\boldsymbol{B}})-\bar{\boldsymbol{B}}=\boldsymbol{W}^{-1}(\bar{\boldsymbol{B}})\bar{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{Z}^{-1}\boldsymbol{U}_s-\bar{\boldsymbol{B}} \quad (48)$$

where $\boldsymbol{Z}=\left[\boldsymbol{I}_M/(2\lambda\sigma^2)+\bar{\boldsymbol{A}}\boldsymbol{W}^{-1}(\bar{\boldsymbol{B}})\bar{\boldsymbol{A}}^{\mathrm{H}}\right]\in\mathbf{C}^{M\times M}$, so the computational complexity of $\boldsymbol{Z}^{-1}$ is $O(M^3)$, the computational complexity of $\bar{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{Z}^{-1}$ is $O((N\tilde{Q})M^2)$; the computational complexity of $\boldsymbol{Z}^{-1}\boldsymbol{U}_s$ is $O(M^2Q)$; the computational complexity of $\bar{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{Z}^{-1}$ is $O((N\tilde{Q})M^2)$; the computational complexity of $\bar{\boldsymbol{A}}\boldsymbol{W}^{-1}(\bar{\boldsymbol{B}})\bar{\boldsymbol{A}}^{\mathrm{H}}$ is $O((N\tilde{Q})M^2)$; the computational complexity of $\bar{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{Z}^{-1}\boldsymbol{U}_s$ is $O((N\tilde{Q})MQ)$. Due to the significantly higher computational complexity of $O((N\tilde{Q})M^2)$ and $O((N\tilde{Q})MQ)$ compared to other terms, thus the proposed method's computational complexity is $O((N\tilde{Q})M^2)+O((N\tilde{Q})MQ)$.

4. Simulation comparison

The primary mission of this section is to verify the performance of the proposed method and compare it with other methods. The selected methods for comparison include the method of sparse representation (SR) [21], the method of block sparse representation (BSR) [22], the method of robust weighted subspace (RWS) [24], focal underdetermined system solver (FOCUSS) [31], the maximum number of iterations for the FOCUSS method is 30, with a termination threshold of 10^{-3} and a norm factor of 1, weighted block sparse DOA estimation based on signal subspace (WBSS) [14], and Cramer-Rao lower bound (CRLB) [32]. In addition, root mean square error (RMSE) is widely used in parameter estimation, and the DOA estimation accuracy of different methods can be compared using RMSE, which is defined as

$$\mathrm{RMSE}_\theta=\sqrt{\frac{1}{\tilde{M}Q}\sum_{n=1}^{\tilde{M}}\sum_{q=1}^{Q}\left(\theta_q-\hat{\theta}_{\tilde{m},q}\right)^2} \quad (49)$$

where $\tilde{M}$ represents the number of Monte Carlo experiments, $\hat{\theta}_{\tilde{m},q}$ denotes the calculated θ_q by the $\tilde{m}$th Monte Carlo experiment. Probability of resolution (PR) can serve as an evaluation criterion for the DOA estimation's stability:

$$\mathrm{PR}=\frac{\sum_{\tilde{m}=1}^{M}P_{\tilde{m}}(\theta)}{\tilde{M}} \quad (50)$$

where the $P_{\tilde{m}}(\theta)$ indicates whether successful estimation of θ_q in the $\tilde{m}$th Monte Carlo experiment. If the DOA estimation of any signal is within the standard range, $P_{\tilde{m}}(\theta)=1$; otherwise, $P_{\tilde{m}}(\theta)=0$. The scenario for $P_{\tilde{m}}(\theta)=1$, i.e., successful DOA estimation, is

$$\left|\theta_q-\hat{\theta}_{\tilde{m},q}\right|\leqslant 0.5^\circ,\ q=1,2,\cdots,Q. \quad (51)$$

