DOA Estimation with Sparse Bayesian Learning Using Hierarchical Half-Cauchy Prior with Spectra Refinement Strategy

Zhendong Chen
School of Informatics, Xiamen University, Xiamen, Fujian, China

Xicheng Lu
School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

Yongfeng Huang
Dingzhao Li
School of Informatics, Xiamen University, Xiamen, Fujian, China

Wei Liu, Senior Member, IEEE
Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, HongKong, China.

Shaohua Hong, Member, IEEE
Haixin Sun, Senior Member, IEEE
School of Informatics, Xiamen University, Xiamen, Fujian, China

***Abstract*—Advancements in sparse signal recovery (SSR) theory have significantly enhanced direction-of-arrival (DOA) estimation performance, a critical aspect of array signal processing with a wide range of applications. This paper transforms the conventional complex-valued sparse representation for array-received signals into a real-valued problem for covariance coefficients by leveraging the real-valued nature of source powers. A two-stage hierarchical sparsity-induced prior based on the half-Cauchy framework is proposed, which is approximated using an inverse-Gamma structural prior. Building on this prior, an iterative variational Bayesian inference solution is developed that admits a closed-form expression. Additionally, a thresholding-block-matrix iteration method and a mixing prior updating strategy to exploit spatial domain sparsity are proposed. Simulation results demonstrate that our approach minimizes Gaussian noise impact on spatial spectral partitioning and exhibits robustness at high grid resolutions and signal-to-noise ratios (SNRs) compared to state-of-the-art DOA estimation algorithms. Experiments on the SWellEx-96 dataset further validate the effectiveness of our method in practical environments.**

***Index Terms*—sparse linear array, DOA estimation, covariance fitting, sparse Bayesian learning.**

I. INTRODUCTION

DIRECTION-of-arrival (DOA) estimation is a pivotal topic in array signal processing, with numerous applications in underwater measurements, wireless communications, and remote sensing [1]–[3]. Recent research has extended to more challenging scenarios, including challenging noise environments [4], [5], one-bit quantization [6], and applications involving acoustic vector sensor arrays [7], [8]. Conventional subspace-based methods including multiple signal classification (MUSIC) [9] and estimation of signal parameters via rotational invariance techniques (ESPRIT) [10], were originally developed to overcome the limitations of conventional beamforming based solutions and achieve super-resolution performance. To increase the degrees of freedoms (DOFs), several array designs with non-uniform structures have been proposed [11]–[14]. The construction of virtual sensors is usually based on the structure of the covariance matrix, and DOA estimation can then be achieved using either well-known spatial smoothing techniques [15] or the Toplitz property of an ideal covariance matrix [16], which allows the degrees of freedom (DOFs) of the array to be associated with the configuration of the array, rather than limited by the number of physical sensors. However, all of the above methods assume that the covariance correlations are accurate and therefore require a significant number of snapshots and an adequate signal-to-noise ratio (SNR).

With the development of compressed sensing (CS) theory [17]–[19], a series of DOA estimation algorithms based on sparse signal recovery (SSR) have been proposed [20]–[23]. The majority of the existing SSR algorithms are based on the concept of multiple measurement vectors (MMV). However, identifying the source in an MMV manner introduces an additional computational burden, particularly when a large number of snapshots are involved. To address this challenge, one category concerns constructing the covariance array under a finite number of snapshots, with the aim of obtaining a more desirable covariance matrix. An interpolation-based covariance reconstruction algorithm is proposed based on nuclear norm minimization in [24]. Subsequently, a structured correlation semi-definite programming (SDP) reconstruction problem for matrix reconstruction is established based on structural priors of the array, with Cramér Rao bound (CRB) derived in [25]. The second class of methods employs the vectorized covariance matrix as the obser-

Manuscript received XXXXX 00, 0000; revised XXXXX 00, 0000; accepted XXXXX 00, 0000.
This work is supported by the National Natural Science Foundation of China under Grant 62271426. *(Corresponding author: Wei Liu and Shaohua Hong).*

Zhendong Chen, Yongfeng Huang, Dingzhao Li, Shaohua Hong and Haixin Sun are with the School of Informatics, Xiamen University, Xiamen, Fujian 361005, China, and also with the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Ministry of Education, Xiamen 361005, China (email: {chenzhendong, huangyongfeng, lidingzhao}@stu.xmu.edu.cn; {hongsh, hxsun}@xmu.edu.cn
Xicheng Lu is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK. (email: xicheng.lu@qmul.ac.uk)
Wei Liu is with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, HongKong, China. (email: wei2.liu@polyu.edu.hk)

vation vector to perform linear regression of the single measurement vector (SMV) model. This approach builds upon the theoretical foundation of sparse representation of covariance matrix [26], and is then extended to the reconstruction of sources through atomic norm minimization (ANM) [27], [28]. Nevertheless, both approaches necessitate the construction of a selection matrix to ensure that the atoms align with the Vandermonde structure.

The majority of optimization-based algorithms previously mentioned incur a high computational complexity, and regularization parameters need to be determined through extensive experimentation. In comparison, the implicit regularization-based sparse Bayesian learning (SBL) is capable of attaining a more refined solution through formulation of a structured sparsity induced prior for the latent variables, and achieve incremental updating of the hyper-parameters in an iterative manner, and alignment of the actual data with the structured prior. It was initially introduced in for MMV models [22], [29], and subsequently has a wide range of priors used in SMV models, including Gaussian [30], Laplace [31], and generalized-t [32]. These methods combine the estimation problems of DOA and noise variance using the same prior by stacking them, ignoring the priority of the two estimation problems, and although the effect of noise can be removed by diagonal cropping, it still has high sidelobes at low SNRs. To address this, a thresholding-block-matrix iterative method is proposed, which leverages the sparse structure of the latent variables and refines the selection of grids in the dictionary by partitioning the posterior covariance matrix and approximating the posterior expectation.

Among the three distribution frameworks—Spike-and-Slab (SSL), exponential, and half-Cauchy—the conventional SSL framework is regarded as the gold standard for sparse Bayesian inference. It represents a mixture of a Bernoulli distribution and another, where the probability density function (PDF) exhibits a characteristic Bernoulli spike at zero and follows the other on the positive axis, and thus this class of point-mass hierarchical priors can naturally separate the significant coefficients from the negligible ones [33]. Previous studies have shown that the SSL framework is effective in filtering signal outliers [34], [35]. Recent research has extended the conventional separable SSL model to continuous SSL in order to achieve a faster convergence [36], [37]. Nevertheless, the primary limitation of this approach is its high computational complexity, as the estimation of each atom in an overcomplete dictionary is a computationally demanding process.

The exponential framework, particularly Gaussian [30] and Laplace distributions [31], has been widely used in sparse Bayesian DOA estimation. These distributions impose convex or non-convex regularization on the sparse Bayesian regression process, enhancing sparsity of the solution. However, the Laplace distribution, while suitable for large-scale signal regression, only provides point estimates about parameters. The Gaussian distribution merely applies ℓ_2 regularization to the signal, lacking the ability to promote sparsity [38]. A superior class of elastic net prior with a combination of ℓ_1 and ℓ_2 regularizations, as outlined in [39], provides direct prior distribution results with dual regularizations, which simplifies the process of inference and effectively addresses the degradation in LASSO performance under high sparsity conditions [40].

In contrast to the two aforementioned frameworks, the half-Cauchy prior is capable of naturally constructing highly non-convex penalty terms. The intermediate value of the contraction coefficient of the distribution, corresponding to continuous relaxation of the log-uniform distribution on SSL, has sufficient probabilistic quality to potentially provide superior regularization and generalization [38].

The contributions of this paper are as follows:

- In this paper, we introduce a half-Cauchy framework for sparse Bayesian learning (SBL) based DOA estimation. Building upon the well-known Horseshoe prior, a novel hierarchical half-Cauchy (HHC) prior is proposed.
- An equivalent hierarchical prior within an exponential distribution framework is designed to approximate the proposed HHC prior in a variational Bayesian manner.
- To mitigate the high spatial spectral sidelobes observed under challenging conditions, a thresholding-block-matrix iterative method is proposed.
- Extensive simulations are conducted to demonstrate the superior performance of the proposed algorithm based on simulated data and real data from a non-uniform acoustic array, as part of the SWellEx-96 ocean acoustic experiment.

The remainder of this paper is structured as follows. In Section II, the overcomplete sparse representation array model is presented, while the proposed sparse Bayesian recovery solution is provided in Section III. Simulation results based on simluated data and real data are provided in Sections IV and V, respectively. Conclusion are drawn in Section VI. Notations of this paper are listed in Table I.

II. DATA MODEL

Consider K narrowband far-field signals $s_1, \cdots, s_K$ with wavelength λ impinging on an M-sensor array with directions $\{\theta_k\}_{k=1}^K \in [-90°, 90°]$. The sensor locations are given by the set $\{d_m\}_{m=1}^M$, with $d_1 = 0$. The signals received by the array, $\mathbf{y}(t) = [x_1(t), \cdots, x_M(t)]^T \in \mathbb{C}^{M\times 1}$ is expressed as:

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad t = 1, 2, \ldots, T, \tag{1}$$

where $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{M\times K}$ is the array manifold matrix, and $\mathbf{a}(\theta_k) = [\exp(-j2\pi d_1 \sin\theta_k/\lambda), \cdots, \exp(-j2\pi d_M \sin\theta_k/\lambda)]^T \in \mathbb{C}^{M\times 1}$ is the steering vector for direction θ_k. The noise $\mathbf{n}(t)$ is i.i.d. complex Gaussian with zero mean and variance σ_n^2. All signals are assumed to be i.i.d. complex Gaussian and uncorrelated with noise. Hence,

Table I: List of Notations

Symbol	Description
$\mathbf{A}$, $\mathbf{a}$, a	Matrices, vectors and scalars
$\circ$	The Khatri-Rao product
$\otimes$	The Kronecker product
$\odot$	Element-wise multiplication (The Hadamard product)
$[\mathbf{A}]_{i,:}$, $[\mathbf{A}]_{:,j}$ and $[\mathbf{A}]_{i,j}$	The i-th row, the j-th column and the (i,j)-th element of matrix $\mathbf{A}$
$[\mathbf{a}]_i$	The i-th element of vector $\mathbf{a}$
$\mathbb{C}^{K\times 1}$	A $K\times 1$ complex vector
$\mathbb{R}^{M\times N}$	An $M\times N$ real matrix
$\mathbf{1}_n$	An all-one $n\times 1$ vector
$\mathbf{I}_n$	An $n\times n$ identity matrix
$\mathcal{C}^+$, $\mathcal{N}$, $\mathcal{CN}$ and $\mathcal{IG}$	The half-Cauchy distribution, the normal distribution, the complex Gaussian distribution and the inversed-Gamma distribution
$\Re$ and $\Im$	The real and imaginary parts of a signal
diag$(\cdot)$	The diagonal operator: for a matrix, it extracts the main diagonal as a vector, and for a vector, it creates a diagonal matrix whose diagonal is the vector
$const.$	Constants
Tr$(\cdot)$	The trace of a matrix
$\{\boldsymbol{\alpha}\backslash\boldsymbol{\beta}\}$	The set difference, excluding elements of $\boldsymbol{\beta}$ from $\boldsymbol{\alpha}$
$\langle\cdot\rangle$	The conditional expectation
$\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$	Rounding down and rounding to the nearest integer
$\mathbf{A}^T$, $\mathbf{A}^H$ and $\mathbf{A}^*$	Transposition, conjugate transposition and conjugate of matrix $\mathbf{A}$

the covariance matrix of the received signal is:

$$\begin{aligned}\mathbf{R}_y = \mathbb{E}\{\mathbf{y}(t)\mathbf{y}^H(t)\} &= \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \sigma_n^2\mathbf{I}_M \\ &= \sum_{k=1}^{K}\sigma_k^2\mathbf{a}(\theta_k)\mathbf{a}^H(\theta_k) + \sigma_n^2\mathbf{I}_M,\end{aligned} \tag{2}$$

where σ_k^2 denotes the variance of the k-th source. By vectorizing $\mathbf{R}_y$, we obtain:

$$\mathbf{r} = \text{vec}(\mathbf{R}_y) = (\mathbf{A}^*\circ\mathbf{A})\mathbf{p} + \sigma_n^2\mathbf{1}_M, \tag{3}$$

where $\mathbf{A}^*\circ\mathbf{A} = [\bar{\mathbf{a}}(\theta_1),\cdots,\bar{\mathbf{a}}(\theta_K)]$, $\bar{\mathbf{a}}(\theta) = \mathbf{a}^*(\theta)\otimes\mathbf{a}(\theta)$, $\mathbf{p} = [\sigma_1^2,\sigma_2^2,\ldots,\sigma_K^2]^T$, and $\mathbf{1}_M = \text{vec}(\mathbf{I}_{M\times M})$. For a finite number of snapshots, the covariance matrix is estimated by:

$$\hat{\mathbf{R}}_y = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}(t)\mathbf{y}^H(t), \tag{4}$$

and the corresponding vectorized covariance is:

$$\hat{\mathbf{r}} = (\mathbf{A}^*\circ\mathbf{A})\mathbf{p} + \sigma_n^2\mathbf{1}_M + \boldsymbol{\sigma}, \tag{5}$$

where $\boldsymbol{\sigma}\in\mathbb{C}^{M^2\times 1}$ denotes the perturbation due to the finite number of snapshots. According to the Central Limit Theorem [41], $\boldsymbol{\sigma}$ asymptotically follows a complex Gaussian distribution:

$$\boldsymbol{\sigma}\sim\mathcal{CN}(\mathbf{0}_{M^2},\frac{1}{T}\hat{\mathbf{R}}_y^T\otimes\hat{\mathbf{R}}_y). \tag{6}$$

To eliminate the diagonal elements of the covariance matrix, a selection matrix $\mathbf{J}\in\{0,1\}^{M(M-1)\times M^2}$ is employed [42]. Therefore, the augmented array structure is:

$$\mathbb{D} = \{d_i - d_j \mid i,j\in\{1,2,\ldots,M\},\ i\neq j\}.$$

The modified observation model becomes:

$$\bar{\mathbf{r}} = \mathbf{J}\hat{\mathbf{r}} = \bar{\mathbf{A}}\mathbf{p} + \bar{\boldsymbol{\sigma}}, \tag{7}$$

where $\bar{\mathbf{A}} = \mathbf{J}(\mathbf{A}^*\circ\mathbf{A})$, and $\bar{\boldsymbol{\sigma}}\sim\mathcal{CN}(\mathbf{0}_{M(M-1)},\mathbf{W})$ with $\mathbf{W} = \frac{1}{T}\mathbf{J}[\hat{\mathbf{R}}_x^T\otimes\hat{\mathbf{R}}_x]\mathbf{J}^T$.

