Stochastic Particle-Based Variational Bayesian Inference for Multi-band Radar Sensing

Zhixiang Hu, An Liu, Senior Member, IEEE, Yubo Wan, Graduate Student Member, IEEE, Tony Xiao Han and Minjian Zhao, Member, IEEE

Abstract—Multi-band fusion is an important technology to improve the radar sensing performance. In the multi-band radar sensing signal model, the associated likelihood function has oscillation phenomenon, which makes it difficult to obtain high-accuracy parameter estimation. To cope with this challenge, we divide the radar target parameter estimation into two stages of coarse estimation and refined estimation, where the coarse estimation is used to narrow down the search range for the refined estimation, and the refined estimation is based on the Bayesian approach to avoid the convergence to a bad local optimum of the likelihood function. Specifically, in the coarse estimation stage, we employ a root MUSIC algorithm to achieve initial estimation. Then, we apply the block stochastic successive convex approximation (SSCA) approach to derive a novel stochastic particle-based variational Bayesian inference (SPVBI) algorithm for the Bayesian estimation of the radar target parameters in the refined stage. Unlike the conventional particle-based VBI (PVBI) in which only the probability of each particle is optimized and the per-iteration computational complexity increases exponentially with the number of particles, the proposed SPVBI optimizes both the position and probability of each particle, and it adopts the block SSCA to significantly improve the sampling efficiency by averaging over iterations. As such, it is shown that the proposed SPVBI can achieve a better performance than the conventional PVBI with a much smaller number of particles and per-iteration complexity. Finally, extensive simulations verify the advantage of the proposed algorithm over various baseline algorithms.

Index Terms—Radar sensing, multi-band fusion, variational Bayesian inference, stochastic successive convex approximation.

I. INTRODUCTION

MULTI-BAND fusion technology is widely applied to high-precision radar sensing, which can achieve the monitoring and identification of targets by obtaining the precise information of target parameters such as the range and scattering coefficient to identify the size, shape, structure and movement of the target. Application scenarios include vehicle-mounted radar imaging, security inspection, medical imaging, remote sensing, inverse synthetic aperture (ISAR) radar imaging, etc. In addition, with the rise of integrated sensing and communication (ISAC) [1], how to use multiple limited communication bands to achieve higher estimation accuracy has become a big challenge. To meet these application requirements, the radar needs to have a high enough resolution and accuracy.

The resolution of radar sensing depends on the bandwidth of the transmitted signal. However, we cannot straightforwardly increase the bandwidth due to the limited spectrum resources. Besides, large bandwidth will bring great pressure to signal acquisition, data transmission and processing, which leads to a high hardware cost. Although there are many improvements in the super-resolution method of single-band signal processing, it is still limited by the fixed signal bandwidth. To address these challenges, the technology of multi-band signal fusion processing has been proposed, that improves the resolution and range accuracy by coherent processing of radar signals in non-overlapping bands.

The existing multi-band radar sensing methods can be divided into three categories [2], i.e., spectral estimation method, sparse signal reconstruction method, and probabilistic inference method.

Spectral Estimation: In [3], the authors proposed band-width extrapolation (BWE) and ultra wide band coherent synthesis techniques for multi-band fusion. They first determine the parameters of an all-pole model, which is an approximation of the signal model. Coherent processing is then performed for the two bands to compensate for their phase differences. After that is the band interpolation or extrapolation. Finally, the range measure of super-resolution can be obtained by using the fitted full-band data. Although this method can effectively improve the range resolution with low complexity, it is sensitive to model error.

For the determination of model parameters, some subspace decomposition methods can be used such as Estimating Signal Parameters via Rotational Invariance Technique (ESPRIT) [4]–[5], Matrix Pencil algorithm [6]–[8] and RELAX algorithm [2], as well as transform domain methods such as aPFFT [9], which are essentially based on spectral search.

In addition, aiming at each processing flow of the algorithm, some methods of frequency band interpolation and extension are proposed, such as auto-regressive (AR) gap filling [10]–[12], minimum entropy criterion [13], CLEAN [14] and GAPES [15]–[16], as well as some methods of estimating the number of scattering centers (i.e. multi-path number), such as [17]–[18].

Sparse Signal Reconstruction: In this method, the signal model is expressed as a linear sparse form, in which the frequency-related scattering coefficient and delay term are discretized into an observation matrix, and the complex scalar is expressed as sparse vectors. The resulting compressed sensing problem can be solved by basis pursuit (BP), orthogonal matching pursuit (OMP) [19], FOCUSS [20] and so on. In
the corresponding solution vector, the number of non-zero elements is the number of scattering centers. In addition, the initial phase and linear phase (i.e., timing synchronization error) can also be aligned by extracting the phase and position of non-zero elements in the solution vector [21]. Compared with spectral estimation method, sparse reconstruction method does not need to determine the number of scattering centers in advance. However, in order to obtain high accuracy estimation, it needs to adopt a large scale observation matrix, which leads to a sharp increase in computational complexity.

**Probabilistic Inference:** In this category, probability estimates of target parameters are inferred based on statistical models. For example, we can obtain the maximum a posteriori (MAP) estimate by methods like expectation-maximization (EM) [22], support vector regression (SVR) [23] and conventional particle-based VBI [24]. Besides, sparse Bayesian learning (SBL) algorithm can also be used to solve the problem under certain sparse prior assumptions [25], and several algorithms have been proposed to improve the SBL, e.g., [26]–[28]. These algorithms usually require additional assumptions about priors, and it is not easy to achieve a good trade-off between estimation accuracy and computational complexity.

To overcome the drawbacks of the existing algorithms, in this paper, we propose a Stochastic Particle-based Variational Bayesian Inference (SPVBI) algorithm for high-accuracy multi-band radar sensing. The main contributions are summarized as follows.

- **Two-stage parameter estimation framework:** To reduce the computational complexity of the SPVBI algorithm, we adopt a two-stage parameter estimation framework. Specifically, a simple but stable MUSIC-based coarse estimation is used to narrow down the search range, so that the complexity of the refined stage can be greatly reduced. In addition, we find that the likelihood function oscillates violently, which is unfavorable for estimation. Therefore, we propose a signal model with bandwidth aperture structure [29] to reduce the degree of oscillation in the coarse estimation stage. Then with the particle positions initialized using the coarse estimation result, the SPVBI fully exploits the high resolution provided by the signal model with the band gap aperture structure [29] to produce an accurate estimation of the target parameters.

- **Stochastic particle-based variational Bayesian inference:** In the proposed SPVBI, three innovative ideas are used to achieve an accurate posterior estimation of the target parameters with reduced complexity and fast convergence speed. First, to avoid making any additional assumptions on the prior/posterior distributions of the target parameters, we adopt the particle-based approximation to transform the multiple integral operation in VBI into multiple weighted summation. Second, the particle positions are also updated in each iteration to minimize the VBI objective function. Such improved degree of freedom can further enhance the performance and accelerate the convergence speed. Finally, to avoid the exponential complexity with the particle number, we extend the SSCA approach in [30] to block SSCA and apply the block SSCA to significantly improve the sampling efficiency of the expectation operator in the VBI iteration by using the average-over-iteration technique.

- **Rigorous convergence analysis of SPVBI:** We prove that SPVBI is guaranteed to converge to a stationary point of the VBI problem, even though the number of samples used to calculate the expectation in each iteration is fixed as a constant that does not increase with the number of adopted particles.

The rest of the paper is organized as follows. In section II, we formulate the system model in the multi-band radar sensing scenario. In Section III, we propose a two-stage estimation framework. In section IV, we propose the SPVBI algorithm, together with the analysis of convergence and complexity. In Section V, we present numerical simulations and performance analysis. Finally, conclusions are presented in Section VI.

## II. SYSTEM MODEL

In the scene of multi-band sensing, we consider two radars that transmit linear frequency modulation (LFM) signals in non-contiguous frequency bands. The discrete received signal model in frequency domain can be formulated as [31]

\[
    r_m^{(n)} = \sum_{k=1}^{K} \alpha_k \left( j \frac{f_{c,m} + nf_{s,m}}{f_{c,m}} \right) e^{j2\pi(f_{c,m}+nf_{s,m})\tau_k} e^{j\phi_m} e^{-2j\pi(f_{c,m}+nf_{s,m})\delta_m + w_m^{(n)}},
\]

where \( m = 1, 2, \ldots, M \) is the frequency band index, \( N_m \) denotes the number of data samples in each band, \( n = 0, 1, \ldots, N_m - 1 \) denotes sample index and \( k = 1, 2, \ldots, K \) denotes the \( k \)-th scattering center. \( w_m^{(n)} \) denotes an additive white Gaussian noise (AWGN) following the distribution \( \mathcal{CN}(0, \sigma_n^2) \). \( \alpha_k \) is a complex scalar carrying the amplitude and phase information of a scattering center, and \( \beta_k \) is the scattering coefficient that characterizes the geometry of scattering center, which is usually an integer multiple of 0.5, e.g., \( \beta_k \in \{-1, -1/2, 0, 1/2, 1\} \). \( f_{s,m} \) and \( f_{c,m} \) are the frequency interval and initial frequency of \( m \)-th frequency band, respectively. \( \tau_k \) is the time delay of the \( k \)-th scattering center.