The 1st experiment compares the RMSE of all methods under different SNRs. The parameter settings are as follows: assuming two narrowband signals are incident on the array, i.e., Q=2. The incident angles of the signal are $\theta_1=-13.1^\circ$ and $\theta_2=43.1^\circ$. The number of elements in the array is M=10, and the mutual coupling coefficients is $\boldsymbol{c}$ = [1, 0.1545−0.4776i], i.e., P=2. The input SNR ranges from −8 dB to 8 dB, with an interval of 2 dB, and the number of snapshots T=200. The relevant parameters in Method 1 are as follows: $\lambda=2$, $\sigma_{\min}=10^{-4}$, $\gamma=0.3$, $\eta=0.3$, $\rho=0.5$. In the following experiments, unless otherwise specified, all experimental parameters are consistent with the first experiment. As shown in Fig. 2, the RMSE of our proposed method reduces gradually in the range of −8−8 dB and achieves optimal DOA estimation at 4 dB. SR, BSR, and RWS belong to the l_1 norm method, FOCUSS belongs to the reweighted l_2 norm method, and the method proposed in this paper belongs to the l_0 method. After preprocessing (eigenvalue decomposition), the influence of noise is significantly reduced. The simulation results show that in an environment with very little noise, the l_0 method can achieve better performance. Furthermore, the dynamic selection of the step size β ensures that the proposed method exhibits more stable and better performance. WBSS also belongs to the l_1 method, but its performance is slightly better than the proposed method when obtaining prior information, indicating that prior information can effectively assist sparse recovery. However, the computation time of WBSS is very long, as shown in Table 1. It should be noted that the 0.1° RMSE of the DOA is inevitable due to the 0.1° deviation between the incident signal and the set grid points, so the RMSE curve is no longer decreasing and there is a big gap between the CRLB and the RMSE of DOA, and with the increase of SNR, the gap becomes bigger. If we want to achieve more ideal RMSE, we need to set denser grid points. However, the calculation time will typically increase significantly as the grid points become denser because the computational complexity of sparse DOA methods is closely linked to the number of grid points. Please refer to the 8th experiment for the specific calculation time. It should be noted that in the proposed method, the estimation of the number of sources is extremely important because the acquisition of the signal subspace is based on the number of sources. If the estimation of the number of sources is inaccurate, it means that the signal subspace obtained is also inaccurate, and the mathematical model of the signal sparsity model is also incorrect, which has a significant negative impact on DOA estimation.

Table 1 Computational complexity and calculation time

Method	Time/s	Computational complexity
The proposed	0.0656	$O((N\tilde{Q})M^2+(N\tilde{Q})MQ)$
SR	1.6014	$O((N\tilde{Q})^3)$
BSR	1.6860	$O((N\tilde{Q}Q)^3)$
RWS	1.6853	$O((N\tilde{Q})^3)$
FOCUSS	0.0875	$O(M(N\tilde{Q})Q+(N\tilde{Q})M^2)$
WBSS	1.9303	$O((N\tilde{Q}Q)^3+NPM(M-Q))$

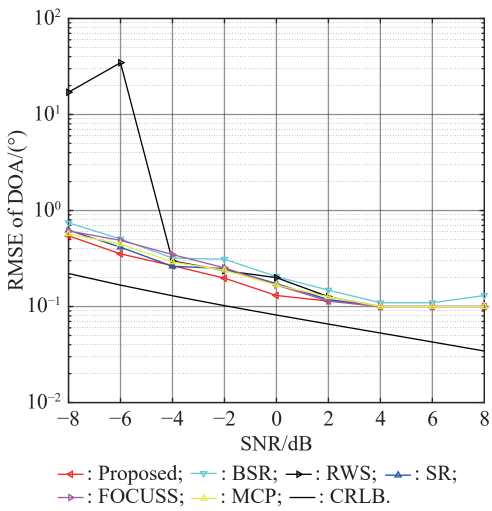

Fig. 2 RMSE of DOA versus SNRs

The 2nd experiment compares the RMSE of all methods under the different snapshots. Assuming that the number of snapshots is 100−1 000, which interval is 100, the input SNR is 1 dB. As shown in Fig. 3, our proposed method performs equally well as the WBSS method and outperforms other algorithms, which achieves optimal DOA estimation within this grid setting at the number of snapshots about 300. As the number of snapshots increases, the RMSE of the DOA estimation of the proposed method becomes better and better. This is because as the number of snapshots increases, the sampling covariance matrix becomes closer to the theoretical covariance matrix, and $\boldsymbol{U}_s$ becomes closer to the theoretical signal subspace. Similar to the first experiment, due to the 0.1° deviation between the incident signal and the set grid points, so the RMSE curve is no longer decreasing when K=300.