To normalize the variance of $\bar{\boldsymbol{\sigma}}$, we pre-multiply both sides of the equation by $\mathbf{W}^{-\frac{1}{2}}$, yielding:

$$\mathbf{r}_w = \mathbf{A}_w\mathbf{p} + \boldsymbol{\sigma}_w, \tag{8}$$

where $\mathbf{r}_w = \mathbf{W}^{-\frac{1}{2}}\bar{\mathbf{r}}$, $\mathbf{A}_w = \mathbf{W}^{-\frac{1}{2}}\bar{\mathbf{A}}$, and $\boldsymbol{\sigma}_w = \mathbf{W}^{-\frac{1}{2}}\bar{\boldsymbol{\sigma}}\sim\mathcal{CN}(\mathbf{0}_{M(M-1)},\mathbf{I}_{M(M-1)})$.

III. SPARSE BAYESIAN RECOVERY

A. Sparse Recovery Modeling

Assume there are N-point sampling grids, $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_1,\cdots,\hat{\theta}_N\}$, uniformly spaced in the DOA range of $[-\pi/2,\pi/2]$, where $N\gg M$. The overcomplete array model is expressed as:

$$\mathbf{r}_w = \mathbf{A}_w\hat{\mathbf{p}} + \boldsymbol{\sigma}_w, \tag{9}$$

where $\mathbf{A}_w = \mathbf{W}^{-\frac{1}{2}}[\bar{\mathbf{a}}(\hat{\theta}_1),\cdots,\bar{\mathbf{a}}(\hat{\theta}_N)]\in\mathbb{C}^{M(M-1)\times N}$, and $\hat{\mathbf{p}}\in\mathbb{R}^{N\times 1}$ is an extension of $\mathbf{p}$, containing only K non-zero elements corresponding to the true DOAs. In practice, due to the finite number of sampling grids, the source angles may not align perfectly with the pre-defined grids. To address this unavoidable off-grid gap, a first-order Taylor expansion-based method, as in [43], is employed:

$$\bar{\mathbf{a}}(\theta_k)\approx\bar{\mathbf{a}}(\hat{\theta}_{n_k}) + (\theta_k - \hat{\theta}_{n_k})\bar{\mathbf{b}}(\hat{\theta}_{n_k}), \tag{10}$$

where $\hat{\theta}_{n_k}$ is the closest grid to θ_k, and $\bar{\mathbf{b}}(\hat{\theta}_{n_k})$ denotes the derivative of $\bar{\mathbf{a}}(\hat{\theta}_{n_k})$ with respect to $\hat{\theta}_{n_k}$. The observation model is then expressed as:

$$\mathbf{r}_w = \boldsymbol{\Phi}_w\hat{\mathbf{p}} + \boldsymbol{\sigma}_w, \tag{11}$$

where $\boldsymbol{\Phi}_w = \mathbf{A}_w + \mathbf{B}_w\text{diag}(\boldsymbol{\beta})$, with $\mathbf{B}_w = \mathbf{W}^{-\frac{1}{2}}[\bar{\mathbf{b}}(\hat{\theta}_1),\cdots,\bar{\mathbf{b}}(\hat{\theta}_N)]\in\mathbb{C}^{M(M-1)\times N}$, and $\boldsymbol{\beta}$ denoting the arcs of the first Taylor expansion. Given that $\hat{\mathbf{p}}$ is a real-valued vector, applying either a complex-valued prior or a real-valued prior will result in a complex-valued regression for $\hat{\mathbf{p}}$. As noted in [30], a real-valued model is preferable:

$$\tilde{\mathbf{r}}_w = \tilde{\boldsymbol{\Phi}}_w\hat{\mathbf{p}} + \tilde{\boldsymbol{\sigma}}_w, \tag{12}$$

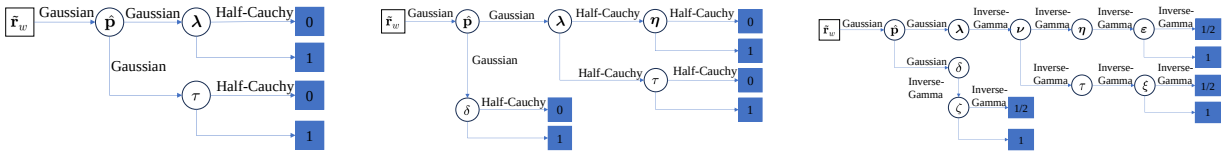
where $\tilde{\mathbf{r}}_w = \begin{bmatrix}\Re(\mathbf{r}_w)\\ \Im(\mathbf{r}_w)\end{bmatrix}$, $\tilde{\boldsymbol{\Phi}}_w = \begin{bmatrix}\Re(\boldsymbol{\Phi}_w)\\ \Im(\boldsymbol{\Phi}_w)\end{bmatrix}$, $\tilde{\boldsymbol{\sigma}}_w = \begin{bmatrix}\Re(\boldsymbol{\sigma}_w)\\ \Im(\boldsymbol{\sigma}_w)\end{bmatrix}$.

B. Sparse Bayesian Learning for DOA Estimation

1. Prior

Given a scale parameter λ with density η, the half-Cauchy distribution is expressed as:

$$p(\lambda|\eta)\sim\mathcal{C}^+(0,\eta) = \frac{2}{\pi\eta\left(1+\frac{\lambda^2}{\eta^2}\right)}. \tag{13}$$



(a) Horseshoe (b) HHC (c) HHC after conversion

Fig. 1. Factor graph of Horseshoe, HHC and HHC after conversion.

It can be observed that the PDF of the half-Cauchy prior exhibits a quadratic decay rate. Due to the divergence at 0, compared to the logarithmic decay rate of the Gaussian prior, the half-Cauchy prior PDF, which concentrates near zero, is theoretically more effective in aggressively shrinking small signals and promoting sparse estimates. Simultaneously, its heavy tails prevent over-shrinking of significant signals [44].

A popular hierarchical shrinkage prior is the Horseshoe prior proposed in [45] for the Gaussian mixture model. Given the observation model from the Gaussian mixture, its hierarchical prior structure is:

$$\begin{aligned} &p(\hat{p}_i|\lambda_i,\tau) \sim \mathcal{N}(0,\lambda_i^2\tau^2), \quad p(\lambda_i) \sim \mathcal{C}^+(0,1), \\ &p(\tau) \sim \mathcal{C}^+(0,1), \quad i=1,\cdots,N, \end{aligned} \tag{14}$$

where τ is the global scale parameter, and λ_i represents the local scale parameters. However, as mentioned in [46], [47], the key to controlling sparsity in the Horseshoe prior lies in the global scale parameter τ, which pulls all variables toward zero. The heavy-tailed local scale parameters λ_i add flexibility by allowing individual coefficients to escape shrinkage, thus highlighting sparse peaks. The standard single-layer Horseshoe prior may be too rigid for complex variable modeling. Therefore, we extend the Horseshoe prior by adding a second layer of half-Cauchy priors to better learn hyperparameters and introduce an additional global shrinkage parameter:

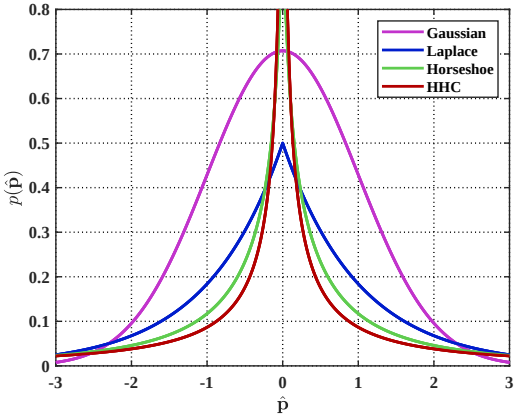
$$\begin{aligned} &p(\hat{p}_i|\lambda_i,\delta) \sim \mathcal{N}(0,\lambda_i^2\delta^2), \\ &p(\lambda_i|\eta_i,\tau) \sim \mathcal{C}^+(0,\eta_i\tau), \quad p(\delta) \sim \mathcal{C}^+(0,1), \\ &p(\eta_i) \sim \mathcal{C}^+(0,1), \quad p(\tau) \sim \mathcal{C}^+(0,1), \quad i=1,\cdots,N, \end{aligned} \tag{15}$$

where λ_i and δ are local and global scale parameters of the first layer, and η_i and τ are those of the second layer. The corresponding factor graph is illustrated in Fig. 1, where white square nodes represent observed variables, circular nodes represent unknown latent variables and hyperparameters, and blue square nodes indicate fixed, predefined hyperparameters.

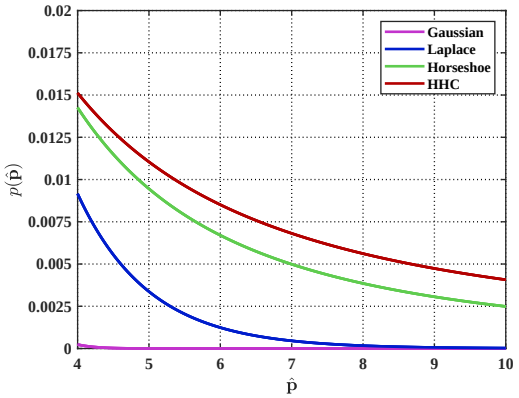
Lemma 1 Let $\delta^2 = 1$ and $\tau^2 = 1$, and then the marginal PDF of the corresponding latent variables is constrained by the following inequality:

$$\frac{1}{\pi^2\sqrt{2\pi}} \ln\left(1+\frac{4}{\hat{p}^2}\right) < p(\hat{p}) \leq \frac{1}{\pi^2|\hat{p}|}. \tag{16}$$

Proof:
The proof of (16) is provided in Appendix A. ■



(a) Local-scale marginal prior densities near the origin.



(b) Local-scale marginal prior densities in the tail region.

Fig. 2. Local-scale PDF comparison between centering and marginal cases.

The PDFs of the HHC, Horseshoe, Gaussian and Laplace on the local scale are shown in Fig. 2, where the proposed prior exhibits the heaviest tail and sharpest peak, implying a good retaining ability for large coefficients and better noise suppression.

2. *Variational Inference*

Due to the complex PDF form of the half-Cauchy prior, it is challenging to directly express its PDF using the expectation maximization (EM) method. Therefore, applying variational inference is a more suitable choice. In variational Bayesian methods [48], the idea is to approximate the true posterior by minimizing the Kullback-Leibler (KL) divergence between the true posterior and a tractable approximate posterior by maximizing the fol-

lowing lower bound of the log-likelihood:

$$\ln p(\mathbf{x};\boldsymbol{\gamma}) = L(q,\boldsymbol{\gamma}) + KL(q||p), \tag{17}$$

where $q(\mathbf{z})$ is the approximate posterior distribution of the latent variable $\mathbf{z}$, $\mathbf{x}$ represents the observed data, and $\boldsymbol{\gamma}$ denotes the fixed parameters associated with the generating latent variables. The term $L(q,\boldsymbol{\gamma}) = \int q(\mathbf{z}) \ln\left\{\frac{p(\mathbf{z},\mathbf{x};\boldsymbol{\gamma})}{q(\mathbf{z})}\right\} d\mathbf{z}$ is the Evidence Lower Bound (ELBO), and $KL(q||p) = -\int q(\mathbf{z}) \ln\left\{\frac{p(\mathbf{z}|\mathbf{x};\boldsymbol{\gamma})}{q(\mathbf{z})}\right\} d\mathbf{z}$ is the KL divergence. Due to the non-negativity of the KL divergence, we have $\ln p(\mathbf{x};\boldsymbol{\gamma}) \geq L(q,\boldsymbol{\gamma})$. Thus, maximizing the ELBO is equivalent to minimizing the KL divergence from $q_j(z_j)$ to $\tilde{p}_j$, i.e.,

$$\max \ln p(\mathbf{x};\boldsymbol{\gamma}) = \max_{\mathbf{z}} L(q,\boldsymbol{\gamma}) = \min_{\forall z_j \in \mathbf{z}} KL(q_j||\tilde{p}_j), \tag{18}$$

where $\tilde{p}_j \triangleq \ln p(\mathbf{x},\mathbf{z};\boldsymbol{\lambda})\rangle_{q_{i\neq j}(z_i)}$ and thus $q_j(z_j)$ can be expressed as:

$$\ln q_j(z_j) = \langle \ln p(\mathbf{x},\mathbf{z};\boldsymbol{\lambda})\rangle_{q_{i\neq j}(z_i)} + const.. \tag{19}$$

Note that the mean-field approximation in variational Bayesian inference assumes that all latent variables in $\mathbf{z}$ are independent, i.e., $q(\mathbf{z}) = \prod_i q_i(z_i)$.