![Fig. 1. An illustration of multi-band distribution in frequency domain.](image)

Due to the instability of Pulse Repeat Frequency (PRF) and transmission delay of synchronization signal, the timing synchronization error \( \delta_m \) between two radars exists. And because of the hardware difference, the signals received by different radars are superimposed with a random initial phase \( \phi_m \). These two imperfect factors are the main obstacles for multi-band signal fusion and need to be calibrated.
III. A TWO-STAGE ESTIMATION FRAMEWORK

In the original signal model, the carrier frequency term \( f_{c,m} \) will lead to the violent oscillation of the likelihood function, where the main-lobe of the likelihood function will become sharp, accompanied by many fluctuating side-lobes, as illustrated by the solid pink line in Fig 2. Consequently, it will be extremely intractable to find the global optimum. When the signal-to-noise ratio (SNR) is not high and there are non-ideal factors, the estimate may deviate to the locally optimal side-lobe, resulting in the degradation of the estimation performance.

Therefore, we customize a two-stage estimation framework, which contains two signal models with bandwidth aperture and band gap aperture, respectively. The explanation of these two apertures will be given in the following sections.

![Fig. 2. An illustration of oscillation phenomena.](image)

A. Coarse Estimation

In the coarse estimation stage, the original signal model can be simplified into the following form:

- Signal model with the bandwidth aperture structure:

  \[
  y_m(n) = \sum_{k=1}^{K} \alpha_{k,m}^r e^{-j2\pi(n-f_{c,m})(\tau_k+\delta_m)} + w_m(n),
  \]

  where \( \alpha_{k,m}^r = \alpha_k^r \) is the complex scalar \( e^{-j2\pi f_{c,m}(\tau_k+\delta_m)} \) and the random phase \( e^{j\phi_m} \) of each band are absorbed into the complex scalar \( \alpha_k^r \), while the bandwidth term \( n_f_{c,m} \) is retained, so that all sub-bands share a bandwidth-dependent delay domain, which we call the bandwidth aperture. In addition, the exponential term \( (j)^{\delta_k} \) of the scattering coefficient relating to the phase can also be absorbed into \( \alpha_k \). Generally, \( (f_{c,m}+n_f_{c,m})^{\delta_k} \) can be approximated as 1 with negligible effect on coarse estimation since \( f_{c,m} \gg n_f_{c,m} \).

As shown by the dotted blue line in Fig 2, the bandwidth aperture smoothes the likelihood function, so that the true value can most likely be found in the peak region of the main lobe, thus obtaining a relatively rough but stable estimate.

Then, we use the root-MUSIC [32] algorithm to roughly estimate the delay, and the complex scalar after absorption is denoted as \( \alpha_{k,m}^r \). The coarse estimate signal model can be written in the following form:

\[
Y_m = X_mA_m + w_m,
\]

where

\[
Y_m = \begin{bmatrix} r_m^{(0)} & r_m^{(1)} & \cdots & r_m^{(N_m-1)} \end{bmatrix}^T,
A_m = \begin{bmatrix} \alpha_1^r & \alpha_2^r & \cdots & \alpha_K^r \end{bmatrix}^T,
X_m(\tau) = \begin{bmatrix} x(\tau_1 + \delta_m), x(\tau_2 + \delta_m), \cdots, x(\tau_K + \delta_m) \end{bmatrix},
w_m = \begin{bmatrix} w_m^{(0)} & w_m^{(1)} & \cdots & w_m^{(N_m-1)} \end{bmatrix}^T.
\]

and \( \tau = [\tau_1, \ldots, \tau_K] \).

Next, we can construct a Hankel matrix for subspace decomposition of the received signal [33], in the following form:

\[
H_m = \begin{bmatrix}
\begin{array}{cccc}
(1) & (1) & \cdots & (L-1) \\
(1) & (2) & \cdots & (L) \\
\vdots & \vdots & \ddots & \vdots \\
(N_m-L) & (N_m-L+1) & \cdots & (N_m-1)
\end{array}
\end{bmatrix},
\]

where \( L \) is the length of correlation window, empirically taken as \( N_m/3 \).

After that, we apply the eigenvalue decomposition to the Hankel matrix of the signal, \( H_m = U_mD_mV_m^H \). The eigenspace composed of the eigenvectors corresponding to the largest \( K \) (i.e. the number of scattering centers) eigenvalues is called the signal subspace and is denoted as \( S_{signal} \). The eigenspace composed of the eigenvectors corresponding to the remaining \( (N_m - K) \) eigenvalues is called the noise subspace and denoted as \( S_{noise} \). For determining the number of scattering centers \( K \), Akaike Information Criterion (AIC) [34] or the Minimum Description Length (MDL) [35] are both efficient methods that generally work well, which will not be described here for conciseness.

Essentially, the idea of root-MUSIC algorithm [32] is to apply the polynomial root-finding method to replace the spectral search of zeros in conventional MUSIC algorithm. Define the polynomial:

\[
f_l(z) = u_l^H z^l, \quad l = K + 1, \ldots, N_m,
\]

where \( u_l^H \) is the \( l \)-th eigenvector of noise subspace \( S_{noise} \), \( z = \hat{z}e^{-j2\pi f_{c,m}z} \). In order to utilize all noise eigenvectors, we wish to find the zeros of the following polynomials:

\[
f(z) = p^H(z)S_{noise}S_{noise}^HP(z).
\]

We rewrite (7) to get the polynomial in terms of \( z \) as:

\[
f(z) = z^{N_m-1}p^T(z^{-1})S_{noise}S_{noise}^HP(z).
\]

Find the roots of the above polynomial, wherein the \( K \) roots in the unit circle whose moduli are closest to 1 contain the information about delay. Denote those roots as \( \hat{\tau}_{k,m} \), \( k = 1, 2 \ldots K \), then the coarse estimate of the delay can be obtained:

\[
\hat{\tau}_{k,m} = \text{arg} \left( \frac{p_{\hat{\tau}_{k,m}}}{-2\pi f_{c,m}} \right).
\]

We can further improve the delay estimation accuracy by combining the results with different SNR in each frequency band:

\[
\hat{\tau}_k = \sum_{m=1}^{M} \frac{SNR_m}{SNR_1 + \cdots + SNR_M} \hat{\tau}_{k,m}.
\]
Also, the estimate of \( \alpha_{k,m} \) can be obtained by the least square (LS) method:
\[
\hat{A}_m = \left( X_H^m(\hat{\tau})X_m(\hat{\tau}) \right)^{-1} X_H^m(\hat{\tau})Y_m. \tag{11}
\]
where \( \hat{\tau} = [\hat{\tau}_1, \ldots, \hat{\tau}_K] \), \( \hat{A}_m = [\hat{\alpha}'_{1,m}, \hat{\alpha}'_{2,m}, \ldots, \hat{\alpha}'_{K,m}]^T \). Then, the signal can be written as an all-pole model \[3\]:
\[
\hat{\tau}_m^{(n)} = \sum_{k=1}^{K} \hat{\alpha}'_{k,m}p_{k,m,n}. \tag{12}
\]

In addition, another advantage of root-MUSIC algorithm is that the phase information can be extracted from the roots to obtain an estimation of those non-ideal factors. By comparing the terms in equations (2) and (12), the difference between the random initial phases of the two bands can be estimated as follows:
\[
\hat{\phi}_m - \hat{\phi}_1 = \frac{1}{K} \sum_{k=1}^{K} \left[ \arg (\hat{\alpha}'_{k,m}) - \arg (\hat{\alpha}'_{1,m}) \right] - 2\pi f_{c1}(\hat{\tau}_{k,1}) + 2\pi f_{c2}(\hat{\tau}_{k,m}). \tag{13}
\]
Then we can also get
\[
\delta_m = \frac{1}{K} \sum_{k=1}^{K} \left[ \arg (p_{k,m}) - 2\pi f_{s,m} - \hat{\tau}_k \right]. \tag{14}
\]

Therefore, in the coarse estimation stage, we can obtain coarse estimate of delay \( \hat{\tau}_k \), complex gain amplitude \( ||\hat{\alpha}_k|| \) (i.e. \( \sum_{m=1}^{M} SNR_m ||\hat{\alpha}_k|| \)), difference between the random initial phases \( \hat{\phi}_m \) and timing synchronization error \( \delta_m \) to serve the refined estimation.

**B. Refined Estimation**

In the stage of refined estimation, we first absorb the term \( e^{-j2\pi f_{c,m} \delta_m} \) into the initial phase \( e^{j\phi_m} \), namely \( e^{j\phi_m} = e^{j\phi_m - 2\pi f_{c,m} \delta_m} \), which also aims to reduce the oscillation degree of the likelihood function in the estimation of \( \delta_m \).

Then, if we take the first frequency band as a reference, the rewritten initial phase \( e^{j\phi_1} \) and the carrier phase \( e^{-j2\pi f_{c1} \tau_k} \) can be absorbed into the complex scalar \( \alpha_k \), and the residual band gap term \( e^{-j2\pi (f_{c_2,m} - f_{c_1,m}) \tau_k} \) is retained in each sub-band signal, which is therefore called the band gap aperture structure.

- Signal model with the band gap aperture structure:
\[
r_{m}^{(n)}(\cdot) = \sum_{k=1}^{K} \alpha_k e^{j\phi_1 - j2\pi f_{c1} \tau_k} \frac{\beta_k}{f_{c,m}} e^{-j2\pi (f_{c,m} + f_{s,m}) \tau_k} e^{j\phi_m - j2\pi f_{s,m} \delta_m} + u_{m}^{(n)}, \tag{15}
\]
where
\[
\alpha_k' = \alpha_k e^{j\phi_1 - j2\pi f_{c1} \tau_k}, \quad \phi_m' = \phi_m - \phi_1, \quad f_{c_2,m}' = f_{c_2,m} - f_{c_1,m}, \quad \phi_{m}' = 0, \quad f_{c,m}' = 0.
\]

In this case, we can exploit the multi-band gain (i.e., the phase rotation caused by the gap between the carrier frequencies) to improve the performance.