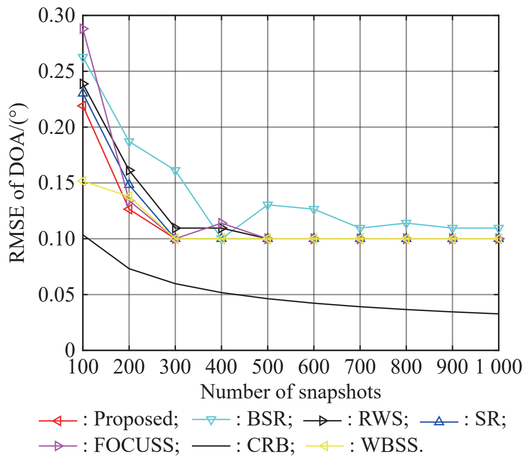

Fig. 3 RMSE of DOA versus the snapshots

The 3rd experiment compares the PR of all methods under the different SNRs. Please refer to the 1st experiment for specific experimental parameters. As shown in Fig. 4, the PR performance of the proposed method improves with increasing input SNR, and in all SNR ranges, except for WBSS, the PR of the proposed method is higher than other methods, this is because the proposed method has the highest PR. In addition, the PR curve of the RWS method exhibits a decreasing trend under high SNR, indirectly suggesting that there is potential for enhancing the stability of this method. Our proposed method achieves a probability of successful DOA estimation of 1 at 4 dB. These phenomena indicate that the proposed DOA estimation method under mutual coupling has better stability under the different SNRs.

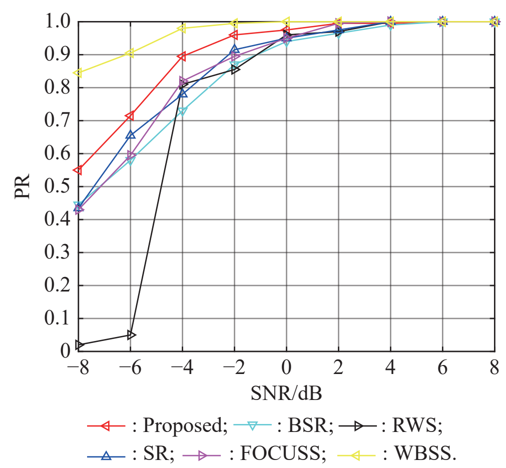

Fig. 4 PR of DOA versus SNRs

The 4th experiment compares the PR of all methods under the different snapshots. Please refer to the 2nd experiment for specific experimental parameters. As shown in Fig. 5, the PR performance of the proposed method improves with the increase in the number of snapshots; moreover, except for WBSS, under any number of snapshots, the proposed DOA estimation method under mutual coupling outperforms other methods in PR, highlighting the DOA estimation stability of the proposed method in different snapshots.