Since the Half-Cauchy prior does not have a conjugate prior structure, we can apply the following mixture representation for the proposed prior, as described in [49]:

$$\begin{aligned} &p(\lambda|\eta) \sim \mathcal{C}^+(0,\eta) \quad \leftrightarrow \quad p(\lambda^2|\chi) \sim \mathcal{IG}\left(\frac{1}{2},\frac{1}{\chi}\right), \\ &p(\chi|\eta^2) \sim \mathcal{IG}\left(\frac{1}{2},\frac{1}{\eta^2}\right). \end{aligned} \tag{20}$$

After some variable substitutions, the hierarchical model in the equivalent exponential framework for the entire regression model is expressed as:

$$\begin{aligned} &p(\tilde{\mathbf{r}}_w|\hat{\mathbf{p}},\tilde{\boldsymbol{\sigma}}_w) \sim \mathcal{N}(\tilde{\boldsymbol{\Phi}}_w\hat{\mathbf{p}}, \sigma_c^2\mathbf{I}_{2M(M-1)\times 2M(M-1)}), \\ &p(\hat{p}_j|\lambda_j,\delta) \sim \mathcal{N}(\hat{p}_j|0,\lambda_j\delta), \\ &p(\lambda_j|\nu_j) \sim \mathcal{IG}\left(\frac{1}{2},\frac{1}{\nu_j}\right), \quad p(\delta|\zeta) \sim \mathcal{IG}\left(\frac{1}{2},\frac{1}{\zeta}\right), \\ &p(\nu_j|\eta_j,\tau) \sim \mathcal{IG}\left(\frac{1}{2},\frac{1}{\eta_j\tau}\right), \quad p(\zeta) \sim \mathcal{IG}\left(\frac{1}{2},1\right), \\ &p(\eta_j|\varepsilon_j) \sim \mathcal{IG}\left(\frac{1}{2},\frac{1}{\varepsilon_j}\right), \quad p(\tau|\xi) \sim \mathcal{IG}\left(\frac{1}{2},\frac{1}{\xi}\right), \\ &p(\varepsilon_j) \sim \mathcal{IG}\left(\frac{1}{2},1\right), \quad p(\xi) \sim \mathcal{IG}\left(\frac{1}{2},1\right), \end{aligned} \tag{21}$$

where $\sigma_c^2 = \frac{1}{2}$ denotes the variance corresponding to covariance matrix. Let $\boldsymbol{\kappa} = \{\hat{\mathbf{p}},\boldsymbol{\lambda},\delta,\boldsymbol{\nu},\zeta,\boldsymbol{\eta},\tau,\boldsymbol{\varepsilon},\xi\}$ represent the set of hyperparameters and latent variables in Eq. (21). From Eq. (19), the iterative updates for each parameter in $\boldsymbol{\kappa}$ are presented as follows:

$$\ln q(\hat{\mathbf{p}}) = \langle \ln p(\tilde{\mathbf{r}}_w,\boldsymbol{\kappa})\rangle_{q_{\{\boldsymbol{\kappa}\setminus\hat{\mathbf{p}}\}}} + const., \tag{22}$$

$$\ln q(\boldsymbol{\lambda}) = \langle \ln p(\tilde{\mathbf{r}}_w,\boldsymbol{\kappa})\rangle_{q_{\{\boldsymbol{\kappa}\setminus\boldsymbol{\lambda}\}}} + const., \tag{23}$$

$$\ln q(\delta) = \langle \ln p(\tilde{\mathbf{r}}_w,\boldsymbol{\kappa})\rangle_{q_{\{\boldsymbol{\kappa}\setminus\delta\}}} + const., \tag{24}$$

$$\ln q(\boldsymbol{\nu}) = \langle \ln p(\tilde{\mathbf{r}}_w,\boldsymbol{\kappa})\rangle_{q_{\{\boldsymbol{\kappa}\setminus\boldsymbol{\nu}\}}} + const., \tag{25}$$

$$\ln q(\zeta) = \langle \ln p(\tilde{\mathbf{r}}_w,\boldsymbol{\kappa})\rangle_{q_{\{\boldsymbol{\kappa}\setminus\zeta\}}} + const., \tag{26}$$

$$\ln q(\boldsymbol{\eta}) = \langle \ln p(\tilde{\mathbf{r}}_w,\boldsymbol{\kappa})\rangle_{q_{\{\boldsymbol{\kappa}\setminus\boldsymbol{\eta}\}}} + const., \tag{27}$$

$$\ln q(\tau) = \langle \ln p(\tilde{\mathbf{r}}_w,\boldsymbol{\kappa})\rangle_{q_{\{\boldsymbol{\kappa}\setminus\tau\}}} + const., \tag{28}$$

$$\ln q(\boldsymbol{\varepsilon}) = \langle \ln p(\tilde{\mathbf{r}}_w,\boldsymbol{\kappa})\rangle_{q_{\{\boldsymbol{\kappa}\setminus\boldsymbol{\varepsilon}\}}} + const., \tag{29}$$

$$\ln q(\xi) = \langle \ln p(\tilde{\mathbf{r}}_w,\boldsymbol{\kappa})\rangle_{q_{\{\boldsymbol{\kappa}\setminus\xi\}}} + const. \tag{30}$$

3. *Variational Inference for Posterior*

For Eq. (22), the expansion yields:

$$\begin{aligned} \ln q(\hat{\mathbf{p}}) &\propto \langle \ln p(\tilde{\mathbf{r}}_w|\hat{\mathbf{p}},\tilde{\boldsymbol{\sigma}}_w) + \ln p(\hat{\mathbf{p}}|\boldsymbol{\lambda},\delta)\rangle_{q_{\{\boldsymbol{\kappa}\setminus\hat{\mathbf{p}}\}}} \\ &\propto -\sigma_c^{-2}(\tilde{\mathbf{r}}_w - \tilde{\boldsymbol{\Phi}}_w\hat{\mathbf{p}})^T(\tilde{\mathbf{r}}_w - \tilde{\boldsymbol{\Phi}}_w\hat{\mathbf{p}}) - \frac{1}{2}\hat{\mathbf{p}}^T\boldsymbol{\Gamma}^{-1}\hat{\mathbf{p}}. \end{aligned} \tag{31}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda})$ and $\boldsymbol{\Gamma} = \delta\boldsymbol{\Lambda}$. It is evident that $q(\hat{\mathbf{p}})$ follows a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, with posterior expectation and covariance given by:

$$\boldsymbol{\mu} = \sigma_c^{-2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Phi}}_w^T\tilde{\mathbf{r}}_w, \tag{32}$$

$$\boldsymbol{\Sigma} = \left(\sigma_c^{-2}\tilde{\boldsymbol{\Phi}}_w^T\tilde{\boldsymbol{\Phi}}_w + \boldsymbol{\Gamma}^{-1}\right)^{-1} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\tilde{\boldsymbol{\Phi}}_w^T\mathbf{D}^{-1}\tilde{\boldsymbol{\Phi}}_w\boldsymbol{\Gamma}, \tag{33}$$

where:

$$\mathbf{D} = \sigma_c^2\mathbf{I}_{M\times M} + \delta\tilde{\boldsymbol{\Phi}}_w\boldsymbol{\Lambda}\tilde{\boldsymbol{\Phi}}_w^T. \tag{34}$$

Under the assumption of sparse latent variables, for most components of $\boldsymbol{\Lambda}$ corresponding to true zeros, the values in the regression process will tend to be small, having little influence on the covariance. During the early stages of Bayesian model iteration, the differences between hyperparameters on the log scale are not significant, resulting in a relatively high learning rate. However, in later stages, the learning rate decreases substantially without significantly affecting the regression results, especially in terms of off-grid estimation. To reduce the sidelobe of the spectra and simplifying the structure of the posterior covariance, we apply a thresholding-block-matrix approximation to $\boldsymbol{\Lambda}$. Ignoring of the auto-correlation of small posteriors, the approximation of $\mathbf{D}$ becomes:

$$\mathbf{D} \approx \mathbf{D}_S = \sigma_c^2\mathbf{I}_{M\times M} + \delta\left[\tilde{\boldsymbol{\Phi}}_w\right]_S\boldsymbol{\Lambda}_S\left[\tilde{\boldsymbol{\Phi}}_w\right]_S^T, \tag{35}$$

where $\boldsymbol{\Lambda}_S = \mathrm{diag}(\boldsymbol{\lambda}_S)$, and the set S is defined as the indices of largest QN values of $\boldsymbol{\lambda}$, with $\boldsymbol{\lambda}_S = [\boldsymbol{\lambda}]_S$ and $\boldsymbol{\lambda}_{S^c} = [\boldsymbol{\lambda}]_{S^c}$; here S^c is the supplementary set of S, and $|S^c|$ and $|S|$ denote the cardinality of the two sets, respectively.

THEOREM 1 *Given a set $F = \{i : i = 1,\ldots,N\}$, with a large-value subset S and its complement S^c, let the selection matrix $\mathbf{G} \in 0,1^{N\times N}$ map from $\boldsymbol{\mu}_A \triangleq \left[\boldsymbol{\mu}_S\ \boldsymbol{\mu}_{S^c}\right]$ to $\boldsymbol{\mu}$ such that $\boldsymbol{\mu} = \mathbf{G}\boldsymbol{\mu}_A$. Then, we have the following approximation:*

$$\boldsymbol{\mu}_A = \sigma_c^{-2}\mathbf{G}\begin{bmatrix}\boldsymbol{\Sigma}_S[\tilde{\boldsymbol{\Phi}}_w]_S^T\tilde{\mathbf{r}}_w \\ \mathbf{0}\end{bmatrix} \tag{36}$$

$$\boldsymbol{\Sigma}_A = \mathbf{G}\begin{bmatrix}\boldsymbol{\Sigma}_S & \boldsymbol{\Sigma}_{Cov} \\ \boldsymbol{\Sigma}_{Cov} & \boldsymbol{\Gamma}_{S^c}\end{bmatrix}\mathbf{G}^T \tag{37}$$

where

$$\boldsymbol{\Sigma}_{Cov} = -\boldsymbol{\Gamma}_S\left[\tilde{\boldsymbol{\Phi}}_w\right]_S^T\mathbf{D}_S^{-1}\left[\tilde{\boldsymbol{\Phi}}w\right]_{S^c}\boldsymbol{\Gamma}_{S^c}. \tag{38}$$

Proof:
Please see Appendix B. ■

For the selection of the dimension of the main dictionary Q, considering the sparsity (or even ultra-sparsity) of the signal in the spatial domain, the maximum effective observation dimension of the covariance matrix is $M(M-1)/2$ due to the Hermitian property (or real symmetry for real-valued covariance matrix). Therefore, to ensure the overcomplete dictionary's efficiency, it is advisable to further reduce its scale. As the iteration progresses, the dictionary dimension is reduced continuously, and the process is stopped when the dimension reaches $M(M-1)\times Q$ or when the hyperparameter iteration threshold converges, where Q represents the pre-defined main dictionary completeness, and *convertiter* denotes the specific iteration to start the spectral refinement.

