Digital Signal Processing 126 (2022) 103460

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Sparse Bayesian multiple sources localization using variational approximation for Laplace priors ☆

Roubing Tang [a], Qiaoling Zhang [a,*], Weiwei Zhang [b], Han Ma [a]

[a] School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou, China
[b] Information Science and Technology College, Dalian Maritime University, Dalian, China

ARTICLE INFO

Article history:
Available online 11 February 2022

Keywords:
Sparse Bayesian learning
Multiple sources localization
Laplace priors
Parametric dictionary learning
Variational approximation

ABSTRACT

Multiple sources localization (MSL) has been playing an important role in various applications of wireless sensor networks. In this paper, sparse representation-based MSL with received signal strength (RSS) as observations is addressed within the sparse Bayesian learning framework. Conventional approaches are usually based on the assumption of Gaussian prior for the sparse coefficients. In this paper, Laplacian prior is employed to model the sparse coefficient for improved sparsity performance. To tackle the problem that Laplace prior is not conjugate to Gaussian likelihood, a variational approximation is employed for Bayesian inference. Finally, a multiple sources localization method is developed. Furthermore, to reduce the computing complexity and promote sparsity, a fast multiple sources localization method is further developed. In the proposed methods, the source locations, sparse coefficients, and all necessary model parameters are adaptively deduced solely from the RSS observations. Meanwhile, by employing Laplacian prior for the sparse coefficients, better sparse recovery can be achieved. Experimental results demonstrate the effectiveness of the proposed methods.

1. Introduction

Multiple sources localization in wireless sensor networks (WSNs) has been playing an important role in many applications [1][2], ranging from commercial, industrial to defense areas. Usually, WSNs comprise many cheap, battery-powered sensors to measure natural conditions like temperature [3], radio signal strength [4], contamination levels [5][6], and moistness [7], etc., from the target sources in the environment. Theoretically, the objective of multiple sources localization (MSL) approaches is to estimate the locations of multiple sources emitting signals such as electromagnetic [8], acoustic [9] and seismic, based on the observations captured by the WSNs.

Over the decades, many literatures have been developed on multiple sources localization in WSNs [10]. Typically, observations employed by localization methods in WSNs mainly include: direction-of-arrival (DOA) [11][12][13], time-of-arrival (TOA) [14][15], angle-of-arrival (AOA) [16] and received-signal-strength (RSS) [17][18]. Among these observations, RSS is more attractive because of its simplicity and low energy cost. Many researchers have investigated this topic from different perspectives in the last few decades, resulting in a large number of RSS-based localization estimation methods.

Conventionally, maximum likelihood estimate (MLE) [19] provides a solution for localization [20]. However, the MLE approaches usually involve exhaustive search, whose computation cost is extremely high, especially for the case of multiple sources [19]. To mitigate the computation complexity of MLE in MSL, the authors in [21] considered a multiresolution (MR) search algorithm and an expectation-maximization (EM) like iterative algorithm to start a coordinate search. Meanwhile, in [22], an alternating projection multiresolution (APMR) algorithm was proposed to reduce the computation complexity. To further reduce the computational complexity and improve the localization performance, an efficient EM algorithm for multiple sources localization is proposed in [23]. In [24], by employing an efficient alternating projection (AP) procedure, a novel EM algorithm was proposed for MSL as well. The method can avoid the multidimensional search by sequentially estimating the location of one single source while fixing the estimates of others obtained from the previous iteration. Additionally, distributed particle filters with Gaussian mixture approximation were proposed to localize and track multiple moving targets in

☆ This document is the results of the research project funded by the National Natural Science Foundation of China (61806178), Zhejiang Provincial Natural Science Foundation of China (LY21F010015), Postdoctoral Science Foundation of China (2020M680932) and the Key Research and Development Program of Zhejiang Province (2021C01047).

* Corresponding author.
E-mail address: qlzhang@zstu.edu.cn (Q. Zhang).

https://doi.org/10.1016/j.dsp.2022.103460
1051-2004/

[25]. The authors in [26][27] reformulated the MSL problem as a unimodality-constrained matrix factorization (UMF) problem and developed a projected gradient algorithm to extract the signature vectors for localizing each of the sources.