After obtaining the above refined estimation signal model, prior to presenting algorithm details, the probability model and objective function are introduced at first.

First of all, the coarse estimate of delay \( \hat{\tau}_k \) can be used as the prior information for the refined estimation stage, assuming that the truth value is evenly distributed in the neighborhood of the coarse estimate \( [\hat{\tau}_k - \Delta \hat{\tau}_k/2, \hat{\tau}_k + \Delta \hat{\tau}_k/2] \). The interval of refined estimation \( \Delta \hat{\tau}_k \) can be determined according to the empirical error of the first stage or based on the Cramér-Rao bound (CRB) analysis. The prior probability distributions for \( \phi'_m \) and \( \delta_m \) are similar. Although a relatively accurate estimate of the amplitude of the complex scalar \( ||\alpha_k|| \) is obtained in the first stage, its phase is still unknown. It may be assumed that the amplitude follows a Gaussian distribution with a small variance near the coarse estimate, and the phase is uniformly distributed from 0 to 2\( \pi \).

In summary, vectorized variables to be estimated in the refined stage are denoted as
\[
\Lambda = \left[ \alpha_1', \ldots, \alpha_K', \tau_1, \ldots, \tau_K, \beta_1, \ldots, \beta_K, \phi_2^{'}, \ldots, \phi_M^{'}, \delta_2, \ldots, \delta_M \right], \tag{16}
\]

the total number of variables is denoted as \( |\Lambda| = J \), and frequency domain measurement of the received signal is denoted as
\[
r = \left[ r_1^{(0)}, r_1^{(1)}, \ldots, r_{N_1-1}^{(0)}, r_{N_1}^{(1)}, \ldots, r_{M}^{(0)}, r_{M}^{(1)}, \ldots, r_{M}^{(N_M-1)} \right]^T. \tag{17}
\]

In the case of AWGN, the logarithmic likelihood function can be written as follow
\[
\ln p(r|\Lambda) = \ln \prod_{m=1}^{M} \prod_{n=0}^{N_m-1} p(r_{m}^{(n)}|\Lambda) = MN_m \ln \frac{1}{\sqrt{2\pi} \hat{\eta}_w} - \sum_{m=1}^{M} \sum_{n=0}^{N_m-1} \frac{1}{2\hat{\eta}_w^2} \left| r_{m}^{(n)} - s_{m}^{(n)}(\Lambda) \right|^2, \tag{18}
\]

\[
s_{m}^{(n)}(\Lambda) = \sum_{k=1}^{K} \alpha_k' \left( j \frac{f_{c,m} + n f_{s,m}}{f_{c,m}} \right) \beta_k e^{-j2\pi (f_{c,m} + n f_{s,m}) \tau_k} e^{j\phi_m - j2\pi f_{s,m} \delta_m}, \tag{19}
\]

where \( s_{m}^{(n)}(\Lambda) \) is the received signal reconstructed from the parameter \( \Lambda \).

The joint posterior probability can be easily deduced from the Bayes equation \( p(\Lambda|r) \propto p(r|\Lambda) p(\Lambda) \), but the marginal posterior probability is not. It is inevitable to integrate other variables except \( \Lambda_j \) to obtain the closed-form expression of the marginal posterior probability \( p(\Lambda_j|r) \). \( \Lambda_j \) represents the \( j \)-th variable to be estimated in \( \Lambda \). In general, closed-form solutions are very difficult to obtain, and the computational complexity is often unacceptable.
Algorithm 1 Two-stage Estimation Scheme

Input: received signals r, multi-band frequency settings.

Stage 1: Coarse Estimation
Construct signal model with the bandwidth aperture structure;
Perform weighted root-MUSIC and LS method;
Provide coarse estimates and prior intervals for Stage 2.

Stage 2: Refined Estimation
Construct signal model with the band gap aperture structure;
Perform SPVBI algorithm to obtain the approximate marginal posterior probability.

Output: Obtain the MAP/MMSE estimates of each variable based on approximate marginal posterior probability.

In summary, we provide a two-stage estimation scheme as shown in Algorithm 1, which divides the estimation process into coarse estimation stage and refined estimation stage.

IV. STOCHASTIC PARTICLE-BASED VARIATIONAL BAYESIAN INFERENCE

A. Problem Formulation based on Variational Bayesian Inference

To obtain the marginal posterior probability distribution of delay, we resort to variational Bayesian inference [36], which can asymptotically approximate the real posterior probability distribution \( p(\Lambda_j|r) \) by iteratively updating the variational probability distribution \( q(\Lambda_j) \).

The expression given by the conventional VBI algorithm contains multiple integral calculation (i.e., expectation), so it is usually intractable to obtain the closed-form expression. To deal with this, most of the existing works make some prior assumptions, such as assuming that the distribution of these variables comes from some distribution families [22] or meets the conjugation condition, but this is usually not accurate enough and subjective.

In [24], the authors proposed a particle-based VBI (PVBI) algorithm to approximate the calculation of expectation by means of importance sampling (IS) method [37], which is a kind of Monte Carlo method. By iteratively updating the weight of particles, the discrete distribution composed of particles can gradually approach the posterior probability distribution. However, based on the conclusion formula ([24], formula (12)) given by VBI, the position of particles cannot be updated. Besides, only a large number of particles can overcome the instability caused by initial random sampling, and ensure that the estimation locally converges to a “good” stationary point. Although parallel computing can speed up the processing, the complexity will rocket as the number of variables and particles increases.

Motivated by the above analysis, we also optimize the particle position and further design a SPVBI algorithm to solve the new problem with much lower per-iteration complexity than the conventional PVBI algorithm.

In the proposed PVBI problem formulation, the marginal posterior probability of the \( j \)-th variable is approximated as the weighted sum of \( N_p \) discrete particles:

\[
q(\Lambda_j;x_j,y_j) = \sum_{p=1}^{N_p} y_{j,p} \delta(\Lambda_j-x_{j,p}), \quad (20)
\]

where \( \delta(\cdot) \) is the impulse function, and \( x_j = [x_{j,1},x_{j,2},...,x_{j,N_p}]^T \) and \( y_j = [y_{j,1},y_{j,2},...,y_{j,N_p}]^T \) are the positions and weights of the particles, respectively. Furthermore, according to the mean field assumption [38]–[39], the approximate posterior probability of each variable can be assumed to be independent of each other, \( q(\Lambda;x,y) = \prod_{j=1}^{J} q(\Lambda_j;x_j,y_j) \), where \( x = [x_1;x_2;...;x_J] \) and \( y = [y_1;y_2;...;y_J] \). Therefore, the positions and weights \( x,y \) should be chosen to minimize the Kullback-Leibler (KL) divergence between the variational probability distribution \( q(\Lambda;x,y) \) and the real posterior probability distribution \( p(\Lambda|r) \) [36], which is defined as

\[
D_{KL}[q\|p] = \int q(\Lambda;x,y) \ln \frac{q(\Lambda;x,y)}{p(\Lambda r)} d\Lambda \quad (21)
\]

Considering that \( p(r) \) is a constant independent of \( q(\Lambda;x,y) \), minimizing the KL divergence is equivalent to solving the following optimization problem:

\[
P: \min_{x,y} L(x,y) \triangleq \int q(\Lambda;x,y) \ln \frac{q(\Lambda;x,y)}{p(\Lambda r)p(\Lambda)} d\Lambda
\]

s.t. \( \sum_{p=1}^{N_p} y_{j,p} = 1, \quad \epsilon \leq y_{j,p} \leq 1, \quad \forall j, p, \)

\( \hat{\Lambda}_j - \Delta \hat{\Lambda}_j/2 \leq x_{j,p} \leq \hat{\Lambda}_j + \Delta \hat{\Lambda}_j/2, \quad \forall j, p, \)

where \( \hat{\Lambda}_j \) is the coarse estimation obtained in the first stage and \( \Delta \hat{\Lambda}_j \) is the range of the estimation error of coarse estimation, and \( \epsilon > 0 \) is a small number. The normalized weight \( y_{j,p} \) represents the probability that the particle is located at position \( x_{j,p} \). Note that it does not make sense to generate particles with very small probabilities in approximate posterior since these particles contribute very little to the MAP/MMSE estimator. Therefore, we restrict the probability of each particle \( y_{j,p} \) to be larger than a small number \( \epsilon \).

The truth value is highly likely to be located in the prior interval \( [\hat{\Lambda}_j - \Delta \hat{\Lambda}_j/2, \hat{\Lambda}_j + \Delta \hat{\Lambda}_j/2] \), so the particle position is searched wherein to accelerate the convergence rate. Initial particles can be generated by sampling according to the prior distribution obtained in the coarse estimation stage.

Adding the optimization of particle positions \( x \) can improve the effectiveness of characterizing the target distribution by discrete particles, and avoid the estimation result falling into the local optimum due to poor initial sampling. Furthermore,
Particle position updating can reduce the number of sampled particles, and thus effectively reduce the computational overhead, which will be discussed in detail in the following sections.

B. SPVBI Algorithm Design based on Block SSCA

Although the particle-based approximation can effectively simplify the integral operation in the objective function $L(x,y)$ of $P$, the number of summations in $L(x,y)$ is still as high as $N_p^J$, that is, with exponential computational complexity. To solve this problem, we propose a stochastic particle-based variational Bayesian inference (SPVBI) algorithm based on the block SSCA to find stationary points of $P$ with lower computational complexity.