4. Variational Inference for Hyperparameters

The iterative solutions for each term in Eq. (23)-(30) are given as follows:

$$\lambda_i^{\text{new}} = \langle\lambda_i\rangle = \frac{1}{2}\langle\delta\rangle^{-1}(\mu_i^2+\Sigma_{i,i}) + \langle\nu_i\rangle^{-1}, \tag{39}$$

$$\delta^{\text{new}} = \langle\delta\rangle = \frac{\sum_{i=1}^{N}\langle\lambda_i\rangle^{-1}(\mu_i^2+\Sigma_{i,i}) + 2\langle\zeta\rangle^{-1}}{N+1}, \tag{40}$$

$$\nu_i^{\text{new}} = \langle\nu_i\rangle = \langle\lambda_i\rangle^{-1} + \langle\tau\rangle^{-1}\langle\eta_i\rangle^{-1}, \tag{41}$$

$$\zeta^{\text{new}} = \langle\zeta\rangle = 1 + \langle\delta\rangle^{-1}, \tag{42}$$

$$\eta_i^{\text{new}} = \langle\eta_i\rangle = \langle\nu_i\rangle^{-1}\langle\tau\rangle^{-1} + \langle\varepsilon_i\rangle^{-1}, \tag{43}$$

$$\tau^{\text{new}} = \langle\tau\rangle = \frac{2\langle\xi\rangle^{-1} + 2\sum_{i=1}^{N}\langle\eta_i^{-1}\rangle\langle\nu_i\rangle^{-1}}{N+1}, \tag{44}$$

$$\varepsilon_i^{\text{new}} = \langle\varepsilon_i\rangle = 1 + \langle\eta_i\rangle^{-1}, \tag{45}$$

$$\xi^{\text{new}} = \langle\xi\rangle = 1 + \langle\tau\rangle^{-1}. \tag{46}$$

For updating the external set, we assume the following structured prior distribution:

$$\begin{aligned} p(\hat{\mathbf{p}}_{S^c}|\mathbf{0},\boldsymbol{\iota}) &\sim \prod_{i\in S^c}\mathcal{N}(0,\iota_i),\\ p(\boldsymbol{\iota}|a,b) &\sim \mathcal{IG}(\boldsymbol{\iota}|a,b), \quad a,b\to 0, \end{aligned} \tag{47}$$

where $\boldsymbol{\iota} \triangleq \boldsymbol{\lambda}_{S^c}$.

Thus, the corresponding variational update equation can be written as:

$$\begin{aligned} &\ln q(\boldsymbol{\iota}) \propto \langle \ln p(\hat{\mathbf{p}}_{S^c}|\mathbf{0},\boldsymbol{\iota}) + \ln p(\boldsymbol{\iota}|a,b)\rangle_{q_{\hat{\mathbf{p}}}(\hat{\mathbf{p}})}\\ &= \sum_{i\in\{S^c\}} -(a+\frac{1}{2}-1)\ln\iota_i - \iota^{-1}\left[\frac{1}{2}(\mu_i^2+\Sigma_{i,i})+b\right], \end{aligned} \tag{48}$$

From the prior distribution of $\boldsymbol{\iota}$, the posterior follows the distribution:

$$q(\boldsymbol{\iota}) = \prod_{i\in S^c}\mathcal{IG}(\iota_i|\tilde{u},\tilde{v}), \tag{49}$$

where:

$$\tilde{u} = a+\frac{1}{2}, \quad \tilde{v}_i \quad = \frac{1}{2}(\mu_i^2+\Sigma_{i,i}) + b. \tag{50}$$

The corresponding variational update is:

$$\iota_i^{\text{new}} = \langle\iota_i\rangle = \frac{\mu_i^2+\Sigma_{i,i}+2b}{2a+1}, \tag{51}$$

where $a,b\to 0$.

Note that in Eqs. (40) and (44), we use a fixed number of all grids N instead of $|S|$ that varies with the number of main dictionary grids. Specifically, in the proposed algorithm, the global scale parameter of the prior mainly affects the suppression of small signals at the main dictionary, while the proposed HHC prior is applied to the constructed main dictionary to perform sparse regression on the corresponding grids to capture the potential signal directions. The approximation of the posterior expectation further combines with the Gaussian prior corresponding to the sub-dictionary to change the trend of variance-related hyperparameters corresponding to the grids of the sub-dictionary, thus making even most of the sub-dictionary's grid converge to zero. In order to reach a smooth iteration, or to get a more substantial spatial power spectrum in the logarithmic domain, we can impose a constraint to apply a correction to the grid points corresponding to the sub-dictionary to keep it nonzero-valued.

C. Correction of the Model Error

In practice, it is challenging to perfectly align the predefined grid points with true DOAs. Due to the high computational cost of CS-based algorithms, increasing the grid density to improve the Bayesian estimation is impractical. Therefore, based on the first-order Taylor expansion in Eq. (12), we consider the model with off-grid corrections. The corresponding posterior is given as:

$$\begin{aligned} \langle\ln p(\tilde{\mathbf{r}}_w|\hat{\mathbf{p}},\tilde{\boldsymbol{\sigma}}_w)\rangle_{q(\hat{\mathbf{p}})} &\propto -\left\langle[\tilde{\mathbf{r}}_w-\tilde{\boldsymbol{\Phi}}_w\hat{\mathbf{p}}]^T[\tilde{\mathbf{r}}_w-\tilde{\boldsymbol{\Phi}}_w\hat{\mathbf{p}}]\right\rangle_{q(\hat{\mathbf{p}})}\\ &\propto -||\tilde{\mathbf{r}}_w-\tilde{\boldsymbol{\Phi}}_w\boldsymbol{\mu}_A||_2^2\\ &\quad -\mathrm{Tr}\left\{\tilde{\boldsymbol{\Phi}}_w\boldsymbol{\Sigma}_A\tilde{\boldsymbol{\Phi}}_w^T\right\}, \end{aligned} \tag{52}$$

where:

$$\mathbf{P} = \tilde{\mathbf{B}}_w^T\tilde{\mathbf{B}}_w \odot (\boldsymbol{\mu}_A\boldsymbol{\mu}_A^T+\boldsymbol{\Sigma}_A), \tag{53}$$

$$\mathbf{v} = \mathrm{diag}(\boldsymbol{\mu}_A)\tilde{\mathbf{B}}_w^T(\tilde{\mathbf{r}}_w-\tilde{\mathbf{A}}_w\boldsymbol{\mu}_A). \tag{54}$$

Thus, we have:

$$\tilde{\boldsymbol{\beta}} = \mathbf{P}^{-1}\mathbf{v}, \tag{55}$$

and if $\mathbf{P}$ is non-invertible:

$$\tilde{\beta}_i = \frac{v_i}{p_{i,i}}. \tag{56}$$

To prevent abnormal updates in angle correction, we impose a threshold constraint:

$$\beta_i^{\text{new}} = \begin{cases} \beta_i^{\text{last}}-\frac{1}{2}\Delta, & \tilde{\beta}_i^{\text{new}} \le \beta_i^{\text{last}}-\frac{1}{2}\Delta,\\ \tilde{\beta}_i^{\text{new}}, & \beta_i^{\text{last}}-\frac{1}{2}\Delta < \tilde{\beta}_i^{\text{new}} < \beta_i^{\text{last}}+\frac{1}{2}\Delta,\\ \beta_i^{\text{last}}+\frac{1}{2}\Delta, & \tilde{\beta}_i^{\text{new}} \ge \beta_i^{\text{last}}+\frac{1}{2}\Delta, \end{cases} \tag{57}$$

the update of the angle set is given by:

$$\boldsymbol{\Theta}^{\text{new}} = \boldsymbol{\Theta}^{\text{old}} + \boldsymbol{\beta}^{\text{new}}. \tag{58}$$

The iterative process solves Eqs. (32), (33), (39)-(46), (53)-(58) until convergence when the variance hyperparameters satisfy $||\boldsymbol{\lambda}_S^{\text{new}} - \boldsymbol{\lambda}_S^{\text{last}}||_2/||\boldsymbol{\lambda}_S^{\text{last}}||_2 < \text{tol}$.

The overall iterative process of the proposed algorithm is summarized in Algorithm 1.

Algorithm 1 The Proposed HHC-VBI Algorithm

Input: $\hat{\mathbf{R}}_y$, Q, N, tol, $maxiter$, $convertiter$.
Output: $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Theta}$.
Initialize:
Vectorize $\hat{\mathbf{R}}_y$, trim, whiten, and regroup the imaginary and real parts of the vector into a new vector.
Initialize hyperparameters: $\delta^{(0)} = \zeta^{(0)} = \tau^{(0)} = \xi^{(0)} = 1$, and for $i = 1, \ldots, N$, set $\lambda_i^{(0)} = \nu_i^{(0)} = \eta_i^{(0)} = \varepsilon_i^{(0)} = 1$; set S^c as an empty set.
Set $converged$ = $False$, and iteration counter $iter$ = 0.
while $converged$ is $False$ **and** $iter < maxiter$ **do**
  if $iter < convertiter$ **then**
    Compute the posterior using Eqs. (32) and (33).
  else
    if $|S^c| < \max(N - QM(M-1), 0)$ **then**
      $|S^c|$++.
    end if
    Set the index of smallest $|S^c|$ parts of $\boldsymbol{\lambda}$ as the corresponding set S^c, and the rest as the set S.
    Average the numbers corresponding to the set S^c in $\boldsymbol{\iota}$
    Compute the posterior using Eqs. (36)-(38).
  end if
  Update hyperparameters for set S using Eqs. (39)-(46).
  Update hyperparameters for set S^c using Eq. (51).
  Perform angle correction using Eqs. (53)-(58).
  if $||\boldsymbol{\lambda}_S^{\text{new}} - \boldsymbol{\lambda}_S^{\text{last}}||_2/||\boldsymbol{\lambda}_S^{\text{last}}||_2 < \text{tol}$ **then**
    Set $converged$ = $True$.
  end if
  $iter$++.
end while

D. Complexity Analysis

Although the proposed algorithm has many hyperparameters, the iterative update expressions of all the hyperparameters are only involved in the inverse of the scalar. Therefore, the computational complexity of this part can be deemed negligible. We proceed to discuss the pre-conversion and post-conversion cases and calculate the computational complexity of their single iterative process, respectively.

1. The Pre-Conversion Process

In the pre-conversion process, the computational complexity is primarily determined by the inverse of the covariance matrix of the posterior with $O(N^3)$ and $O(2NM(M-1))$ for the posterior of the expectation, represented by $O(6N+2)$ for the variational inference of the hyperparametric part, and by $O(KM(M-1))$ for the parametric inference of the off-grid part. Consequently, the total complexity for a single iteration is $O(N^3 + 2NM(M-1) + 6N + 2 + KM(M-1))$.

2. The Conversed Process

The computational complexity of the inference of the posterior is $O(|S|^3 + 4|S|M^2(M-1)^2 + 2|S||S^c|M(M-1) + 2|S|^2M(M-1))$, and the complexity of the variational inference of the hyperparameter is reduced to $O(6|S| + 2 + 2|S^c|)$, while the off-grid part remains unchanged. Consequently, the total computational complexity is $O(|S|^3 + 4|S|M^2(M-1)^2 + 2|S||S^c|M(M-1) + 2|S|^2M(M-1) + 6|S| + 2 + 2|S^c| + KM(M-1))$.

IV. SIMULATIONS AND DISCUSSIONS

In this section, a series of experiments are carried out to verify the performance of the proposed algorithm in different scenarios. Seven existing algorithms are selected for comparison, including SS-MUSIC [50], Real-VBI [30], untrimmed-SBL [43], ℓ_1-SRACV [26], CSPICE [51], OGSBI [22] and JSBL [52].

A. Simulation with Comparative Methods

1. Simulation Setup

Both the co-prime array (CPA) and the nested array (NA) are considered. The performance metric for evaluation is the root mean square error (RMSE), defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N_{MC}K}\sum_{n_{MC}=1}^{N_{MC}}\sum_{k=1}^{K}\left(\check{\theta}_{n_{MC},k} - \theta_{n_{MC},k}\right)^2}, \tag{59}$$

where N_{MC} represents the number of Monte Carlo simulations and $\check{\theta}_{n_{MC},k}$ denotes the estimated angle for the k-th source in the n_{MC}-th Monte Carlo trial. The detailed simulation settings are listed in Table II.

In the case of on-grid methods e.g. ℓ_1-SRACV, a one-step grid refinement method is devised in order to determine the direction. Once the solution $\mathbf{p}$ has been obtained for each overcomplete grid, the following steps are undertaken:

$$\mathbf{R}_{over} = \bar{\mathbf{A}}\mathbf{p}\bar{\mathbf{A}}^H. \tag{60}$$

Subsequently, the spatial spectrum of the constructed covariance array is reconstructed with a resolution of 0.01°, and the resulting spectral peaks are taken to represent the estimated angles.

2. RMSE v.s. SNR

The SNR is selected between -10dB and 18dB, with an increment of 2dB. The simulation results are presented in Fig. 3, which shows that the proposed algorithm achieves the lowest RMSE ranging from -10 dB to 18 dB.

Table II: Parameter settings in simulations

Parameters	Values
Monte Carlo trials	500
Hyperparameters initialization	1
Array configurations	A CPA with $M_1 = 3$, $M_2 = 5$ and an NA with $M_1 = 3$, $M_2 = 4$.
Q	1
$convertiter$	200
$maxiter$	1000
tol	1e-5
SNR	4dB
Number of snapshots	200
Directions of incident signals	$[-31°, -5°, 45°]$ with a standard normal perturbation.
Resolution	0.01°(SS-MUSIC) 2°(other methods)
Ratio of final angular search steps for [52]	1000

In comparison to the proposed algorithm, untrimmed-SBL frequently achieve the iterative maximum at high SNR within the context of CPA and NA.

3. RMSE v.s. Number of Snapshots

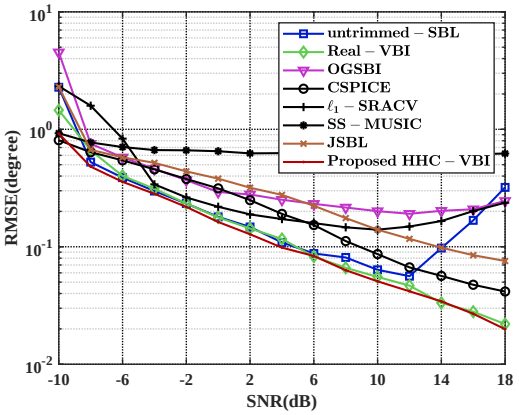
The number of snapshots is considered in steps of 50 between 50 and 500. As illustrated in Fig. 4, all algorithms exhibit enhanced performance as the number of snapshots increases. The proposed algorithm with Real-VBI achieves the best one, while the performance of untrimmed-SBL is found to be inferior.