Recently, sparse representation-based localization has attracted much attention, which formulates the sources localization as a joint sparse signal recovery and parametric learning problem, and solves it within the Bayesian learning framework [28]. In [29], a variational Bayesian EM algorithm for multiple sources localization was proposed, where the sparse signal and parameter were iteratively updated in the variational Bayesian expectation step and maximization step, respectively. In [30], the localization problem was reformulated as a sparse recovery problem that recovered two sparse support-related vectors, and a three-level hierarchical prior model was imposed on the sparse signals to learn their supports. In [31], the localization problem was transformed to reconstruct a block-sparse signal and learn its inner correlation structure. Then, a novel block-sparse reconstruction algorithm was proposed for jointly estimating the block-sparse signal and learning its inner correlation structure so as to estimate locations and varying powers of multiple targets simultaneously. To reduce the computational cost and the effect of the off-grid problem, [32] provided a fast iterative update for the hyperparameters of the sparse probabilistic model and used the Taylor expansion to approximate the true parametric dictionary respectively. In all, these methods use Gaussian distribution as their priors for sparse coefficients in the sparse signal recovery subproblem of MSL. Note that the Laplace prior has been proved better for sparse recovery [33]. And the Laplace prior has been employed to model the sparse coefficients for the DOA estimations of multiple sources [34][35].

In this paper, the RSS-based MSL problem is addressed within the sparse Bayesian learning framework. Specifically, the MSL problem is reformulated as a joint sparse signal recovery and parametric dictionary learning problem; in order to achieve better sparse signal recovery, which is furthermore helpful for the final localization, the Laplace prior is employed to model the sparse coefficients. To address the problem that the Laplace prior is not conjugate with the Gaussian distribution and Bayesian inference is intractable, variational approximations are made so as to carry on Bayesian inference further; subsequently, a sparse Bayesian multiple sources localization (SBMSL) algorithm is developed. Moreover, in order to reduce the computational complexity of SBMSL and promote sparsity, a fast algorithm referred to as fast sparse Bayesian multiple sources localization (FastSBMSL) is further developed. In the proposed methods, the source locations, sparse coefficients, and all necessary model parameters are adaptively deduced solely from the RSS observations. Meanwhile, by employing Laplacian prior for the sparse coefficients, better sparse recovery can be obtained compared to Gaussian prior.

The rest of this paper is organized as follows. In section 2, the signal superposition model is first presented. Then the sparse-based multiple sources localization problem is formulated. In section 3, statistical models are first presented for observations, sparse coefficients, and parametric dictionary. Then, based on these models, a SBMSL algorithm is developed within the sparse Bayesian learning framework. Particularly, to tackle the problem that the Laplace prior is not conjugate to Gaussian likelihood, a lower bound on the Laplacian prior with the form of Gaussian is developed via the local variational approximation. Subsequently, based on the variational approximation for the Laplace prior, the SBMSL algorithm is developed. To reduce the computational complexity of the SBMSL, a fast algorithm is further developed. In section 4, simulations are performed to evaluate the proposed methods, and the related results are analyzed as well. In section 5, some conclusions are given.

Table 1
Notations.

Notation	Description
$\mathbb{L}$	sources locations
K	number of sources
$\mathbb{S}$	sensors locations
M	number of sensors
$\mathbf{y}$	the noisy measurements
P_k	the transmitted power of the k-th source target at a reference distance d_0
η	path loss coefficient
$\boldsymbol{\Phi}$	the dictionary matrix
$\mathcal{G}$	candidate grid points
N	number of candidate grid points
$\boldsymbol{\omega}$	sparse representation coefficients
$\boldsymbol{\lambda}$	hyper-parameter of $\boldsymbol{\omega}$
c, d	hyper-parameters of $\boldsymbol{\lambda}$
β^{-1}	the noise variance
a, b	hyper-parameters of β
ς	the variational parameters
$\boldsymbol{\varepsilon}$	contact points of variational lower bounds