Specifically, we divided the optimization variables into $2J$ blocks $x_1$, $y_1$, $x_2$, $y_1$, ..., $x_J$, $y_J$. Starting from an initial point $x^{(0)}$, $y^{(0)}$, the SPVBI algorithm alternatively optimizing each block until convergence. Let $x_j^{(t)}$, $y_j^{(t)}$, and $x_j^{(t+1)}$, $y_j^{(t+1)}$ denote the blocks $x_j$, $y_j$ before and after the update in the $t$-th iteration, respectively. Then in the $t$-th iteration, the $(2j-1)$-th block $x_j$ is updated by solving the following subproblem:

$$
P_{x_j} : \min_{x_j} L_j^{(t)} (x_j, y_j^{(t)})
\quad \text{s.t.} \quad \hat{\Lambda}_j - \Delta \hat{\Lambda}_j / 2 \leq x_{j,p} \leq \hat{\Lambda}_j + \Delta \hat{\Lambda}_j / 2, \forall p
$$

where $L_j^{(t)} (x_j, y_j) = L (x^{(t)}, y^{(t)})$ with $x^{(t)} = [x_1^{(t)}, \ldots, x_{j-1}^{(t)}, x_{j+1}^{(t)}, \ldots, x_J^{(t)}]$ and $y^{(t)} = [y_1^{(t)}, \ldots, y_{j-1}^{(t)}, y_{j+1}^{(t)}, \ldots, y_J^{(t)}]$. In other words, $L_j^{(t)} (x_j, y_j)$ is the objective function $L (x,y)$ when fixing all other variables as the latest iterate and only treating $x_j$, $y_j$ as variables. It can be shown that

$$L_j^{(t)} (x_j, y_j) = \sum_{p=1}^{N_p} y_{j,p} \ln y_{j,p} - \sum_{p=1}^{N_p} y_{j,p} \ln p (x_{j,p}, p) + \sum_{p=1}^{N_p} \sum_{p_j=1}^{N_p} \sum_{p_{j+1}=1}^{N_p} \ldots \sum_{p_J=1}^{N_p} \bar{y}_{j-p} \ln p (r | x_{j-p}, x_{j+p}), \tag{22}
$$

where $\bar{y}_{j-p} = \prod_{i \neq j} y_{i,p}$, and $\tilde{x}_{j-p} = \{x_{i,p} \}_{i \neq j}$, which involves a summation of $N_p^{J-1}$ terms. As such, the complexity of directly solving $P_{x_j}$ is unacceptable.

To overcome this challenge, we reformulate $P_{x_j}$ to a stochastic optimization problem as

$$P_{x_j} : \min_{x_j} L_j^{(t)} (x_j, y_j^{(t)}) \triangleq \mathbb{E}_{\Lambda_{\omega_j}} [g_j^{(t)} (x_j, y_j; \Lambda_{\omega_j})]
\quad \text{s.t.} \quad \hat{\Lambda}_j - \Delta \hat{\Lambda}_j / 2 \leq x_{j,p} \leq \hat{\Lambda}_j + \Delta \hat{\Lambda}_j / 2, \forall p,
$$

where $\Lambda_{\omega_j}$ represents all the other variables except $\Lambda_j$ and $\mathbb{E}_{\Lambda_{\omega_j}} [\cdot]$ represents the expectation operator over the probability distribution of variable $\Lambda_{\omega_j}$, and

$$g_j^{(t)} (x_j, y_j; \Lambda_{\omega_j}) = \sum_{p=1}^{N_p} y_{j,p} \ln y_{j,p} - \sum_{p=1}^{N_p} y_{j,p} \ln p (x_{j,p}, p) \times [\ln p (x_{j,p}) + \ln p (r | \Lambda_{\omega_j}, x_{j,p})], \tag{23}
$$

Then following the idea of SSCA, we replace the objective function $L_j^{(t)} (x_j, y_j^{(t)})$ in $P_{x_j}$ with a simple quadratic surrogate objective function

$$T_{x_j}^{(t)} (x_j) \triangleq (f_j^{(t)})^T (x_j - x_j^{(t)}) + \Gamma_{x_j} \| x_j - x_j^{(t)} \|^2, \tag{24}
$$

and obtain an intermediate variable $x_j^{(t)}$ by minimizing the surrogate objective function as

$$x_j^{(t)} = \arg \min_{x_j} T_{x_j}^{(t)} (x_j)
\quad \text{s.t.} \quad \hat{\Lambda}_j - \Delta \hat{\Lambda}_j / 2 \leq x_{j,p} \leq \hat{\Lambda}_j + \Delta \hat{\Lambda}_j / 2, \forall p,
$$

where $\Gamma_{x_j}$ can be any positive number, $f_j^{(t)}$ is an unbiased estimator of the gradient $\nabla_{x_j} L_j^{(t)} (x_j, y_j)$, which is updated recursively as follows:

$$f_j^{(t)} = (1 - \rho^{(t)} ) f_j^{(t-1)} + \frac{\rho^{(t)}}{B} \sum_{b=1}^{B} \nabla_{x_j} g_j^{(t)} (x_j^{(t)}, y_j^{(t)}; \Lambda_{\omega_j}^{(b)}), \tag{26}
$$

where $\{ \Lambda_{\omega_j}^{(b)} ; b = 1, \ldots, B \}$ is a mini-batch of $B$ samples generated by the distribution $\prod_{j' \neq j} q (\Lambda_{j'}; x_{j'}, y_{j'}$) with $x_{j'} = x_{j'}^{(t+1)}$, $y_{j'} = y_{j'}^{(t+1)}$, $\forall j' < j$ and $x_{j'} = x_{j'}^{(t)}$, $y_{j'} = y_{j'}^{(t)}$, $\forall j' > j$, and $\rho^{(t)}$ is a decreasing step size that will be discussed later and we set $f_{x_j}^{(0)} = 0$. Finally, the updated $x_j$ is given by

$$x_j^{(t+1)} = (1 - \gamma^{(t)} ) x_j^{(t)} + \gamma^{(t)} x_j^{(t)}, \tag{27}
$$

where $\gamma^{(t)}$ is another decreasing step size that will be discussed later, and there is a closed-form solution for $x_j^{(t)} \triangleq [\bar{x}_{j,1}^{(t)}, \bar{x}_{j,2}^{(t)}, \ldots, \bar{x}_{j,N_p}^{(t)}]^T
\begin{align}
\bar{x}_{j,p}^{(t)} = \left\{ \begin{array}{ll}
\hat{\Lambda}_j - \Delta \hat{\Lambda}_j / 2, & \bar{x}_{j,p}^{(t)} < \hat{\Lambda}_j - \Delta \hat{\Lambda}_j / 2,
\hat{\Lambda}_j - \Delta \hat{\Lambda}_j / 2 \leq \bar{x}_{j,p}^{(t)} < \hat{\Lambda}_j + \Delta \hat{\Lambda}_j / 2, \\
\hat{\Lambda}_j + \Delta \hat{\Lambda}_j / 2, & \bar{x}_{j,p}^{(t)} \geq \hat{\Lambda}_j + \Delta \hat{\Lambda}_j / 2,
\end{array} \right.
\end{align}

p = 1, ..., N_p,
\text{where} \bar{x}_j^{(t)} \triangleq \left[ \bar{x}_{j,1}^{(t)}, \bar{x}_{j,2}^{(t)}, \ldots, \bar{x}_{j,N_p}^{(t)} \right]^T = x_j^{(t)} - \frac{1}{\mathbb{E}_{x_j} f_{x_j}^{(t)}}.
$$

Similarly, in the $t$-th iteration, the $(2j)$-th block $y_j$ is updated by solving the following subproblem:

$$P_{y_j} : \min_{y_j} L_j^{(t)} (x_j^{(t+1)}, y_j)
\quad \text{s.t.} \quad \sum_{p=1}^{N_p} y_{j,p} = 1, \quad \epsilon \leq y_{j,p} \leq 1, \forall p.
$$

In stead of solving $P_{y_j}$ directly, we first construct a simple quadratic surrogate objective function

$$T_{y_j}^{(t)} (y_j) \triangleq (f_{y_j}^{(t)})^T (y_j - y_j^{(t)}) + \Gamma_{y_j} \| y_j - y_j^{(t)} \|^2, \tag{29}
$$
and obtain an intermediate variable $\mathbf{y}^{(t)}_j$ by minimizing the surrogate objective function as

$$\mathbf{y}^{(t)}_j = \arg\min \mathcal{L}^{(t)}_{y'_j}(\mathbf{y}_j)$$  \hspace{1cm} (30)

subject to $\sum_{p=1}^{N_{\rho}} y_{j,p} = 1$, $\epsilon \leq y_{j,p} \leq 1$, $\forall p$,

where $\Gamma_{y_j}$ can be any positive number, $f^{(t)}_{y'_j}$ is an unbiased estimator of the gradient $\nabla_{\mathbf{y}'} L^{(t)}_{y'_j}(x^{(t-1)}_j, y^{(t)}_j)$, which is updated recursively as follows

$$f^{(t)}_{y'_j} = \left(1 - \rho^{(t)}\right) f^{(t-1)}_{y'_j} + \frac{\rho^{(t)}}{B} \sum_{b=1}^{B} \nabla_{y'_j} g^{(t)}_{y_j}(x^{(t+1)}_j, y^{(t)}_j; \Lambda^{b}_{y_j})$$ \hspace{1cm} (31)

and we set $f^{(t-1)}_{y'_j} = 0$. Finally, the updated $y_j$ is given by

$$y_j^{(t+1)} = \left(1 - \gamma^{(t)}\right) y_j^{(t)} + \gamma^{(t)} f^{(t)}_{y'_j}.$$ \hspace{1cm} (32)

To ensure the convergence of the algorithm, the step sizes $\rho^{(t)}$ and $\gamma^{(t)}$ must satisfy the following conditions.