4. RMSE v.s. Number of Sources

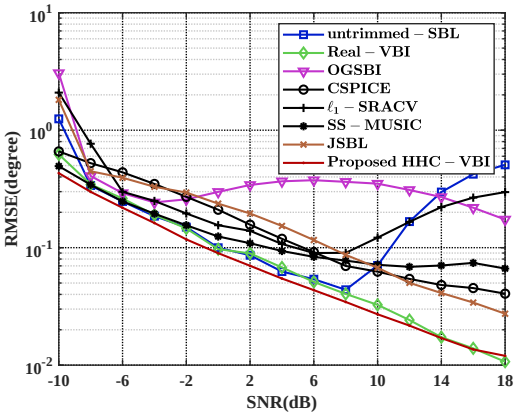
Now consider arrays of larger dimensions, wherethe augmented CPA of $M_1 = 4$, $M_2 = 5$ with 12 sensors, and NA of $M_1 = M_2 = 6$ are employed. To ensure completeness of the dictionary of steering vectors, the resolution is set to $r = 1°$. The set of incident directions is uniformly distributed between $[-60°, 60°]$. In addition, a perturbation that obeys the Gaussian distribution is applied to each incident angle to ensure randomness of the experiment. The number of incident sources ranges between 1 and 16 for the Monte Carlo experiments. The results are presented in Fig. 5. The proposed algorithm has the best performance for all source number conditions. The sparse Bayesian-like algorithm exhibits robust performance for both ultra-sparse and sparse signals, where the RMSE of the proposed algorithm, OGSBI, untrimmed-SBL, and Real-VBI, increases gradually with the source number.

5. RMSE v.s. Resolution

The experimental conditions are analogous to those described earlier. Spacing of grids $r = 0.5°, 1°, 1.5°, 2°, 2.5°, 3°, 4°, 5°$ and $6°$ are chosen to further validate the performance. The simulation results are presented in Fig. 6. The proposed algorithm demonstrates superior estimation capabilities in both overcomplete and undercomplete dictionaries.



(a) RMSE of DOA estimation versus SNR for CPA at $M_1 = 3, M_2 = 5$.



(b) RMSE of DOA estimation versus SNR for NA at $M_1 = 3, M_2 = 4$

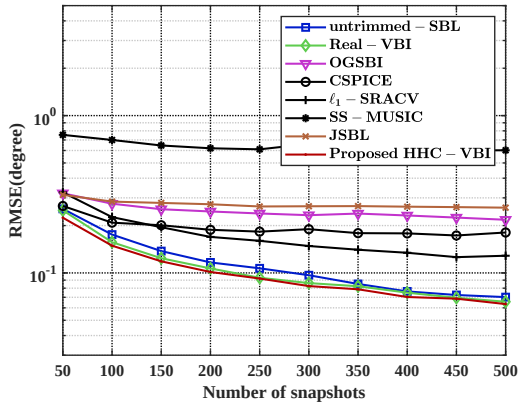
Fig. 3. Influence of SNR for CPA and NA at $T = 200$, $r = 2°$, $Q = 1$ and $convertiter = 200$.

6. RMSE v.s. Numbers of Iterations

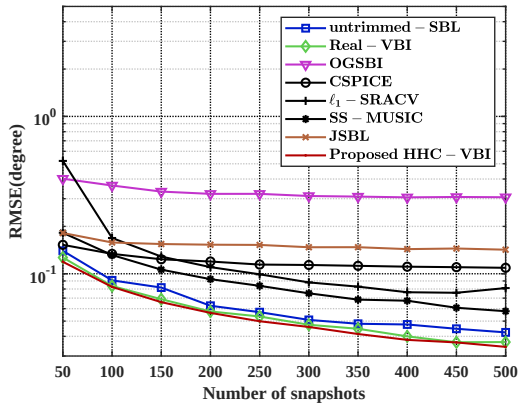
The number of iterations is changed between 10 and 200, with a step size interval of 10. The experimental results are presented in Fig. 7. It can be seen that the proposed algorithms can converge after approximately 80 iterations in both NA and CPA scenarios. Furthermore, the evolution of corresponding normalized spatial power spectrum of the proposed algorithm is also examined as the number of iterations increases. In the comparison experiments, $convertiter = 100$, $Q = 1$, and the number of iterations $maxiter$ = 100, 150, 200, 250, and 300. The normalized spatial power spectrum for different iterations are shown in Fig. 8, which demonstrate that the proposed algorithm with a mixing prior iteration strategy effectively reduces the amplitude of the sidelobe with increase in the number of iterations.

B. Relationship between Thresholds of the Main Dictionary and RMSE

This subsection demostrates the impact of completeness of the main dictionary of the proposed algorithm

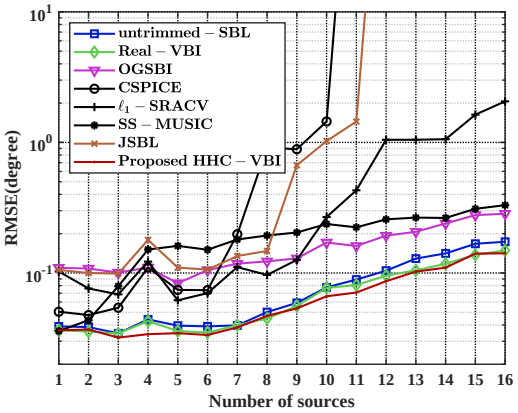


(a) RMSE of DOA estimation versus number of snapshots for CPA at $M_1 = 3, M_2 = 5$.

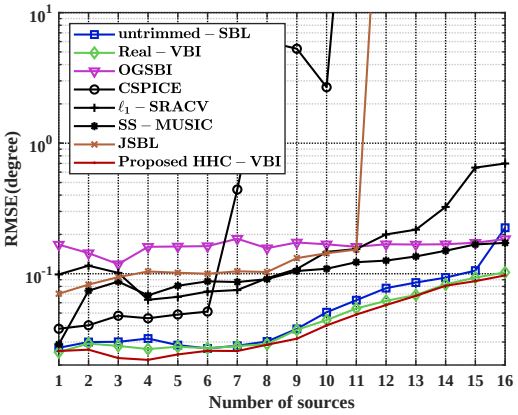


(b) RMSE of DOA estimation versus number of snapshots for NA at $M_1 = 3, M_2 = 4$.

Fig. 4. Influence of number of snapshots for CPA and NA at $T = 200$, $r = 2°$, $Q = 1$ and $convertiter = 200$.



(a) RMSE of DOA estimation versus number of sources for augment CPA at $M_1 = 4, M_2 = 5$.



(b) RMSE of DOA estimation versus number of sources for NA at $M_1 = 6, M_2 = 6$.

Fig. 5. Influence of number of sources for augment CPA and NA at $T = 200$, $r = 2°$, $Q = 1$ and $convertiter = 200$.

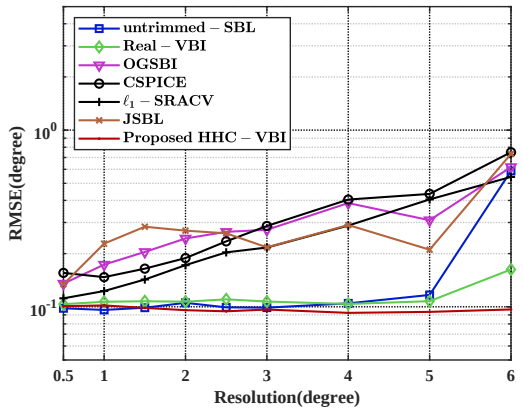
on the performance with different grid resolutions and different array types. Fig. 9 presents the performance under varying degrees of main dictionary completeness of $Q = 0.5$, 1, 1.5 and 2, and grid resolutions of 0.5°, 1°, 1.5° and 2°. Similarly, we further verify the effect of the scale of S^c in S on power spectrum estimation for the same number of iterations. As shown in Fig. 10, the power spectrum sidelobe of the proposed algorithm decreases as the size of S^c increases, which is a consequence of the robustness of the proposed prior on an undercomplete dictionary, and the spatial spectrum on both CPA and NA reach a low amplitude except fot the part near the direction of the souces, at around $Q = 0.5$. In order to ensure convergence of the algorithm, we do not consider it below $Q = 0.5$.

V. EXPERIMENT WITH SWELLEX-96 DATASET

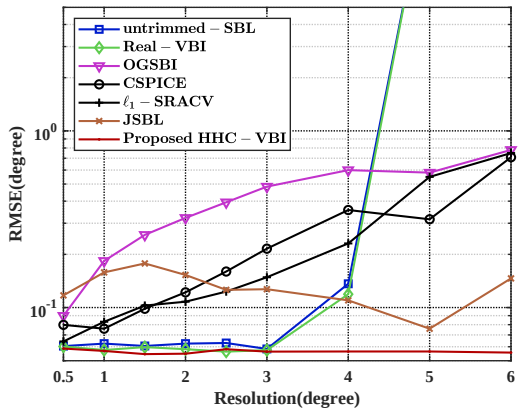
This section examines the performance of the proposed algorithm in comparison with others in a real marine environment, using the SWellEx-96 underwater acoustic dataset [53].

A. Dataset

The SWellEx-96 experiment was conducted on 10-18 May, 1996, approximately 12 km from the tip of Point Loma near San Diego, California. During the conduct of Event S5 which lasting 75 minutes, the towed sound source passes sequentially through the HLA South, HLA North, VLA and TLA, transmitted various sources between 49 Hz and 400 Hz. The GPS recordings for single-tone towed sound sources were obtained at a frequency of every one minute. The two horizontal arrays (HLA North and South) recorded only the last 50 minutes. For the corresponding array for DOA estimation, HLA South was employed which is a 32-element horizontal non-uniform array deployed at a depth of 198 metres underwater with a sampling frequency of 3276.8 Hz and a total array aperture of 255 m [53]. In order to construct sparse arrays, we utilize array elements 1, 2, 3, 4, 7, 8, 9 and 10 of the HLA-South array for DOA estimation. The east incidence is defined as the absolute direction of 0°. The positive of the absolute and relative directions is considered to be counter-clockwise. The incidence direction of the array is

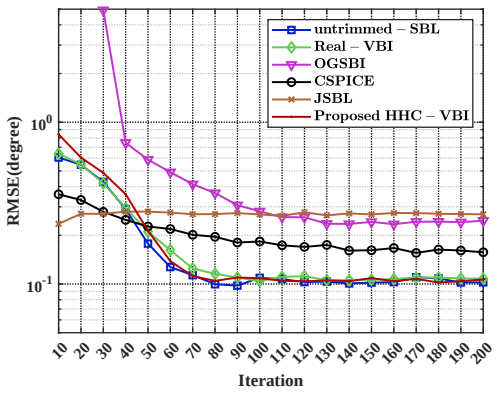


(a) RMSE of DOA estimation versus resolution for initial grid for CPA at $M_1 = 3, M_2 = 5$.

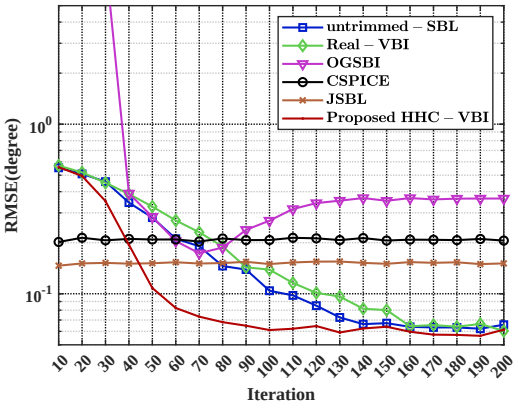


(b) RMSE of DOA estimation versus resolution for initial grid for NA at $M_1 = 3, M_2 = 4$.

Fig. 6. Influence of resolution for initial grid for CPA and NA at $T = 200$, $r = 2°$, $Q = 1$ and $convertiter = 200$.



(a) RMSE of DOA estimation versus max iterations for CPA at $M_1 = 3, M_2 = 5$.



(b) RMSE of DOA estimation versus max iterations for NA at $M_1 = 3, M_2 = 4$.

Fig. 7. Influence of max iterations for CPA and NA at $T = 200$, $r = 2°$, $Q = 1$ and $convertiter = 200$.

between -90° and 90°. Consequently, the relative direction of -90° of the array is defined as the incidence direction of the 9th array element to the 1st array element, which is 137.28°. Additionally, the first array element is identified as the location of the GPA record for HLA South.