Notations used in this paper: $\|\cdot\|$ is the Euclidean norm, $\mathrm{tr}(\cdot)$ represents trace; $\langle\cdot\rangle$ represents expectation; $\circ$ represents Hadamard product. Given a vector $\mathbf{h}$, $\|\mathbf{h}\|_1, \|\mathbf{h}\|_2, \|\mathbf{h}\|_F$ represents the l_1 norm, l_2 norm and Frobenius norm of the vector respectively; $\mathbf{h}_i, \mathbf{h}_{-i}$ represents the i-th element of the vector $\mathbf{h}$ and the vector $\mathbf{h}$ without the i-th component respectively. Given a matrix $\mathbf{A}, \mathbf{A}^{-1}, \mathbf{A}^T, \mathbf{A}'_a$ represents inverses, transpose and the partial differential of a respectively; $\mathbf{A}_i, \mathbf{A}^j, \mathbf{A}_{i,j}$ represents the i-th row of the matrix, j-th column of the matrix and (i, j)-th element of the matrix respectively. And the notations used through the paper are listed in Table 1.

2. Problem formulation

In this section, we state the problem of multiple sources localization (MSL) as a joint dictionary learning (DL) and sparse signal reconstruction (SSR) problem. Given current dictionary parameters, we can estimate the sparse coefficients $\boldsymbol{\omega}$ using Bayesian inference methods [36]. Then, according to $\boldsymbol{\omega}$, we can update the dictionary parameters.

2.1. Signal superposition model

Consider multiple sources localized in a two-dimensional square area covered by a wireless sensor network. Each target source is equipped with radio transmission equipment that periodically emits signals perceived by some sensors deployed in the area. And the signals from all targets are superimposed at the sensor. To capture this fact, we model each received signal strength (RSS) as the superimposition of the received signal strengths from all targets, which is called the signal superposition model.

Assume that in a two-dimensional region of interest (ROI) $\mho$, there are M sensors with known locations $\mathbb{S} = \{\boldsymbol{s}_i = (u_i^s, v_i^s)\}_{i=1}^M$ and K sources with unknown locations $\mathbb{L} = \{\boldsymbol{l}_k = (u_k^l, v_k^l)\}_{k=1}^K$. The RSS of sensor i is specified by

$$y_i = g_i \sum_{k=1}^{K} P_k f(s_i, l_k, \eta) + \varepsilon_i \tag{1}$$

where g_i is the gain factor of sensor i, which will be set to 1 hereafter for simplicity since it can be calibrated in advance [32]. P_k is the transmitted power of the k-th source at a reference distance d_0, ε_i is the observation noise of the i-th sensor, and $f(\cdot)$ is an energy attenuation function, defined as [37]

$$f(s_i, l_k, \eta) = \begin{cases} 1, & \text{if } d < d_0 \\ \left(\frac{d_0}{\|\boldsymbol{l}_k - \boldsymbol{s}_i\|}\right)^{\eta}, & \text{otherwise} \end{cases} \tag{2}$$

where η is the path loss exponent (PLE), and $\|\cdot\|$ denotes the Euclidean norm.

Denote $\mathbf{y}$ as the stacked vector from all sensors' observations, i.e., $\mathbf{y} = [y_1, y_2, ..., y_M]^T$. The matrix-vector representation of the observation model of all sensors is

$$\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{x} + \mathbf{e} \tag{3}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{M\times K}$, $\boldsymbol{\Psi}_{i,k} = f(s_i, l_k, \eta)$, $\boldsymbol{x} = [P_1, P_2, \ldots, P_K]^T$, and $\mathbf{e} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_K]^T$. The objective of MSL is to estimate the locations of multiple sources from noisy observations $\mathbf{y}$.