**Assumption 1 (Assumptions on step sizes).**

1. $\rho^t \rightarrow 0$, $\sum_t \rho^t = \infty$, $\sum_t (\rho^t)^2 < \infty$.
2. $\lim_{t \rightarrow \infty}\gamma^t / \rho^t = 0$.

A typical choice of $\rho^t, \gamma^t$ that satisfies Assumption 1 is $\rho^t = \mathcal{O}(t^{-\kappa_1}), \gamma^t = \mathcal{O}(t^{-\kappa_2})$, where $0.5 < \kappa_1 < \kappa_2 \leq 1$.

---

**C. Summary of the SPVBI Algorithm**

The overall SPVBI algorithm is summarized in Algorithm 2. The mini-batch size $B$ can be chosen to achieve a trade-off between the per-iteration complexity and the total number of iterations. Thanks to the idea of averaging over iterations as in (26), (27), (31) and (32), a constant value of the mini-batch size $B$ is usually sufficient to achieve a fast convergence, e.g., in the simulations, we set $B = 10$. Compared to the conventional PVBI algorithm which needs to calculate a summation of $N^J_{t-1}$ terms when updating one block in each iteration, the proposed SPVBI only requires to solve a simple quadratic programming problem which only involves the calculation of $B$ gradients, where $B$ can be much smaller than $N^J_{t-1}$. Moreover, the addition of updating particle position allows the algorithm to converge more quickly and flexibly to stable solutions. As such, the proposed SPVBI algorithm has both lower per-iteration complexity and faster convergence speed than the conventional PVBI.

The proposed SPVBI is an extension of the SSCA framework in [30]. There are two key differences. First, the SSCA in [30] constructs a single surrogate function to update all the variables simultaneously in each iteration, while the SPVBI allows block-wise update, which is often used in VBI-type algorithms [36]. Second, the distribution of the random state in the original SSCA framework is assumed to be independent of the optimization variable, i.e., is control independent, while the distribution of the random state in the SPVBI is control dependent. For example, the random state $\Lambda_{x,j}$ in the $t$-th iteration follows the distribution $\prod_{j \in \mathcal{J}_x} q(\Lambda_{x,j} ; x'_j, y'_j)$, which depends on the value of the current optimization variables and is changing over iterations. Therefore, SPVBI can be viewed as an extension of the SSCA framework in [30] to enable block-wise update and control-dependent random state. As such, the convergence analysis is also more challenging. In the next subsection, we shall provide the convergence analysis for SPVBI.

**D. Convergence Analysis**

In this section, we will prove the convergence of the SPVBI algorithm. First, we present a key lemma which gives some important properties of the surrogate functions.

**Lemma 1 (Properties of the surrogate functions).** For each iteration $t = 1, 2, \ldots$ and each block $x_j, y_j, j = 1, 2, \ldots, J$, we have

1. $\mathcal{L}^{(t)}_{y'_j}(x_j)$ and $\mathcal{L}^{(t)}_{y'_j}(y_j)$ are uniformly strongly convex in $x_j$ and $y_j$, respectively.
Algorithm 2 Stochastic Particle-based VBI Algorithm

Input: received signals r, number of scattering centers K, multi-band frequency settings, prior information for Λ, step-size sequences \{ρ(t)\} and \{γ(t)\}.

Initialization: The initial particle is generated according to the prior information obtained by coarse estimation, and the initial weight is set as \(1/N_p, \{x_{j,p}, y_{j,p}\}_{p=1}^{N_p} : j = 1, 2, \ldots, J\).

While not converge do \((t \rightarrow \infty)\)

For \(j = 1 : J\)

Generate a mini-batch of realization based on the \(\{x_{i,p}, y_{i,p}\}_{p=1}^{N_p} : i \neq j\);

Update \(f^{(t)}_{x_j}\) or \(f^{(t)}_{y_j}\);

Solving surrogate optimization problem \(P_{x_j}\) or \(P_{y_j}\);

Update the particle position and weight of variable \(\Lambda_j\);

end

Output: Find the position \(\tilde{x}_j, \tilde{y}_j = 1, 2, \ldots, J\) of the particle with the maximum weight \(y^{(\text{max})}_j\), which is the MAP estimate.

2) For any \(x_j \in X\) and \(y_j \in Y\), the function \(T^{(t)}_{x_j}(x_j)\) and \(T^{(t)}_{y_j}(y_j)\), their derivative, and their second order derivative are uniformly bounded.

3) \(T^{(t)}_{x_j}(x_j)\) and \(T^{(t)}_{y_j}(y_j)\) are Lipschitz continuous function w.r.t. \(x_j\) and \(y_j\), respectively, Moreover,

\[
\limsup_{t_1, t_2 \rightarrow \infty} \left| T^{(t_1)}_{x_j}(x_j) - T^{(t_2)}_{x_j}(x_j) \right| \leq B_x \left\| x^{(t_1)} - x^{(t_2)} \right\|, \\
\limsup_{t_1, t_2 \rightarrow \infty} \left| T^{(t_1)}_{y_j}(y_j) - T^{(t_2)}_{y_j}(y_j) \right| \leq B_y \left\| y^{(t_1)} - y^{(t_2)} \right\|, \\
\forall x_j \in X, \forall y_j \in Y, \text{ for some constants } B_x > 0 \text{ and } B_y > 0.
\]

4) Consider a subsequence \(\{x^{(t_i)}, y^{(t_i)}\}_{i=1}^{\infty}\) converging to a limit point \(x^*, y^*\). There exist uniformly differentiable functions \(\tilde{f}_{x_j}(x_j)\) and \(\tilde{f}_{y_j}(y_j)\) such that

\[
\lim_{t \rightarrow \infty} T^{(t)}_{x_j}(x_j) = \tilde{f}_{x_j}(x_j), \forall x_j \in X, \\
\lim_{t \rightarrow \infty} T^{(t)}_{y_j}(y_j) = \tilde{f}_{y_j}(y_j), \forall y_j \in Y.
\]

Moreover, we have

\[
\left\| \nabla x \tilde{f}_{x_j}(x_j^*) - \nabla x L(x^*, y^*) \right\| = 0, \\
\left\| \nabla y \tilde{f}_{y_j}(y_j^*) - \nabla y L(x^*, y^*) \right\| = 0.
\]

Please refer to Appendix A for the proof. With the Lemma 1, we are ready to prove the following main convergence result.

Theorem 2 (Convergence of SPVBI). Starting from a feasible initial point \(x^{(0)}, y^{(0)}\), let \(\{x^{(t)}, y^{(t)}\}_{t=1}^{\infty}\) denote the iterates generated by Algorithm 2. Then every limiting point \(x^*, y^*\) of \(\{x^{(t)}, y^{(t)}\}_{t=1}^{\infty}\) is a stationary point of original problem \(P\) almost surely.

Please refer to Appendix B for the proof.

E. Complexity Analysis

In this subsection, we analyze the complexity of SPVBI algorithm. In the coarse estimation stage, weighted root-MUSIC algorithm is adopted to narrow down the range, which mainly includes subspace decomposition and polynomial rooting, with the complexity of \(O(MN_m^3)\) and \(O(M(2N_m - 1)^3)\), respectively. In addition, the computational complexity of LS method is approximately \(O(N_mK^2 + K^3)\).

In the refined estimation stage, for PVBI algorithm, as the number of associated variables and particles increases, the computational complexity increases exponentially. Its per-iteration complexity order is \(O(J(N_p)^{-1}N_{LH})\), where \(N_{LH} = O(KMN_m)\) represents the average number of floating point operations (FLOPs) required to compute the dominant logarithmic likelihood value. Through mini-batch sampling and minimization of quadratic surrogate objective functions, the per-iteration complexity of SPVBI algorithm can be reduced to \(O(2J(N_pBN_{grad} + N_p^3))\), where \(BN_{grad}\) and \(N_p^3\) represent the complexity of computing a mini-batch of gradients and quadratic programming search, respectively, where \(N_{grad} = O(KMN_m)\) represents the average number of FLOPs required to compute a gradient. It can be seen that SPVBI greatly reduces the computational complexity compared with conventional PVBI.

V. Numerical Simulation and Performance Analysis

In this section, simulations are conducted to demonstrate the performance of the proposed algorithm. We compare the proposed algorithm with the following baseline algorithms:

1) Baseline 1 (root MUSIC, R-MUSIC) [3]: The conventional root MUSIC algorithm has relatively high accuracy in delay estimation, so it is adopted here for single-band data, mainly to show the gain brought by combining the results of multiple bands.

2) Baseline 2 (Weighted root MUSIC, WR-MUSIC) [29]: For each individual band, root MUSIC algorithm is adopted, and the estimation results of each band are fused using maximum ratio combination to obtain a more accurate estimation. The WR-MUSIC algorithm is adopted in the coarse estimation stage. See Section III-A for relevant procedures.

3) Baseline 3 (Spectral estimation, SE) [3]: The parameters of the approximate all-pole model are estimated by spectral search methods such as root MUSIC algorithm, and the unknown frequency band data is interpolated according to the model to improve the resolution.

4) Baseline 4 (SBL) [26]: The information required for coherent compensation can be extracted from the SBL solution. Then, by interpolating data between non-contiguous bands, more accurate estimate can be obtained, where the number of atoms in the dictionary is set to be 5000.