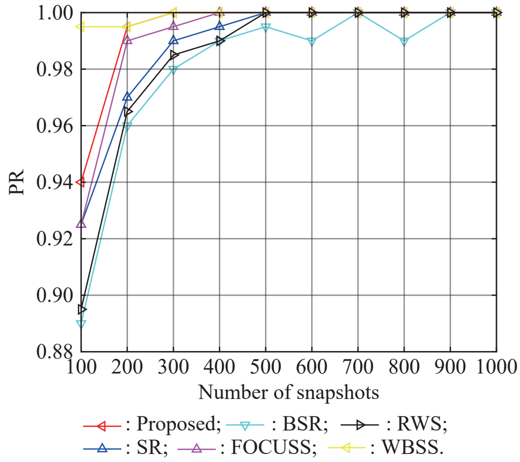

Fig. 5 PR of DOA versus snapshots

The 5th experiment aims to verify the effectiveness of the proposed method under various numbers of mutual coupling coefficients. The number of mutual coupling coefficients is set to P=1, i.e., $\boldsymbol{c}$ = [1]; P=2, i.e., $\boldsymbol{c}$ = [1, 0.1545−0.4776i];P=3,i.e.,$\boldsymbol{c}$=[1,0.1545−0.4776i,0.1345+0.1070i], other parameter settings remain the same as in the 1st experiment. As shown in Fig. 6, when P=1, the performance is relatively optimal compared to P=2 and P=3; When P=3, the performance is relatively poor compared to P=2 and P=1. When SNR=2 dB, the number of mutual coupling coefficients reaches the optimal RMSE at P=1 and P=2. When SNR=8 dB, the number of mutual coupling coefficients approaches the optimal RMSE at P=3.

The 6th experiment compares the estimation performance under different DOAs. Set θ_1 = 5.1°, θ_2 ranging from 14.1° to 26.1°, with an interval of 3°, SNR = 5 dB, and K=200. As shown in Fig. 7, when θ_2=20.1°, the RMSE of θ_1 and θ_2 are both close to 1°; when θ_2=23.1°, the RMSE of θ_1 and θ_2 reaches the optimal RMSE. The simulation shows that the proposed method is more suitable for big DOA intervals.

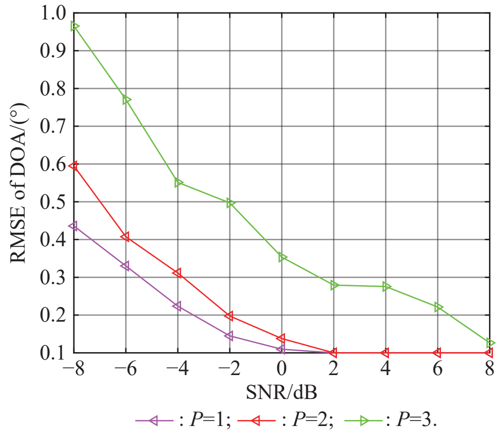

Fig. 6 RMSE of DOA under different numbers of mutual coupling coefficients

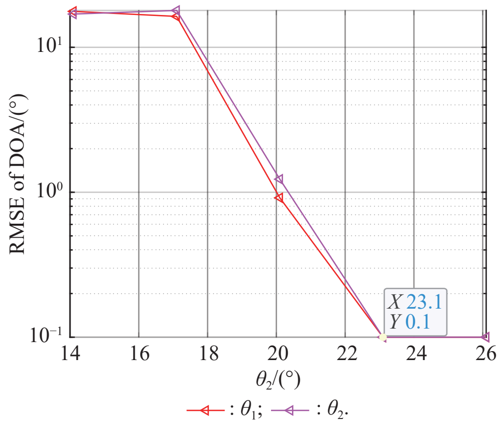

Fig. 7 RMSE under different DOA of incident signals

The 7th experiment compares the computation time of all algorithms. Except for a fixed SNR of 4 dB, all other experimental conditions are consistent with the 1st experiment. The computer settings used are as follows: Intel Core i5-12500, 3.0 GHz processor, 16 GB RAM. As shown in Table 1, the proposed method and FOCUSS are significantly ahead of SR, BSR, and RWS in computational time. This is because SR, BSR, RWS, FOCUSS, and our proposed method require multiple iterations, but our method and FOCUSS have lower computational complexity, resulting in shorter calculation times. The computational complexity of FOCUSS is comparable to that of the proposed method, but the computational speed of FOCUSS is slightly slower than that of the proposed algorithm. This is because the calculation of FOCUSS requires setting corresponding thresholds and maximum iteration times. This experiment shows that under the current parameter settings, the proposed algorithm has higher computational efficiency.