2.2. Sparsity-based multiple sources localization

Assume that the region covering sensors and sources is discretized into a set of N candidate grid points $\mathcal{G} = \{\boldsymbol{g}_j = (u_j, v_j), j = 1, \cdots, N\}$, which contains source locations. Generally, the number of sources K is far less than that of the candidate grid points N, i.e., $K \ll N$. Then, the MSL model of (3) can be reformulated as the following sparse representation model [38]

$$\mathbf{y} = \boldsymbol{\Phi}(\mathcal{G}, \eta)\boldsymbol{\omega} + \mathbf{e} \tag{4}$$

where $\boldsymbol{\Phi}(\mathcal{G}, \eta) \in \mathbb{R}^{M\times N}$ is the so-called sparse dictionary matrix parametrized by $\mathcal{G}$ and η, $\boldsymbol{\Phi}_{i,j}(\mathcal{G}, \eta) = f(s_i, g_j, \eta)$, $\boldsymbol{\omega} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_N]^T$ are the sparse representation coefficients, $\boldsymbol{\omega}_i = P_k$ when source k is located at $\boldsymbol{g}_i$, and otherwise $\boldsymbol{\omega}_i = 0$. In this way, the MSL problem is transformed into a sparse signal recovery (SSR) problem [29], and sparse-based localization approaches can be applied, where source locations can be estimated by searching the sparsest solution.

Traditional sparse-based localization approaches implicitly assumed that source localizations fall exactly on predefined grid points $\mathcal{G}$. In practice, source localizations could probably derive from the grid points $\mathcal{G}$, referred to as the off-grid problem, which deteriorates the localization performance greatly. Moreover, the parameter η may differ from that in the RSS observations [32], or even unknown [20] and time-varying. Usually, both the grid points $\mathcal{G}$ and the PLE η can be rationally seen as unknown variables to be estimated based on the RSS observations. Therefore, the MSL problem can be cast into a joint optimization problem comprising two subproblems [32]: (a) how to estimate the dictionary parameters $\mathcal{G}$ and η for fixed coefficients $\boldsymbol{\omega}$, so-called the parametric dictionary learning (PDL); (b) how to estimate the sparse representation coefficients $\boldsymbol{\omega}$ for fixed dictionary parameters $\mathcal{G}$ and η, namely, the SSR problem.

The PDL subproblem has no mature solutions, and an alternative approximation can be obtained in an iterative manner [32]. Denote $\mathcal{G}^{(t)}$ and $\eta^{(t)}$ as the estimated dictionary parameters at iteration t, $\mathcal{G}^\star$ and $\eta^\star$ as the optimal points. A feasible solution for the off-grid problem can be obtained via the first-order Taylor expansion-based model

$$\begin{aligned} &\boldsymbol{\Phi}(\mathcal{G}^\star, \eta^\star) \approx \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}'_u\left(\mathcal{G}^{(t)}, \eta^{(t)}\right)\mathrm{diag}(\boldsymbol{\delta_u}) \\ &+\boldsymbol{\Phi}'_v\left(\mathcal{G}^{(t)}, \eta^{(t)}\right)\mathrm{diag}(\boldsymbol{\delta}_v) + \delta_\eta\boldsymbol{\Phi}'_\eta\left(\mathcal{G}^{(t)}, \eta^{(t)}\right) \end{aligned} \tag{5}$$

where $\boldsymbol{\delta}_g = [\boldsymbol{\delta}_u, \boldsymbol{\delta}_v]$ is the offset between $\mathcal{G}^{(t)}$ and $\mathcal{G}^\star$, $\boldsymbol{\Phi}_0 = \boldsymbol{\Phi}\left(\mathcal{G}^{(t)}, \eta^{(t)}\right)$, and $\boldsymbol{\Phi}'_u$, $\boldsymbol{\Phi}'_v$ and $\boldsymbol{\Phi}'_\eta$ are defined as the partial differential of u, v and η. Then the PDL subproblem can be resolved via the following iterations until it converges

$$\mathcal{G}^{(t+1)} = \mathcal{G}^{(t)} + \boldsymbol{\delta}_g, \quad \eta^{(t+1)} = \eta^{(t)} + \delta_\eta \tag{6}$$

For the SSR subproblem, Gaussian prior are usually employed to model the sparse coefficients $\boldsymbol{\omega}$ and then conventional Bayesian inference can performed [29]. Yet, it has been shown that Laplacian prior on the coefficients $\boldsymbol{\omega}$ is better for sparse recovery [33], compared with Gaussian prior. Therefore, a Laplacian prior is adopted to model the sparse coefficients $\boldsymbol{\omega}$ herein, and the related inference is further developed (for detail in Section 3.3.1).