In the proposed SPVBI, each variable is equipped with 10 particles, and the size of mini-batch \(B\) is 10. The step size sequence is set as follows: \(\rho(t) = 5/(5 + t)^{0.9}, \rho(0) = 1\);
\[ \gamma^{(t)} = 15/(15 + t)^1, \gamma^{(0)} = 1. \] Unless otherwise specified, the experiment was repeated 400 times.

In the simulations, we consider both narrow-band and wide-band scenarios. We first describe the common setup for both scenarios. There are two scattering centers. The amplitude of \( \alpha_k \) is 1 and 0.5, and \( \beta_k \) is \(-0.5\) and 0.5, respectively. In addition, the phase of \( \alpha_k \) and initial phase \( \phi_m \) are uniformly generated within \([0, 2\pi]\).

In the narrow-band scenario, the received signals come from two non-adjacent frequency bands with a bandwidth of 40MHz, and the frequency step is 0.8MHz. The initial frequency is set to 2.4GHz and 2.54GHz, respectively. There are two scattering centers with delays of 50ns and 500ns. The timing synchronization error \( \delta_m \) is generated following a Gaussian distribution \( \mathcal{N}(0, 0.1ns^2) \).

In the wide-band scenario, the bandwidth of the two non-contiguous frequency bands is 0.5GHz, and the frequency step is 20MHz. The initial frequency is set to 15GHz and 16GHz, respectively. The reference delays of the two scattering centers are 10ns and 20ns, respectively. The timing synchronization error \( \delta_m \) is generated following a Gaussian distribution \( \mathcal{N}(0, 1ns^2) \).

### A. Analysis of Convergence Performance

In this section, we will focus on the performance of convergence. The conventional PVBI algorithm is too complicated to be simulated under the scenario of this paper (i.e., too many variables need to be estimated). Therefore, we will focus on the benefits of updating particle positions by comparing the proposed SPVBI with and without updating particle positions.

#### Impact of the SNR:

As can be seen in Fig 6 and Table I, when the SNR increases from 12dB to 15dB, the performance of all algorithms is improved. At a certain SNR, the CDF curve of SPVBI is more inclined to the upper left than that of WR-MUSIC, indicating that the estimation error is smaller in most cases. The performance of the SBL algorithm for delay estimation is inferior to WR-MUSIC, even when the dictionary size is already quite large (i.e. 5000 atoms).

#### Impact of the band gap:

This subsection presents the impact of the band gap. We changed the initial frequency of the second band from 2.54GHz to 2.49GHz, so the band gap reduces to 50MHz.

In Fig 7 and Table II, with the increase of frequency band gap, the performance of other algorithms remains almost unchanged except that of SPVBI algorithm is further improved.

#### Impact of the SNR:

As can be seen in Fig 6 and Table I, when the SNR increases from 12dB to 15dB, the performance of all algorithms is improved. At a certain SNR, the CDF curve of SPVBI is more inclined to the upper left than that of WR-MUSIC, indicating that the estimation error is smaller in most cases. The performance of the SBL algorithm for delay estimation is inferior to WR-MUSIC, even when the dictionary size is already quite large (i.e. 5000 atoms).

#### Impact of the band gap:

This subsection presents the impact of the band gap. We changed the initial frequency of the second band from 2.54GHz to 2.49GHz, so the band gap reduces to 50MHz.

In Fig 7 and Table II, with the increase of frequency band gap, the performance of other algorithms remains almost unchanged except that of SPVBI algorithm is further improved.

1SE algorithm is designed to improve the resolution by reconstructing the full-band data based on the estimated parameters from the R-MUSIC algorithm, and its delay estimation accuracy is consistent with R-MUSIC.
which indicates that the proposed algorithm does utilize the multi-band gap aperture gain.

\[
\text{Impact of the factor } \delta_m: \text{ Next comes the impact of timing synchronization error } \delta_m. \text{ In Fig 8 and Table III, as the variance of } \delta_m \text{ increases, the performance of all algorithms degrades significantly.}
\]

\[
C. \text{ Resolution Performance}
\]

In the application scenarios of wide-band radar, such as radar imaging and target feature extraction, ultra-high resolution is required. In this case, it is often desirable to reconstruct the full-band data from the available non-adjacent band data to improve the resolution. However, whether high resolution can be achieved will depend on the accuracy of the data reconstruction. Therefore, for the SPVBI algorithm and Baseline 3 and 4, we compare the root-mean-square error (RMSE) of data reconstruction to indirectly show the resolution performance$^2$.

The RMSE between the estimated full-band data and the true full-band data can be calculated via the following equation:

\[
\text{RMSE} = \sqrt{\frac{\sum_{n=1}^{N} (r(n) - \hat{r}(n))^2}{N}},
\]

where $N$ indicates the number of frequency points in the full-band.

\[
\text{In Fig 9, we show the RMSE of full-band data reconstruction under different SNR. In the case of low SNR, SE algorithm is less affected by noise due to its simple model and few parameters to be estimated. However, with the increase of SNR, compared with other multi-band fusion algorithms, the full-band data reconstructed by SPVBI is more accurate, which implies that the proposed algorithm can obtain more accurate estimation of the signal parameters. The RMSE performance of SBL algorithm is worse than that of SPVBI and its complexity is higher.}
\]

\[
\text{In addition, Fig 10 shows the high resolution range profile (HRRP) reconstructed by full-band data of different algorithms. It can be found that the HRRP of the multi-band fusion algorithm (i.e. SE, SBL, SPVBI) is narrower than that of the single-band reconstruction, so the resolution is higher. Meanwhile, the RMSE of the full-band data reconstructed by SPVBI algorithm is smaller, so the peak point of SPVBI is closer to the true value as can be seen from the rectangular box.}
\]

$^2$The (weighted) root MUSIC algorithm can achieve a good delay estimation accuracy, but it cannot obtain a good estimation of all parameters to reconstruct the full-band data. Therefore, we do not compare with Baseline 1 and 2 in the wide-band scenario.
D. Comparison of Computational Complexity

In Table IV, we analyze the per-iteration complexity order of other mainstream multi-band radar sensing algorithms, and then compare them numerically with SPVBI algorithm under a typical setting. In the SE algorithm, there is a nonlinear least square fitting step, which is solved by Levenberg-Marquarell (LM) algorithm with a complexity order of $O\left(\mathcal{N}_m^2\right)$. In the SBL algorithm, $M_{ob}$ is the number of atoms. Typical Settings are as follows: $J = 8$, $N_p = 10$, $M = 2$, $N_m = 50$, $B = 10$, $N_{LH} = 3900$, $N_{\text{grad}} = 1500$, $M_{ob} = 5000$.

| Algorithms | Complexity order per iteration | Typical values |
|------------|--------------------------------|----------------|
| PVBI       | $O\left(J \cdot (N_p)^{J-1} \cdot N_{LH}\right)$ | 3.12 $\times$ 10^{11} |
| SPVBI      | $O\left(2J \cdot (N_p B N_{\text{grad}} + N_p^2)\right)$ | 2.416 $\times$ 10^{6} |
| WR-MUSIC   | $O\left(M N_m^3 + M (2N_m - 1)^3\right)$ | 2.19 $\times$ 10^{6} |
| SE         | $O_{\text{WR-MUSIC}} + O\left(N_m^3\right)$ | 2.316 $\times$ 10^{6} |
| SBL        | $O\left(M \cdot M_{ob}\right)$ | 2.5 $\times$ 10^{11} |

It can be found that SPVBI algorithm can achieve good trade-off between performance and complexity. PVBI algorithm can only be applied in scenarios with few variables, and once the number of variables is large, the complexity will be unacceptable. Although the computational complexity of SPVBI is higher than the WR-MUSIC and SE algorithm, the estimation accuracy and resolution are greatly improved. Compared with the SBL algorithm, the complexity of SPVBI is greatly reduced and the performance is also improved.

VI. Conclusions

In this paper, we proposed a novel high-accuracy algorithm for multi-band radar sensing. To overcome the difficulty caused by the oscillation of likelihood function, two signal models transformed from the original multi-band signal model are adopted in a two-stage estimation framework. The coarse estimation stage helps to reduce the computational complexity by narrowing down the estimation range, and the refined estimation stage can take full advantage of the carrier phase information between different frequency bands (i.e., multi-band gap gain) to further improve estimation performance. Moreover, the SPVBI algorithm based on block SSCA can transform the computation of expectation with exponential complexity in the conventional PVBI into solving stochastic optimization problems, so that the convergence can be guaranteed theoretically and the computational complexity can be greatly reduced through mini-batch random sampling and averaging over iterations. Simulation results show that the proposed algorithm achieved good performance with acceptable complexity in different scenarios, and adding the particle position update can speed up convergence and reduce the number of particles required. It is worth mentioning that the proposed SPVBI framework has good generalization ability and is expected to be applied to more general parameter estimation scenarios.

APPENDIX

A. Proof of Lemma 1

We first introduce the following preliminary result.