The 8th experiment compares the computation time of different methods under different grid sizes. Fixed

SNR=4 dB, the grid sizes are 0.2°, 0.3°, 0.4°, 0.5°, and 1°, respectively. The other simulation conditions are the same as the 1st experiment. As shown in Fig. 8, even with different grids, the computation time of the proposed method is still comparable to that of FOCUSS, and the proposed method is slightly faster than FOCUSS. Compared to the SR, BSR, RWS, FOCUSS, and WBSS, the proposed method is the fastest.

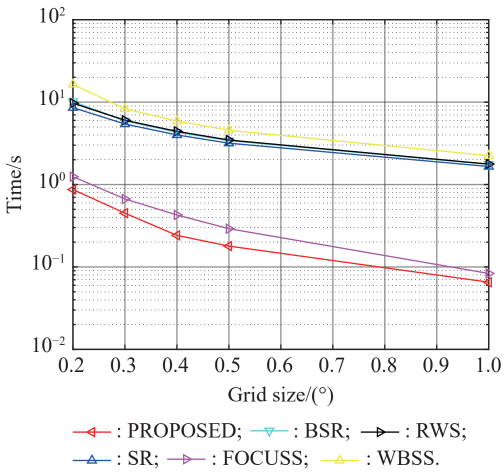

Fig. 8 Computation time versus grid size

5. Conclusions

In this paper, we propose a Newton-like method for recovering sparse signals and achieving DOA estimation under unknown mutual coupling. In the proposed method, a block sparse model about array mutual coupling is established based on the signal subspace. Secondly, a new gradient direction, similar to the Newton's method, is designed based on the mapping $\zeta(\bar{\boldsymbol{B}})$ and uses matrix inverse lemma to reduce its computational complexity. Finally, the advantage of the proposed method in DOA estimation accuracy is demonstrated through simulation.

References

[1] YIN Y T, WANG Y X, DAI T T, et al. DOA estimation based on smoothed sparse reconstruction with time modulated linear arrays. IEEE Trans. on Signal Processing, 2024. DOI: 10.1016/j.sigpro.2023.109229.

[2] RAIGURU P, SUSANTA K S, SAHANI M, et al. Machine learning-aided sparse direction of arrival estimation. IEEE Sensors Journal, 2024, 24 (22): 38125–38134.

[3] JIN J, HE D, SHUANG W, et al. Off-grid DOA estimation method based on sparse Bayesian learning with clustered structural-aware prior in formation. IEEE Trans. on Vehicular Technology, 2024, 73 (4): 5469–5483.

[4] SUN S L, LIU S, WANG J, et al. Joint polarization and DOA estimation based on improved maximum likelihood estimator and performance analysis for conformal array. Journal of Systems Engineering and Electronics, 2023, 34 (6): 1490–1500.

[5] WU X M, YANG Z, WEI Z Q, et al. Direction-of-arrival estimation for constant modulus signals using a structured matrix recovery technique. IEEE Trans. on Wireless Communications, 2024, 23 (4): 3117–3130.

[6] PENG S, CHEN B, YANG M. Joint sparse recovery for direction of arrival based on the generalized music criterion. Digital Signal Processing, 2022, 122: 103382.

[7] LEITE W S, LAMARE R C, ZAKHAROV Y, et al. Direction finding with sparse subarrays: Design, algorithms, and analysis. IEEE Trans. on Aerospace and Electronic Systems, 2024, 60(6): 8149–8165.

[8] REN A K, WU Q, LIANG P Y, et al. Off-grid DOA estimation based on coherent accumulation and weighted block sparse Bayesian. Journal of Systems Engineering and Electronics, 2026, 37(2): 327–336.

[9] HAN K, NEHORAI A. Improved source number detection and direction estimation with nested arrays and using jackknifing. IEEE Trans. on Signal Processing, 2013, 61(23): 6118–6128.

[10] MALIOUTOV D, CETIN M, WILLSKY A A. Sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Trans. on Signal Processing, 2005, 53(8): 3010–3022.

[11] YANG Z, XIE L H, ZHANG C S. Off-grid direction of arrival estimation using sparse Bayesian inference. IEEE Trans. on Signal Processing, 2013, 61(1): 38–43.