B. DOA Estimation Results

For DOA estimation, we take the single-tone signal with a frequency of $f_0 = 148$Hz, a 50% overlap rate for FFT with a dimension of $N_{FFT} = 819$, and a number of snapshots of $P = 99$. The sound speed is 1490m/s. We compare the proposed method with CBF, Minimum Variance Distortionless Response (MVDR) [54], OGSBI [22], untrimmed-SBL [43] Real-VBI [30] and JSBL [52]. Each frame of DOA estimation corresponds to a time interval of $N_{FFT} \times (P+1)/2 \times 0.5/f_s \approx 6.24s$. The parameters are listed in Table III. The trajectory of the whole event and DOA estimation comparision are presented in Figs. 11 and 12, where the solid red line and the solid black vertical line indicate the true trajectory and direction, respectively. It can be observed that during the initial 15 minutes, all the methods exhibit considerable divergence

Table III: Parameter settings in the experiment

Parameters	**Values**
Hyperparameters Initialization	1
$convertiter$	100
Q	1
$maxiter$	500
tol	1e-4
Resolution	2°(Bayesian methods) 0.01°(CBF and MVDR)
Ratio of final angular search steps for [52]	1000

from the actual trajectory of the signal, predominantly due to the Doppler effect and the deterioration in direction-finding capabilities in the direction of the array end-fire regions, and at a later stage, the proposed method gives a clear and precise trajectory map. Furthermore, in order to ascertain the efficacy and computational efficiency of the proposed algorithm in comparison with the other algorithms, the average absolute error (as defined in (61)) for the final 30 minutes of real data, which was not affected by the Doppler effect, is presented in Table IV. When the estimated time point does not align with the time point

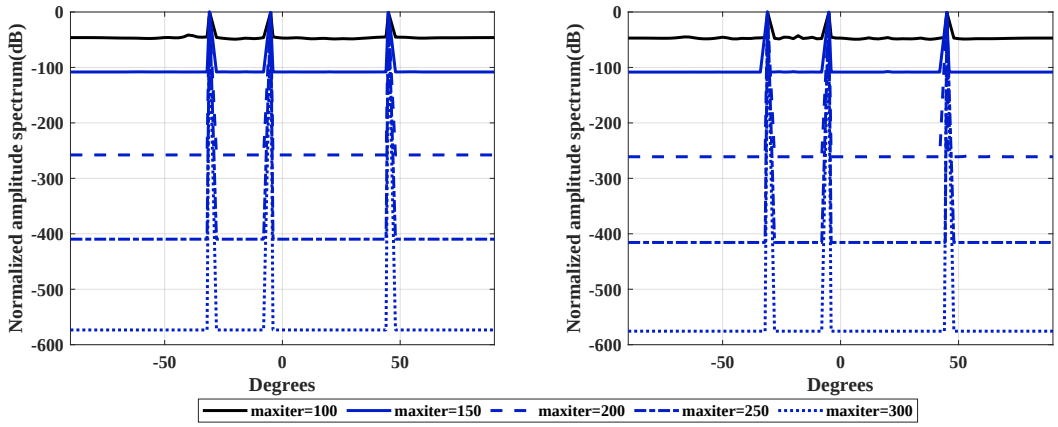


Fig. 8. Diffrernt number of iterations with nornalized amplitude spectrum at $Q = 1$, $r = 2°$, $T = 200$, $convertiter = 100$ and SNR $= 4$dB: CPA at $M_1 = 3, M_2 = 5$ (left); NA at $M_1 = 3, M_2 = 4$ (right).

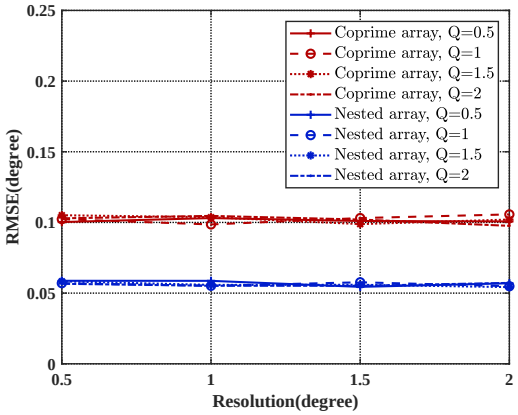


Fig. 9. Influence of overcompleteness of the main dictionary with dense angle grids for CPA at $M_1 = 3, M_2 = 5$ and NA at $M_1 = 3, M_2 = 4$,$T = 200$, and SNR $= 4$dB, and convertiter=200.

indicated by the GPS, a linear interpolation is applied to estimate the corresponding direction. Specifically, let $\hat{\theta}^{(t)}$ denote the interpolated estimated direction at time instance t, and $\theta_{\text{GPS}}^{(t)}$ denote the GPS-indicated (reference) direction at the same time instance. The average absolute error (denoted as Err_{avg}) is then defined as

$$Err_{\text{avg}} = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{\theta}^{(t)} - \theta_{\text{GPS}}^{(t)}\right|, \tag{61}$$

We also present the running times of the comparison algorithms for the entire 50-minute event in table IV, running on a system with a CPU of i7-13700HX, a 32GB RAM, and with a 64-bit Win11 system. The results demonstrate that the proposed algorithm is capable of estimating the incoming angle in real environments and has the lowest runtime requirement than the other Bayesian-type algorithms.

Table IV: Average DOA estimation error and CPU running times of different methods

Algorithm	Average DOA error (deg)	CPU time (sec)
CBF	2.9943	9.7171
MVDR	2.8214	7.3758
OGSBI	2.9344	195.7306
untrimmed-SBL	2.9512	65.8680
Real-VBI	3.5241	56.2624
JSBL	2.8149	150.2713
HHC-VBI	**2.7580**	**44.9097**

VI. CONCLUSION

In this paper, the half-Cauchy framework of distributions, together with a new hierarchical prior is introduced for sparse Bayesian DOA estimation based on arbitrary linear arrays. In comparison to the well-known Horseshoe prior in the half-Cauchy framework of distributions and the Laplace and Gaussian prior in the exponential framework of distributions, the local scales of the PDF of the proposed framework of distributions exhibit the sharpest spikes and heaviest tails. This indicates that the proposed prior is capable of obtaining a sparser solution. Furthermore, the proposed block matrix iteration method is capable of controlling the dimension of the main dictionary through a posterior covariance matrix approximation, thereby enabling the control of Bayesian-like algorithms with high dictionary correlation under overcomplete dictionaries and high ill-conditionality. In order to address the issue of high power spectrum sidelobes, a mixing prior updating strategy based on HHC and a Gaussian distribution is proposed. The results based on simulations and real datasets demonstrate that the proposed real-valued prior is capable of achieving faster convergence with a sparser solution.

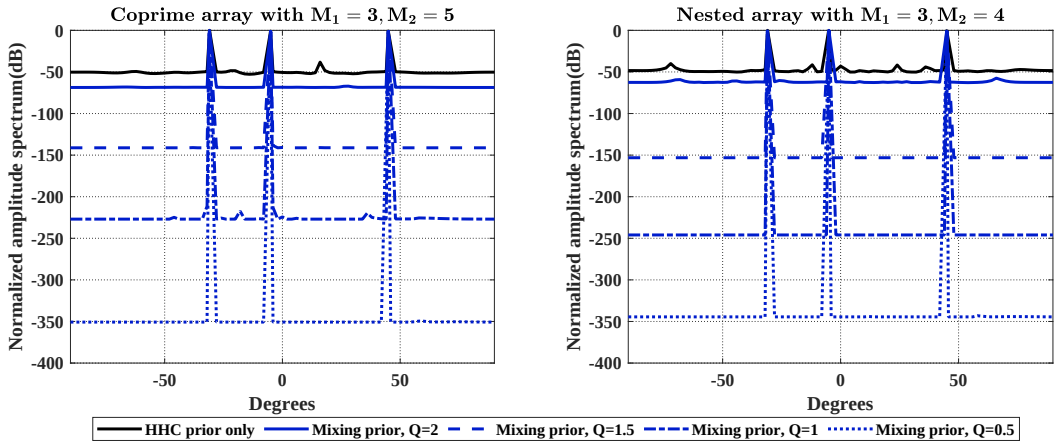
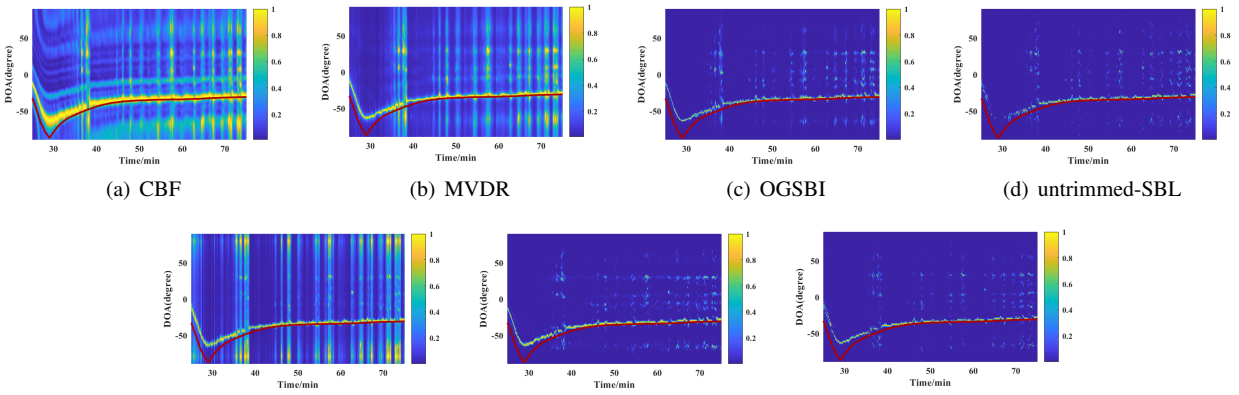


Fig. 10. Diffrernt completeness with nornalized amplitude spectrum at $r = 2°$, $T = 200$, $convertiter = 100$, maxiter=200 and SNR $= 4$dB. CPA at $M_1 = 3, M_2 = 5$ (left); NA at $M_1 = 3, M_2 = 4$ (right).



(a) CBF (b) MVDR (c) OGSBI (d) untrimmed-SBL

(e) Real-VBI (f) JSBL (g) The proposed HHC-VBI

Fig. 11. DOA tracking results by different algorithms.

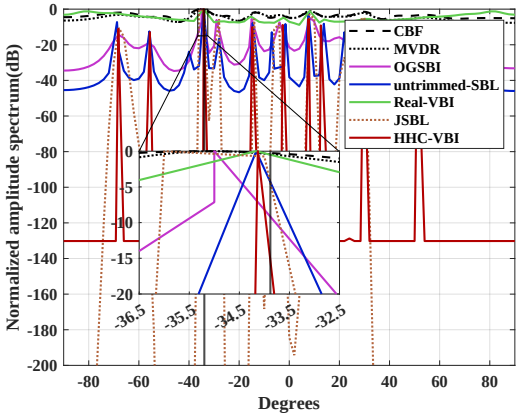


Fig. 12. Spectrum at 362^{nd} frame, with $T \approx 62.84$min.

Appendix A
Proof of Approximation for Local Scale PDF

The conditional priors at the local scale can be expressed as:

$$p(\hat{p}_i|\lambda_i) \sim \mathcal{N}(0, \lambda_i^2), p(\lambda_i|\eta_i) \sim \mathcal{C}^+(0, \eta_i), p(\eta_i) \sim \mathcal{C}^+(0, 1). \tag{62}$$

Proof:

First, the marginal probability density function of λ_i is given by:

$$\begin{aligned} p(\lambda_i) &= \int_0^{+\infty} p(\lambda_i|\eta_i)p(\eta_i)d\eta_i \\ &= \int_0^{+\infty} \frac{2}{\pi\eta_i\left(1+\frac{\lambda_i^2}{\eta_i^2}\right)} \frac{2}{\pi(1+\eta_i^2)} d\eta_i = \frac{4}{\pi^2}\frac{\ln\lambda_i}{\lambda_i^2-1}. \end{aligned} \tag{63}$$

Now, integrating $p(\hat{p}_i|\lambda_i)p(\lambda_i)$ over λ_i gives:

$$p(\hat{p}_i) = \frac{4}{\pi^2\sqrt{2\pi}} \int_0^{\infty} \frac{1}{\lambda_i} \exp\left(-\frac{1}{2\lambda_i^2}\hat{p}_i^2\right) \frac{\log(\lambda_i)}{\lambda_i^2-1} d\lambda_i. \tag{64}$$

Substituting $\zeta = 1/\lambda_i^2$, we then get:

$$p(\hat{p}_i) = \frac{1}{\pi^2\sqrt{2\pi}} \int_0^{\infty} \exp\left(-\frac{\zeta}{2}\hat{p}_i^2\right) \frac{\log(\zeta)}{\zeta-1} d\zeta. \tag{65}$$

Lemma 2 For $\forall\zeta > 0$, the following inequality holds:

$$\frac{2}{1+\zeta} \leq \frac{\ln\zeta}{\zeta-1} \leq \frac{1}{\sqrt{\zeta}}, \quad \forall\zeta > 0. \tag{66}$$

The equation holds when $\zeta = 1$; when $\zeta > 0$ and $\zeta \neq 1$ the following inequality holds according to L'Hôpital's law:

$$\frac{1}{2}\zeta^{-\frac{3}{2}} < \frac{\ln\zeta + \zeta^{-1} - 1}{(\zeta-1)^2} < \frac{2}{(1+\zeta)^2} \tag{67}$$