**Lemma 3.** Given subproblem $\mathcal{P}_{x_j}$ and $\mathcal{P}_{y_j}$ under Lemma 1, suppose that the step sizes $\gamma$ and $\gamma'$ are chosen according to Assumption 1. Let $\{x^{(t)}, y^{(t)}\}$ be the sequence generated by Algorithm 2. Then, the following holds

$$\lim_{t \to \infty} \left| f^{(t)}_{x_j} - \nabla x_j L^{(t)}_j \left( x^{(t)}_j, y^{(t)}_j \right) \right| = 0, \text{w.p.1.} \quad (40)$$

$$\lim_{t \to \infty} \left| f^{(t)}_{y_j} - \nabla y_j L^{(t)}_j \left( x^{(t+1)}_j, y^{(t)}_j \right) \right| = 0, \text{w.p.1.} \quad (41)$$

**Proof:** Lemma 3 is a consequence of ([41], Lemma 1). We only need to verify that all the technical conditions therein are satisfied. Specifically, Condition (a) of ([41], Lemma 1) is satisfied because $\mathcal{X}$ and $\mathcal{Y}$ are compact and bounded. Condition (b) of ([41], Lemma 1) follows from the boundedness of the instantaneous gradient $\nabla y_j^{(t)}$. Conditions (c)–(d) immediately come from the step-size rule (1) in Assumption 1. Although the control-dependent random states are not identically distributed over iterations, the distributions (i.e., positions and weights of particles) change slowly at the rate of order $O\left(\gamma\right)$, so we have $\left| \nabla L_j^{(t+1)} - \nabla L_j^{(t)} \right| = O\left(\gamma\right)$. Plusing the step-size rule 2) in Assumption 1, Condition (e) of ([41], Lemma 1) is also satisfied.

Using this result, since $\nabla L_j^{(t)} \left( x^{(t)}_j, y^{(t)}_j \right)$ is obviously bounded, then $f^{(t)}_{x_j}$ and $f^{(t)}_{y_j}$ are bounded. As can be seen, the surrogate function adopted is a convex quadratic function with box constraints. Therefore, 1)-3) in Lemma 2 follow directly from the expression of the surrogate function in (24) and (29).

For 4) in Lemma 2, the proof is similar to that of ([30], Lemma 1). Due to 1)-3) in Lemma 2, the families of functions $\{ \overline{P}_{x_j}(x_j) \}$ and $\{ \overline{P}_{y_j}(y_j) \}$ are equicontinuous. Moreover, they are bounded and defined over a compact set $\mathcal{X}$ and $\mathcal{Y}$. Hence the Arzela–Ascoli theorem [42] implies that, by restricting to a subsequence, there exists uniformly continuous functions $\overline{f}_{x_j}(x_j)$ and $\overline{f}_{y_j}(y_j)$ such that (35) and (36) in Lemma 2-4) are satisfied.
Clearly, we have
\[ \nabla_{x_j/y_j} L(x^*, y^*) = \nabla_{x_j/y_j} L_j(x^*, y^*) \] almost surely. And because of \( \lim_{t \to \infty} f_{x_j}^{(t)}(x_j^*) = \nabla_{x_j} f_{x_j}(x_j^*) \) and Lemma 3, we further have
\[ \left\| \nabla_{x_j} f_{x_j}(x_j^*) - \nabla_{x_j} L(x^*, y^*) \right\| = 0, \] \[ \left\| \nabla_{y_j} f_{y_j}(y_j^*) - \nabla_{y_j} L(x^*, y^*) \right\| = 0. \] (42)
\[ \nabla_{y_j} f_{y_j}(y_j^*) \] and Lemma 3, we have that
\[ \left\| \nabla_{x_j} f_{x_j}(x_j^*) - \nabla_{x_j} L(x^*, y^*) \right\| = 0, \] \[ \left\| \nabla_{y_j} f_{y_j}(y_j^*) - \nabla_{y_j} L(x^*, y^*) \right\| = 0. \] (43)
\[ \left\| \nabla_{y_j} f_{y_j}(y_j^*) - \nabla_{y_j} L(x^*, y^*) \right\| = 0. \] (44)

B. Proof of Theorem 2

It is easy to see that each iteration of Algorithm 2 is equivalent to optimizing the following surrogate function
\[ \tilde{f}^{(t)}(x, y) = \sum_{j=1}^{J} \left[ \tilde{f}_{x_j}^{(t)}(x_j) + \tilde{f}_{y_j}^{(t)}(y_j) \right]. \] (45)
Moreover, from Lemma 1, we have
\[ \lim_{t \to \infty} \tilde{f}^{(t)}(x, y) = \tilde{f}(x, y) \] \[ = \sum_{j=1}^{J} \left[ f_{x_j}(x_j) + f_{y_j}(y_j) \right]. \] (46)
Using Lemma 1 and the similar analysis as in the proof of ([30], Theorem 1), we have that
\[ \left\{ x^*, y^* \right\} = \arg \min \tilde{f}(x, y) \] \[ s.t. \quad h_j(x, y) \leq 0, \quad j = 1, \ldots, J \] \[ H_j(x, y) = 0, \quad j = 1, \ldots, J, \] (47)
where \( h_j(x, y) \) and \( H_j(x, y) \) represent inequality constraints and equality constraints in the original problem, respectively. The KKT condition of problem (47) implies that there exist \( \lambda_1, \ldots, \lambda_J \) and \( \mu_1, \ldots, \mu_J \) that
\[ \nabla \tilde{f}(x^*, y^*) + \sum_{j=1}^{J} \mu_j \nabla h_j(x^*, y^*) + \sum_{j=1}^{J} \lambda_j \nabla H_j(x^*, y^*) = 0 \] \[ h_j(x^*, y^*) \leq 0, \quad h_j(x^*, y^*) = 0, \quad j = 1, \ldots, J \] \[ \mu_j \geq 0, \quad j = 1, \ldots, J \] \[ \mu_j h_j(x^*, y^*) = 0, \quad j = 1, \ldots, J. \] (48)
Finally, it follows from Lemma 1 and (48) that \( \left\{ x^*, y^* \right\} \) also satisfies the KKT condition of Problem \( \mathcal{P} \). This completes the proof.

C. Derivation of Particle Position Gradient

The general formula for the gradient of position is
\[ \frac{\partial g_j^{(t)}(x_j^{(t)}, y_j^{(t)}, \Lambda_{(x_j^{(t)}, y_j^{(t)})})}{\partial x_j^{(t)}} = -y_j^{(t)} \left[ \frac{\partial \ln p}{\partial x_j^{(t)}} + \frac{\partial \ln p}{\partial y_j^{(t)}} \right], \] (49)
where the two terms in brackets are derived for different variables as follows.

1) Complex scalar \( \alpha'_k \):
\[ \frac{\partial g_j^{(t)}(x_j^{(t)}, y_j^{(t)}, \Lambda_{(x_j^{(t)}, y_j^{(t)})})}{\partial \alpha_k} = \begin{cases} \frac{y_{j,p}^{(t)}}{2\eta_{\alpha}} \left[ \sum_{n=0}^{N_{m}-1} h_{k,n,m}^{(p)} \left( r_{m}^{(n)} - \sum_{k=1}^{K} \alpha'^{(t)}_{k} h_{k,n,m}^{(b)} \right) \right]^{*} \end{cases} \] (50)
\[ h_{k,n,m}^{(b)} = e^{i\phi_{m}^{(b)}} \left( \frac{f_{c,m} + n f_{s,m}}{f_{c,m}} \right) \theta_{k}^{(b)} e^{-j2\pi(f_{c,m} + n f_{s,m})r_{m}^{(b)}} e^{-j2\pi(n f_{s,m})\delta_{k}^{(b)}}. \] (51)
where the superscript \( (b) \) indicates that its value is generated by the \( b \)-th sample of a mini-batch.

2) Scattering coefficient \( \beta_k \): \( \beta_k \) only takes some fixed discrete values of \( \{-1, -1/2, 0, 1/2, 1\} \), which in physical sense correspond to the geometry of the scatterer as corner, edge, top, simple curved surface and flat plate [31], so there is no need to optimize the position of the particles here.

3) Random initial phase \( \phi_{m}^{(b)} \):
\[ \frac{\partial g_j^{(t)}(x_j^{(t)}, y_j^{(t)}, \Lambda_{(x_j^{(t)}, y_j^{(t)})})}{\partial \phi_{m,p}^{(b)}} = \begin{cases} \frac{y_{j,p}^{(t)}}{\eta_{\phi}^{2}} \sum_{n=0}^{N_{m}-1} \ln \left( \frac{r_{m}^{(n)} h_{k,n,m}^{(b)} e^{i2\pi f_{c,m} + n f_{s,m} \delta_{k}^{(b)}}}{r_{m}^{(n)} h_{k,n,m}^{(b)} e^{i2\pi f_{c,m} + n f_{s,m} \delta_{k}^{(b)}}} \right) \end{cases} \] (52)
\[ h_{k,n,m}^{(b)} = \sum_{k=1}^{K} \alpha'^{(t)}_{k} \left( \frac{f_{c,m} + n f_{s,m}}{f_{c,m}} \right) \theta_{k}^{(b)} e^{-j2\pi(f_{c,m} + n f_{s,m})r_{m}^{(b)}} e^{-j2\pi(n f_{s,m})\delta_{k}^{(b)}}. \] (53)

4) Time delay offset \( \delta_{m} \):
\[ \frac{\partial g_j^{(t)}(x_j^{(t)}, y_j^{(t)}, \Lambda_{(x_j^{(t)}, y_j^{(t)})})}{\partial \delta_{m,p}} = \begin{cases} \frac{y_{j,p}^{(t)}}{\eta_{\delta}^{2}} \sum_{n=0}^{N_{m}-1} 2\pi(n f_{s,m}) \ln \left( \frac{r_{m}^{(n)} h_{k,n,m}^{(b)} e^{i2\pi f_{c,m} + n f_{s,m} \delta_{k}^{(b)}}}{r_{m}^{(n)} h_{k,n,m}^{(b)} e^{i2\pi f_{c,m} + n f_{s,m} \delta_{k}^{(b)}}} \right) \end{cases} \] (54)
\[ h_{k,n,m}^{(b)} = \sum_{k=1}^{K} \alpha'^{(t)}_{k} \left( \frac{f_{c,m} + n f_{s,m}}{f_{c,m}} \right) \theta_{k}^{(b)} e^{-j2\pi(f_{c,m} + n f_{s,m})r_{m}^{(b)}} e^{-j2\pi(n f_{s,m})\delta_{k}^{(b)}}. \] (55)