[12] CHENG P, CHEN Z M, CAO Z X, et al. A new atomic norm for DOA estimation with gain-phase errors. IEEE Trans. on Signal Processing, 2020, 68: 4293–4396.

[13] WANG X P, WANG W, LIU J, et al. A sparse representation scheme for angle estimation in monostatic MIMO radar. Signal Processing, 2014, 104: 258–263.

[14] LIU Y L, YIN Y Z, LU H M, et al. A novel weighted block sparse DOA estimation based on signal subspace under unknown mutual coupling. Electronics, 2024, 13(9): 1790.

[15] WEISS A, FRIEDLANDER B. Mutual coupling effects on phase-only direction finding. IEEE Trans. on Antennas and Propagation, 1992, 40(5): 535–541.

[16] ZHENG Z, YANG C L. Direction-of-arrival estimation of coherent signals under direction-dependent mutual coupling. IEEE Communications letters, 2021, 25(1): 147–151.

[17] YE Z F, LIU C. On the resiliency of music direction finding against antenna sensor coupling. IEEE Trans. on Antennas and Propagation, 2008, 56(2): 371–380.

[18] TIAN Y, WANG R, CHEN H, et al. Real-valued DOA estimation utilizing enhanced covariance matrix with unknown mutual coupling. IEEE Communications Letters, 2022, 26(4): 912–916.

[19] LIAO B, ZHANG Z G, CHAN S C. DOA estimation and tracking of ULAS with mutual coupling. IEEE Trans. on Aerospace and Electronic Systems, 2012, 48(1): 891–905.

[20] GE Q C, ZHANG Y, WANG Y D. A low complexity algorithm for direction of arrival estimation with direction-dependent mutual coupling. IEEE Communications Letters, 2020, 24(1): 90–94.

[21] DAI J S, ZHAO D, JI F X. A sparse representation method for DOA estimation with unknown mutual coupling. IEEE Antennas and Wireless Propagation Letters, 2012, 11:

1210–1213.

[22] WANG Q, DOU T D, CHEN H, et al. Effective block sparse representation algorithm for DOA estimation with unknown mutual coupling. IEEE Communications Letters, 2017, 21(12): 2622–2625.

[23] WANG X P, MENG D D, HUANG M X, et al. Reweighted regularized sparse recovery for DOA estimation with unknown mutual coupling. IEEE Communications Letters, 2019, 23(2): 290–293.

[24] MENG D D, WANG X P, HUANG M X, et al. Robust weighted subspace fitting for DOA estimation via block sparse recovery. IEEE Communications Letters, 2022, 24(3): 563–567.

[25] MENG D D, LI X, WANG W. Robust sparse recovery based vehicles location estimation in intelligent transportation system. IEEE Trans. on Intelligent Transportation Systems, 2024, 25(1): 1023–1032.

[26] MENG D D, WANG X P, HUANG M X, et al. Block rank sparsity-aware DOA estimation with large-scale arrays in the presence of unknown mutual coupling. Digital Signal Processing, 2019, 94: 96–104.

[27] HYDER M M, MAHATA K. A scalable distributed video coder using compressed sensing. Proc. of the Annual IEEE India Conference, 2009. DOI: 10.1109/INDCON.2009.5409352.

[28] MOHIMANI H, BABAIE M. A fast approach for overcomplete sparse decomposition based on smoothed l0 norm. IEEE Trans. on Signal Processing, 2009, 57(1): 289–301.

[29] HYDER M M, MAHATA K. Direction-of-arrival estimation using a mixed $l_{2,0}$ norm approximation. IEEE Trans. on Signal Processing, 2010, 58(9): 4646–4655.

[30] HYDER M M, MAHATA K. An improved smoothed l_0 approximation algorithm for sparse representation. IEEE Trans. on Signal Processing, 2010, 58(4): 2194–2205.

[31] GORODNITSKY I, RAO B. Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm. IEEE Trans. on Signal Processing, 1997, 45(3): 600–616.

[32] FRIEDLANDER B, WEISS A. Direction finding in the presence of mutual coupling. IEEE Trans. on Antennas and Propagation, 1991, 39(3): 273–284.