■

From the right-hand side of Lemma 2, we can derive:

$$\begin{aligned} p(\hat{p}_i) &< \frac{1}{\pi^2}\int_0^\infty \zeta^{-1/2}\exp\left(-\frac{\zeta}{2}\hat{p}_i^2\right)d\zeta \\ &= \frac{\Gamma(1/2)}{\pi^2(\hat{p}_i^2/2)^{1/2}} = \frac{1}{\pi^2|\hat{p}_i|}. \end{aligned} \tag{68}$$

Similarly, based on Lemma 2 and the Cauchy principal value inequality [45], we have

$$\frac{\exp(-t)}{2}\ln\left(1+\frac{2}{t}\right) < E_1(t) < \exp(-t)\ln\left(1+\frac{1}{t}\right), \tag{69}$$

where $E_1(t)$ is defined as:

$$E_1(t) \triangleq \int_1^\infty \frac{1}{z}\exp(-zt)dz. \tag{70}$$

Let $\zeta \to \zeta + 1$, and thus we obtain:

$$\begin{aligned} p(\hat{p}_i) &\geq \frac{1}{\pi^2\sqrt{2\pi}}\int_0^\infty \exp\left(-\frac{\zeta}{2}\hat{p}_i^2\right)\frac{2}{\zeta+1}d\zeta \\ &= \frac{2\exp(\hat{p}_i^2/2)}{\pi^2\sqrt{2\pi}}\int_1^\infty \frac{1}{\zeta}\exp\left(-\frac{\zeta\hat{p}_i^2}{2}\right)d\zeta \\ &= \frac{2\exp(\hat{p}_i^2/2)}{\pi^2\sqrt{2\pi}}E_1(\hat{p}_i^2/2) > \frac{1}{\pi^2\sqrt{2\pi}}\ln\left(1+\frac{4}{\hat{p}_i^2}\right). \end{aligned} \tag{71}$$

Appendix B
Proof of Block Covariance Approximation

For the set of large latent variables and their complement, assuming sparsity of the latent variables, the following results hold:

$$\hat{\mathbf{p}}_S \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_S), \tag{72}$$

$$\hat{\mathbf{p}}_{S^c} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_{S^c}). \tag{73}$$

Under the assumption of (21), the corresponding posterior expectations are:

$$\boldsymbol{\mu}_S = \sigma_c^{-2}\mathbf{\Sigma}_S\left[\tilde{\mathbf{\Phi}}_w\right]_S^T\tilde{\mathbf{r}}_w = \mathbf{\Gamma}_S\left[\tilde{\mathbf{\Phi}}_w\right]_S^T\tilde{\mathbf{r}}_w, \tag{74}$$

$$\boldsymbol{\mu}_{S^c} = \sigma_c^{-2}\mathbf{\Sigma}_{S^c}\left[\tilde{\mathbf{\Phi}}_w\right]_{S^c}^T\tilde{\mathbf{r}}_w = \mathbf{\Gamma}_{S^c}\left[\tilde{\mathbf{\Phi}}_w\right]_{S^c}^T\tilde{\mathbf{r}}_w \approx \mathbf{0}. \tag{75}$$

Note that if $\boldsymbol{\mu}_S$ contains all valid signal components, then $\boldsymbol{\mu}_{S^c}$ plays a minimal role in the performance of regression after the learning rate of the Bayesian model has reached a steady state. Consequently, the approximate form of the final obtained posterior expectation is:

$$\boldsymbol{\mu}_A = \mathbf{G}\begin{bmatrix}\boldsymbol{\mu}_S \\ \mathbf{0}\end{bmatrix} = \mathbf{G}\begin{bmatrix}\sigma_c^{-2}\mathbf{\Sigma}_S[\tilde{\mathbf{\Phi}}_w]_S^T\tilde{\mathbf{r}}_w \\ \mathbf{0}\end{bmatrix}. \tag{76}$$

Thus, based on equation (12), the cross-covariance between the observations and the complement of latent variables is:

$$Cov[\tilde{\mathbf{r}}_w, \hat{\mathbf{p}}_{S^c}] = E[(\tilde{\mathbf{\Phi}}_w\mathbf{G}\hat{\mathbf{p}}_F)(\mathbf{p}_{S^c})^T] = \left[\tilde{\mathbf{\Phi}}_w\right]_{S^c}^T\mathbf{\Gamma}_{S^c}. \tag{77}$$

The covariance of the posterior is expressed as:

$$\begin{aligned} \mathbf{\Sigma}_A &= E[(\hat{\mathbf{p}}-\boldsymbol{\mu})(\hat{\mathbf{p}}-\boldsymbol{\mu})^T] \\ &= \mathbf{G}\begin{bmatrix}E[(\hat{\mathbf{p}}_S-\boldsymbol{\mu}_S)(\hat{\mathbf{p}}_S-\boldsymbol{\mu}_S)^T] & E[(\hat{\mathbf{p}}_S-\boldsymbol{\mu}_S)\hat{\mathbf{p}}_{S^c}^T] \\ E[\hat{\mathbf{p}}_{S^c}(\hat{\mathbf{p}}_S-\boldsymbol{\mu}_S)^T] & E[\hat{\mathbf{p}}_{S^c}\hat{\mathbf{p}}_{S^c}^T]\end{bmatrix}\mathbf{G}^T \\ &= \mathbf{G}\begin{bmatrix}\mathbf{\Sigma}_S & \mathbf{\Sigma}_{Cov} \\ \mathbf{\Sigma}_{Cov} & \mathbf{\Gamma}_{S^c}\end{bmatrix}\mathbf{G}^T, \end{aligned} \tag{78}$$

where the cross-correlation of solutions between the main dictionary and sub-dictionary is shown as:

$$\begin{aligned} \mathbf{\Sigma}_{Cov} &= Cov(\hat{\mathbf{p}}_S, \hat{\mathbf{p}}_{S^c}) = \left\langle\hat{\mathbf{p}}_S\hat{\mathbf{p}}_{S^c}^T\right\rangle - \left\langle\hat{\mathbf{p}}_S\right\rangle\left\langle\hat{\mathbf{p}}_{S^c}\right\rangle \\ &= -\mathbf{\Gamma}_S\left[\tilde{\mathbf{\Phi}}_w\right]_S^T\mathbf{D}_S^{-1}\left[\tilde{\mathbf{\Phi}}_w\right]_{S^c}\mathbf{\Gamma}_{S^c}. \end{aligned} \tag{79}$$

Appendix C
Variational Inference of Hyperparameters

C.1 Variational inference of λ_i

To the Eq. (23), we have:

$$\begin{aligned} \ln q(\boldsymbol{\lambda}) &\propto \left\langle\ln p(\hat{\mathbf{p}}|\boldsymbol{\lambda},\delta) + \ln p(\boldsymbol{\lambda}|\boldsymbol{\nu})\right\rangle_{q_{\{\boldsymbol{\kappa}\setminus\boldsymbol{\lambda}\}}(\{\boldsymbol{\kappa}\setminus\boldsymbol{\lambda}\})} \\ &= \sum_{j=1}^N -(1+1)\ln\lambda_j \\ &\quad - \lambda_j^{-1}\left(\frac{1}{2}\left\langle\delta\right\rangle^{-1}\left\langle\hat{p}_j^2\right\rangle + \left\langle\nu_j\right\rangle^{-1}\right), \end{aligned} \tag{80}$$

From the prior structure of $\boldsymbol{\lambda}$ with the above equation, its posterior obeys a distribution of

$$q(\boldsymbol{\lambda}) = \prod_{i=1}^N \mathcal{IG}(\lambda_i|\tilde{a},\tilde{b}_i), \tag{81}$$

where:

$$\begin{cases}\tilde{a} = 1, \\ \tilde{b}_i = \frac{1}{2}\left\langle\delta\right\rangle^{-1}(\mu_i^2 + \Sigma_{i,i}) + \left\langle\nu_j\right\rangle^{-1}.\end{cases} \tag{82}$$

Hence we have:

$$\lambda_i^{new} = \left\langle\lambda_i\right\rangle = \frac{\tilde{b}_i}{\tilde{a}} = \frac{1}{2}\left\langle\delta\right\rangle^{-1}(\mu_i^2 + \Sigma_{i,i}) + \left\langle\nu_j\right\rangle^{-1}. \tag{83}$$

C.2 Variational inference of δ

Expand Eq. (24):

$$\begin{aligned} \ln q(\delta) &\propto \left\langle\sum_{i=1}^N \ln p(\hat{p}_i|\lambda_i,\delta) + \ln p(\delta|\zeta)\right\rangle_{q_{\{\boldsymbol{\kappa}\setminus\delta\}}(\{\boldsymbol{\kappa}\setminus\delta\})} \\ &= -(\frac{N+1}{2}+1)\ln\delta \\ &\quad - \delta^{-1}\left[\frac{1}{2}\sum_{i=1}^N\left\langle\lambda_i\right\rangle^{-1}\left\langle\hat{p}_i^2\right\rangle + \left\langle\zeta\right\rangle^{-1}\right]. \end{aligned} \tag{84}$$

From the prior structure of δ and the above equation, the distribution that its posterior obeys is:

$$q(\delta) = \mathcal{IG}(\delta|\tilde{c}, \tilde{d}), \tag{85}$$

where:

$$\begin{cases} \tilde{c} = \frac{N+1}{2}, \\ \tilde{d} = \frac{1}{2}\sum_{i=1}^{N} \langle \lambda_i \rangle^{-1} (\mu_i^2 + \Sigma_{i,i}) + \langle \zeta \rangle^{-1}. \end{cases} \tag{86}$$

Hence we have:

$$\delta^{new} = \langle \delta \rangle = \frac{\sum_{i=1}^{N} \langle \lambda_i \rangle^{-1} (\mu_i^2 + \Sigma_{i,i}) + 2\langle \zeta \rangle^{-1}}{N+1}. \tag{87}$$

C.3 Variational inference of ν_i and η_i

For Eq. (25), we have:

$$\begin{aligned} \ln q(\boldsymbol{\nu}) &\propto \left\langle \sum_{i=1}^{N} \ln p(\lambda_i|\nu_i) + \ln p(\nu_i|\eta_i, \tau) \right\rangle_{q_{\{\boldsymbol{\kappa}\setminus\boldsymbol{\nu}\}}(\{\boldsymbol{\kappa}\setminus\boldsymbol{\nu}\})} \\ &= -\sum_{i=1}^{N} \left[2\ln \nu_i + \nu_i^{-1} (\langle \lambda_i \rangle^{-1} + \langle \tau \rangle^{-1} \langle \eta_i \rangle^{-1}) \right]. \end{aligned} \tag{88}$$

We similarly model its posterior as an Inversed-Gamma distribution with:

$$q(\boldsymbol{\nu}) = \prod_{i=1}^{N} \mathcal{IG}(\nu_i|\tilde{e}, \tilde{f}_i), \tag{89}$$

where

$$\begin{cases} \tilde{e} = 1, \\ \tilde{f}_i = \langle \lambda_i \rangle^{-1} + \langle \tau \rangle^{-1} \langle \eta_i \rangle^{-1}, \end{cases} \tag{90}$$

is obtained from the prior of ν_i, and hence we have:

$$\nu_i^{new} = \langle \nu_i \rangle = \langle \lambda_i \rangle^{-1} + \langle \tau \rangle^{-1} \langle \eta_i \rangle^{-1}. \tag{91}$$

Similarly, with an expansion of Eq. (27) and based on its a prior structure, we have:

$$\eta_i^{new} = \langle \eta_i \rangle = \langle \nu_i \rangle^{-1} \langle \tau \rangle^{-1} + \langle \varepsilon_i \rangle^{-1}. \tag{92}$$

C.4 Variational inference of τ

Expanding Eq. (28), we have:

$$\begin{aligned} \ln q(\tau) &\propto \left\langle \sum_{i=1}^{N} \ln p(\nu_i|\eta_i, \tau) + \ln p(\tau|\xi) \right\rangle_{q_{\{\boldsymbol{\kappa}\setminus\tau\}}(\{\boldsymbol{\kappa}\setminus\tau\})} \\ &= -(\frac{3}{2} + \frac{N}{2}) \ln \tau \\ &\quad - \tau^{-1} (\langle \xi \rangle^{-1} + \sum_{i=1}^{N} \langle \eta_i^{-1} \rangle \langle \nu_i \rangle^{-1}), \end{aligned} \tag{93}$$

by employing the prior structure of τ and the aforementioned equation, the distribution of its posterior obeys:

$$q(\tau) = \mathcal{IG}(\tau|\tilde{m}, \tilde{n}), \tag{94}$$

where

$$\begin{cases} \tilde{m} = \frac{N}{2} + \frac{1}{2}, \\ \tilde{n} = \langle \xi \rangle^{-1} + \sum_{i=1}^{N} \langle \eta_i^{-1} \rangle \langle \nu_i \rangle^{-1}, \end{cases} \tag{95}$$

and thus we have:

$$\tau^{new} = \langle \tau \rangle = \frac{\langle \xi \rangle^{-1} + \sum_{i=1}^{N} \langle \eta_i^{-1} \rangle \langle \nu_i \rangle^{-1}}{\frac{N}{2} + \frac{1}{2}}. \tag{96}$$

C.5 Vatiational inference of ε_i, ζ and ξ

For Eq. (29), we have:

$$\begin{aligned} \ln q(\boldsymbol{\varepsilon}) &\propto \left\langle \sum_{i=1}^{N} \ln p(\eta_i|\varepsilon_i) + \ln p(\varepsilon_i) \right\rangle_{q_{\{\boldsymbol{\kappa}\setminus\boldsymbol{\varepsilon}\}}(\{\boldsymbol{\kappa}\setminus\boldsymbol{\varepsilon}\})} \\ &= -\sum_{i=1}^{N} \left[2\ln \varepsilon_i + \varepsilon_i^{-1} (1 + \langle \eta_i \rangle^{-1}) \right], \end{aligned} \tag{97}$$

Similarly, the posterior distribution of ε_i is represented based on the prior distribution as:

$$q(\varepsilon_i) = \mathcal{IG}(\varepsilon_i|\tilde{o}, \tilde{t}_i). \tag{98}$$

where:

$$\begin{cases} \tilde{o} = 1, \\ \tilde{t}_i = 1 + \langle \eta_i \rangle^{-1}, \end{cases} \tag{99}$$

and thus the close-form expression of ε_i is given by:

$$\varepsilon_i^{new} = \langle \varepsilon_i \rangle = 1 + \langle \eta_i \rangle^{-1}. \tag{100}$$

In a similar manner, by extending Eq. (26) and Eq. (30), we can derive a closed-form solution with:

$$\zeta^{new} = \langle \zeta \rangle = 1 + \langle \delta \rangle^{-1}, \tag{101}$$

$$\xi^{new} = \langle \xi \rangle = 1 + \langle \tau \rangle^{-1}. \tag{102}$$

REFERENCE

[1] H. Xu, M. Jin, Q. Guo, T. Jiang, and Y. Tian, "Direction of arrival estimation with gain-phase uncertainties in unknown nonuniform noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 6, pp. 9686–9696, 2023.