5) Delay \( \tau_k \):
\[ \frac{\partial g_j^{(t)}(x_j^{(t)}, y_j^{(t)}, \Lambda_{(x_j^{(t)}, y_j^{(t)})})}{\partial \tau_{k,p}} = \begin{cases} \frac{y_{j,p}^{(t)}}{\eta_{\tau}^{2}} \sum_{n=0}^{N_{m}-1} 2\pi(f_{c,m} + n f_{s,m}) \ln \left( \frac{r_{m}^{(n)} h_{k,n,m}^{(b)} e^{i2\pi(f_{c,m} + n f_{s,m})r_{m}^{(p)}}}{r_{m}^{(n)} h_{k,n,m}^{(b)} e^{i2\pi(f_{c,m} + n f_{s,m})r_{m}^{(p)}}} \right) \end{cases} \] (56)
\[ h_{k,n,m}^{(b)} = \alpha'^{(t)}_{k} \left( \frac{f_{c,m} + n f_{s,m}}{f_{c,m}} \right) \theta_{k}^{(b)} e^{-j2\pi(n f_{s,m})\delta_{k}^{(b)}}. \] (57)
D. Derivation of Particle Weight Gradient

The general formula for the gradient of weight is

$$
\frac{\partial j(t)}{\partial y_{j,p}} \left( x_{j}^{(t+1)} , y_{j}^{(t)} , \Lambda_{j}^{(b)} \right)
$$

$$
= \ln y_{j,p} - \ln p \left( x_{j}^{(t+1)} ; y_{j}^{(t)} , \Lambda_{j}^{(b)} \right) + 1.
$$

(59)

REFERENCES

[1] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, “Integrated sensing and communications: Toward dual-functional wireless networks for 6g and beyond,” IEEE Journal on Selected Areas in Communications, vol. 40, no. 6, pp. 1728–1767, 2022.

[2] W. Jiang, J. Liu, J. Yang, X. Zhang, and W. Li, “A novel multiband fusion method based on a modified relax algorithm for high-resolution and anti-non-gaussian colored clutter microwave imaging,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–12, 2022.

[3] K. Cuomo, J. Pion, and J. Mayhan, “Ultra-wideband coherent processing,” IEEE Transactions on Antennas and Propagation, vol. 47, no. 6, pp. 1094–1107, 1999.

[4] S. Rouquette and M. Najim, “Estimation of frequencies and damping factors by two-dimensional esprit type methods,” IEEE Transactions on Signal Processing, vol. 49, no. 1, pp. 237–245, 2001.

[5] D. Xiong, J. Wang, L. Zhao, Z. Yuan, and M. Gao, “Sub-band mutual-coherence compensation in multiband fusion isar imaging,” IET Radar, Sonar & Navigation, vol. 13, no. 7, pp. 1056–1062, 2019.

[6] X.-L. Yong Qiang Zou, Xun Zhang Gao and Y. X. Liu, “A matrix pencil algorithm based multiband iterative fusion imaging method,” Scientific Reports, vol. 6, no. 1, 2016.

[7] S. Zhang, J. Yang, P. Ge, T. Feng, Y. Du, and H. Jiang, “Multiband radar signal phase reference synthesis based on matrix pencil method,” in 2023 IEEE 4th Advanced Information, Manage, Electronic and Automation Control Conference (IMCEC), vol. 4, 2021, pp. 1–7.

[8] J. Wang, P. Aubry, and A. Yarovsky, “Wavenumber-domain multiband signal fusion with matrix-pencil approach for high-resolution imaging,” IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 7, pp. 4037–4049, 2018.

[9] J. Tian, J. Sun, G. Wang, Y. Wang, and W. Tan, “Multiband radar signal coherent fusion processing with iaa and apfi,” IEEE Signal Processing Letters, vol. 20, no. 5, pp. 463–466, 2013.

[10] P. Hu, S. Xu, W. Wu, and Z. Chen, “Sparse subband isar imaging based on autoregressive model and smoothed $\theta$ algorithm,” IEEE Sensors Journal, vol. 18, no. 22, pp. 9315–9323, 2018.

[11] B. Hussein, A. Malacarne, D. Onori, S. Pinna, F. Laghezza, G. Meloni, S. Maresca, F. Scotti, P. Ghelfi, and A. Bogoni, “Performance analysis of auto-regressive uwb synthesis algorithm for coherent sparse multi-band radars,” in International Conference on Radar Systems (Radar 2017), 2017, pp. 1–6.

[12] P. van Dorp, R. Ebeling, and A. G. Huizing, “High resolution radar methods on compressed sensing,” in 2019 International Applied Computational Electromagnetics Society Symposium - China (ACES), vol. 1, 2019, pp. 1–2.

[13] I. Gorodnitsky and B. Rao, “Sparse signal reconstruction from limited data using focus: a re-weighted minimum norm algorithm,” IEEE Transactions on Signal Processing, vol. 45, no. 3, pp. 600–616, 1997.

[14] Y. Zou, X. Gao, and X. Li, “A sparse representation and gtd model parameter estimation based multiband radar signal coherent compensation method,” in 2016 CIE International Conference on Radar (RADAR), 2016, pp. 1–4.

[15] F. Zhou and X. Bai, “High-resolution sparse subband imaging based on bayesian learning with hierarchical priors,” IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 8, pp. 4568–4580, 2018.

[16] Y. Zhang, T. Wang, H. Zhao, Y. Zhang, and H. Zhao, “Multiple radar subbands fusion algorithm based on support vector regression in complex noise environment,” IEEE Transactions on Antennas and Propagation, vol. 66, no. 1, pp. 381–392, 2018.

[17] B. Zhou, Q. Chen, H. Wymeersch, P. Xiao, and L. Zhao, “Variational inference-based positioning with nondeterministic measurement accuracies and reference location errors,” IEEE Transactions on Mobile Computing, vol. 16, no. 10, pp. 2955–2967, 2017.

[18] F. Ye, F. He, and J. Zhu, “Multiband radar signal fusion based on gtd model,” in 8th European Conference on Synthetic Aperture Radar, 2010, pp. 1–4.

[19] H. H. Zhang and R. S. Chen, “Coherent processing and superresolution technique of multi-band radar data based on fast sparse bayesian learning algorithms,” IEEE Transactions on Antennas and Propagation, vol. 62, no. 12, pp. 6217–6227, 2014.

[20] R. Zhang, X. Bui, and J. Zhao, “Multiband passive isar processing based on bayesian compressive sensing,” in 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), 2019, pp. 1–4.

[21] D. Xiong, J. Wang, L. Zhao, and M. Gao, “Bssl-based multiband fusion isar imaging,” The Journal of Engineering, vol. 2019, no. 19, pp. 6039–6042, 2019.

[22] T. Kazar, G. J. M. Janssen, J. Romme, and A.-J. van der Veen, “Delay estimation for ranging and localization using multiband channel state information,” IEEE Transactions on Wireless Communications, vol. 21, no. 4, pp. 2591–2607, 2022.

[23] A. Liu, V. K. N. Lau, and B. Kananian, “Stochastic successive convex approximation for non-convex constrained stochastic optimization,” IEEE Transactions on Signal Processing, vol. 67, no. 16, pp. 4189–4203, 2019.

[24] L. Potter, D.-M. Chiang, R. Carriere, and M. Gerry, “A gtd-based parametric model for radar scattering,” IEEE Transactions on Antennas and Propagation, vol. 43, no. 10, pp. 1058–1067, 1995.

[25] B. D. Rao and K. S. Hari, “Performance analysis of root-music,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, no. 12, pp. 1939–1949, 1989.

[26] S. Y. Kung, K. S. Arun, and D. Rao, “State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem,” J. Opt. Soc. Amer. vol. 73, no. 12, pp. 1799–1811, 1983.

[27] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 33, no. 2, pp. 387–392, 1985.

[28] M. Wax and I. Ziskind, “Detection of the number of coherent signals by the mdl principle,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, no. 8, pp. 1190–1196, 1989.

[29] C. W. Fox and S. J. Roberts, “A tutorial on variational bayesian inference,” Artificial Intelligence Review, vol. 38, no. 2, pp. 85–95, 2012.

[30] M. Vemula, M. F. Bugallo, and P. M. Djuric, “Sensor self-localization with beacon position uncertainty,” Signal Processing, vol. 89, no. 6, pp. 1144–1154, 2009.

[31] G. Parisi and R. Shanks, “Statistical field theory,” 1988.

[32] M. J. Beal, “Variational algorithms for approximate bayesian inference,” Phd Thesis University of London, 2003.

[33] A. Ruszczyński, “Feasible direction methods for stochastic programming problems,” Mathematical Programming, vol. 19, no. 1, pp. 220–229, 1980.

[34] X. Wang and F. Jiang, “Multi-band synthesis of wideband radar based on compressed sensing,” in 2019 International Applied Computational Electromagnetics Society Symposium - China (ACES), vol. 1, 2019, pp. 1–2.

[35] D. Xiong, J. Wang, X. Qi, and M. Gao, “A coherent compensation method for multiband fusion imaging,” in 2017 IEEE Radar Conference (RadarConf), 2017, pp. 1024–1027.