[2] B. Liao and S.-C. Chan, "Direction finding in partly calibrated uniform linear arrays with unknown gains and phases," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 217–227, 2015.

[3] C. Qian, L. Huang, M. Cao, J. Xie, and H. C. So, "PUMA: An improved realization of mode for DOA estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 5, pp. 2128–2139, 2017.

[4] B. Liao, S.-C. Chan, L. Huang, and C. Guo, "Iterative methods for subspace and DOA estimation in nonuniform noise," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3008–3020, 2016.

[5] F. Chen, D. Yang, and S. Mo, "A DOA estimation algorithm based on eigenvalues ranking problem," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–15, 2023.

[6] X. Huang and B. Liao, "One-bit MUSIC," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 961–965, 2019.

[7] F. Chen, D. Yang, S. Mo, X. Huang, M. Wang, Z. Zhu, and Y. Li, "A method for estimating the direction of arrival without knowing the source number using acoustic vector sensor arrays," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–16, 2024.

[8] F. Chen, D. Yang, and S. Mo, "A DOA estimation algorithm based on eigenvalues ranking problem," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–15, 2023.

[9] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[10] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[11] Q. Shen, W. Liu, W. Cui, S. Wu, and P. Pal, "Simplified and enhanced multiple level nested arrays exploiting high-order difference co-arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3502–3515, 2019.

[12] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, 1968.

[13] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, 2010.

[14] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, 2010.

[15] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 4, pp. 806–811, 1985.

[16] X. Wu, W.-P. Zhu, and J. Yan, "A Toeplitz covariance matrix reconstruction approach for direction-of-arrival estimation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8223–8237, 2017.

[17] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[18] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[19] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[20] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[21] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, 2010.

[22] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, 2012.

[23] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5145–5157, 2016.

[24] C.-L. Liu, P. Vaidyanathan, and P. Pal, "Coprime coarray interpolation for DOA estimation via nuclear norm minimization," in *Proc. IEEE Int. Symp. Circuits and Syst. (ISCAS)*, Montréal, QC, CAN, 2016, pp. 2639–2642.

[25] C. Zhou, Y. Gu, Z. Shi, and M. Haardt, "Structured Nyquist correlation reconstruction for DOA estimation with sparse arrays," *IEEE Trans. Signal Process.*, vol. 71, pp. 1849–1862, 2023.

[26] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4489–4493, 2011.

[27] T. Chen, L. Shi, and L. Guo, "Gridless direction of arrival estimation exploiting sparse linear array," *IEEE Signal Process. Lett.*, vol. 27, pp. 1625–1629, 2020.

[28] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. D. Zhang, "Direction-of-arrival estimation for coprime array via virtual array interpolation," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5956–5971, 2018.

[29] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 1–11, 2012.

[30] J. Dai and H. C. So, "Real-valued sparse Bayesian learning for DOA estimation with arbitrary linear arrays," *IEEE Trans. Signal Process.*, vol. 69, pp. 4977–4990, 2021.

[31] N. Li, X.-K. Zhang, B. Zong, F. Lv, J. Xu, and Z. Wang, "An off-grid direction-of-arrival estimator based on sparse Bayesian learning with three-stage hierarchical Laplace priors," *Signal Process.*, vol. 218, p. 109371, 2024.

[32] J. Yang, Y. Yang, and J. Lu, "A variational Bayesian strategy for solving the DOA estimation problem in sparse array," *Digit. Signal Process.*, vol. 90, pp. 28–35, 2019.

[33] F. Scheipl, L. Fahrmeir, and T. Kneib, "Spike-and-slab priors for function selection in structured additive regression models," *J. Am. Stat. Assoc.*, vol. 107, no. 500, pp. 1518–1532, 2012.

[34] Q. Wan, H. Duan, J. Fang, H. Li, and Z. Xing, "Robust Bayesian compressed sensing with outliers," *Signal Process.*, vol. 140, pp. 104–109, 2017.

[35] R. Zheng, X. Xu, Z. Ye, and J. Dai, "Robust sparse Bayesian learning for DOA estimation in impulsive noise environments," *Signal Process.*, vol. 171, p. 107500, 2020.

[36] V. Ročková, "Bayesian estimation of sparse signals with a continuous spike-and-slab prior," *Ann. Stat.*, vol. 46, no. 1, pp. 401 – 437, 2018.

[37] V. Ročková and E. I. George, "The spike-and-slab lasso," *J. Am. Stat. Assoc.*, vol. 113, no. 521, pp. 431–444, 2018.

[38] C. Louizos, K. Ullrich, and M. Welling, "Bayesian compression for deep learning," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[39] Q. Li and N. Lin, "The Bayesian elastic net," *Bayesian Anal.*, vol. 5, no. 1, pp. 151 – 170, 2010.

[40] R. Bai and M. Ghosh, "On the beta prime prior for scale parameters in high-dimensional Bayesian regression models," *Stat. Sin.*, vol. 31, no. 2, pp. 843–865, 2021.

[41] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digit. Signal Process.*, vol. 8, no. 3, pp. 185–210, 1998.

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

[43] F. Chen, J. Dai, N. Hu, and Z. Ye, "Sparse Bayesian learning for off-grid DOA estimation with nested arrays," *Digit. Signal Process.*, vol. 82, pp. 187–193, 2018.

[44] A. Armagan, M. Clyde, and D. Dunson, "Generalized beta mixtures of gaussians," *Adv. Neural Inf. Process. Syst.*, vol. 24, 2011.

[45] C. M. Carvalho, N. G. Polson, and J. G. Scott, "The horseshoe estimator for sparse signals," *Biometrika*, vol. 97, no. 2, pp. 465–480, 2010.

[46] A. Bhadra, J. Datta, N. G. Polson, and B. Willard, "Lasso meets horseshoe," *Stat. Sci.*, vol. 34, no. 3, pp. 405–427, 2019.

[47] N. G. Polson and J. G. Scott, "On the half-cauchy prior for a global scale parameter," *Bayesian Anal.*, vol. 7, no. 4, pp. 887–902, 2012.

[48] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, 2008.

[49] M. P. Wand, J. T. Ormerod, S. A. Padoan, and R. Frühwirth, "Mean field variational Bayes for elaborate distributions," *Bayesian Anal.*, vol. 6, no. 4, pp. 847–900, 2011.

[50] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, 2010.

[51] S. Cai, G. Wang, J. Zhang, K.-K. Wong, and H. Zhu, "Efficient direction of arrival estimation based on sparse covariance fitting criterion with modeling mismatch," *Signal Process.*, vol. 137, pp. 264–273, 2017.

[52] Karimi, Mahmood and Zare, Mohammadreza and Derakhtian, Mostafa, "Sparse Bayesian learning with Jeffreys' noninformative prior for off-grid DOA estimation," *Signal Process.*, vol. 230, p. 109809, 2025.

[53] J. Murray and D. Ensberg, "The SWellEx-96 Experiment," 1996. [Online]. Available: https://swellex96.ucsd.edu

[54] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

Zhendong Chen is currently pursuing the M.S. degree with the School of Informatics, Xiamen University, Xiamen, China. His current research interests include sparse signal recovery and its application in array signal processing.

Xicheng Lu received the B.Sc. degree in electrical engineering from Fuzhou University and the M.Sc. degree in electronic and electrical engineering from The University of Sheffield. He is currently pursuing the Ph.D. degree with the School of Electronic Engineering and Computer Science, Queen Mary University of London. His research interests include low-bit quantization, phase retrieval, and their applications in array signal processing.

Yongfeng Huang is currently pursuing the M.S. degree with the School of Informatics, Xiamen University, Xiamen, China. His current research interests include sparse signal recovery, signal processing and deep learning.

Dingzhao Li is currently pursuing the Ph.D. degree with the School of Informatics, Xiamen University, Xiamen, China. His current research interests include communication specific emitter identification, class-incremental learning, prototype learning and deep learning.

Wei Liu (S'01-M'04-SM'10) received his BSc (Space Physics) and LLB (Intellectual Property Law) degrees from Peking University, China, in 1996 and 1997, respectively, MPhil from the Department of Electrical and Electronic Engineering, University of Hong Kong in 2001, and PhD from the School of Electronics and Computer Science, University of Southampton, UK, in 2003. He then worked as a postdoc first at Southampton and later at Imperial College London. From 2005 to 2023, he was a Lecturer/Senior Lecturer at the Department of Electronic and Electrical Engineering, University of Sheffield, UK and from 2023 to 2024, a Reader at the School of Electronic Engineering and Computer Science, Queen Mary University of London (Visiting Professor from 2024 to 2027). Since 2024, he has been a Professor at the Department of Electrical and Electronic Engineering, Hong Kong Polytechnic University. He has published 470+ journal and conference papers, six book chapters, and two research monographs titled "Wideband Beamforming: Concepts and Techniques" (Wiley, 2010) and "Low-Cost Smart Antennas" (Wiley, 2019), respectively. His research interests are mainly focused on sensor array and multichannel signal processing and its various applications, such as robotics and autonomous systems, human computer interface, radar, sonar, and wireless communications.

He is a member of the Applied Signal Processing Systems Technical Committee (Vice-Chair for 2025-2026) of the IEEE Signal Processing Society (SPS), the Digital Signal Processing Technical Committee (Chair for 2022-2024) of the IEEE Circuits and Systems Society, and the IEEE SPS Education Board (2024-2026, Chair of its Educational Conference Program Committee), and a former member of the Sensor Array and Multichannel Signal Processing Technical Committee of the IEEE SPS (Chair for 2021-2022), the IEEE SPS Technical Directions Board (2021-2022), and the IEEE SPS Conference Board and its Executive Subcommittee (2022-2023). He also acted as an associate editor for IEEE Trans. on Signal Processing, IEEE Access, and Journal of the Franklin Institute, and an Executive Associate Editor-in-Chief of the journal Frontiers of Information Technology and Electronic Engineering; currently he is an associate editor for IEEE Transactions on Circuits and Systems I: Regular Papers and IEEE Antennas and Wireless Propagation Letters. He is an IEEE Distinguished Lecturer for the Aerospace and Electronic Systems Society (2023-2026).

Shaohua Hong (Member, IEEE) received the B.Sc. degree in electronics and information engineering and the Ph.D. degree in electronics science and technology from Zhejiang University, Hangzhou, China, in 2005 and 2010, respectively.

He is currently an Associate Professor with the Department of Informatics and Communication Engineering, Xiamen University, Xiamen, China. His research interests include coding and modulation, array signal processing, specific emitter identification, and nonlinear signal processing. Dr. Hong has been an Editor of KSII Transactions on Internet and Information Systems since 2015.

Haixin Sun (Senior Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Shandong University of Science and Technology, Shandong, China, in 1999 and 2003, respectively, and the Ph.D. degree in communication engineering from the Institute of Acoustic, Chinese Academy of Science, Shanghai, China, in 2006.

From March 2012 to April 2013, he was with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA. He is currently a Professor and Doctorial Tutor with the School of Informatics, Xiamen University, Fujian, China.

Dr. Sun is a member of the Institute of Electronics, Information and Communication Engineers. He was the recipient of the Huawei Fellowship of Xiamen University in 2010, the Faculty of Engineering Excellence Award of Xiamen University, and the Third Prize of Chinese Army Scientific and Technological in 2017.