Overall, the MSL problem can be reformulated as the following joint optimization problem [39]

$$\left(\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\delta}}_g, \hat{\delta}_\eta\right) = \underset{\boldsymbol{\omega}, \boldsymbol{\delta}_g, \delta_\eta}{\arg\min}\, \mathcal{R}(\boldsymbol{\omega}) \tag{7}$$

$$\text{s.t. } \boldsymbol{\Phi}_0 = \boldsymbol{\Phi}\left(\mathcal{G}^{(t)}, \eta^{(t)}\right), \tag{8a}$$

$$\begin{aligned} \boldsymbol{\Phi} = \boldsymbol{\Phi}_0 &+ \boldsymbol{\Phi}'_u\left(\mathcal{G}^{(t)}, \eta^{(t)}\right)\mathrm{diag}(\boldsymbol{\delta}_u) + \delta_\eta\boldsymbol{\Phi}'_\eta\left(\mathcal{G}^{(t)}, \eta^{(t)}\right) \\ &+ \boldsymbol{\Phi}'_v\left(\mathcal{G}^{(t)}, \eta^{(t)}\right)\mathrm{diag}(\boldsymbol{\delta}_v), \end{aligned} \tag{8b}$$

$$\|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\omega}\|_F < \zeta, \tag{8c}$$

$$\boldsymbol{\delta_u} \in [\boldsymbol{LB}_u, \boldsymbol{UB}_u], \boldsymbol{\delta_v} \in [\boldsymbol{LB}_v, \boldsymbol{LB}_v], \delta_\eta \in \left[LB_\eta, UB_\eta\right] \tag{8d}$$

where $\mathcal{R}(\boldsymbol{\omega})$ denotes the row sparsity of $\boldsymbol{\omega}$, i.e., the number of non-zero rows. In this way, our goal is to iteratively estimate the sparse coefficients $\boldsymbol{\omega}$, the dictionary parameters $\mathcal{G}$ and η from the noisy measurements $\mathbf{y}$.

3. Multiple sources localization via sparse Bayesian learning

In this section, statistical models for the MSL problem are first presented: the Gaussian model for the observation likelihood, the Laplacian prior for the sparse coefficients, and the dictionary model for the PLE parameter and source locations. Considering that the Laplacian prior is not conjugate to the Gaussian likelihood and is intractable for Bayesian inference, a variational approximation is made for the Laplace prior. Based on the variational approximation, Bayesian references for both sparse coefficients and dictionary parameters are then presented, yielding the SBMSL algorithm. Furthermore, a fast solution for SBMSL, named as FastSBMSL is developed.

3.1. Bayesian model

3.1.1. Observation (noise) model

Assume that the observation noises of all individual sensors are i.i.d. Gaussian random variables. The observation model can be expressed as

$$p(\mathbf{y}|\boldsymbol{\omega}, \beta) = N(\mathbf{y}|\boldsymbol{\Phi}\boldsymbol{\omega}, \beta^{-1}\mathbf{I}) \tag{9}$$

where β^{-1} is the noise variance, and a conjugate Gamma prior with hyper-parameters a and b is imposed on β

$$p(\beta; a, b) = \mathrm{Gamma}(\beta|a, b) \tag{10}$$

where $a > 0$ is the shape parameter, and $b > 0$ is the scale parameter.

3.1.2. Sparse model

As has been shown [33][35] that, compared to Gaussian priors [30], Laplace priors enforce the sparsity constraint more heavily on the coefficients. Meanwhile, Laplace prior is the prior that promotes sparsity to the largest extent while being log-concave [40], which is helpful to eliminate local-minima since it leads to unimodal posterior distributions [33][40][41]. Therefore, a Laplace prior is employed herein to model the coefficients $\boldsymbol{\omega}$:

$$p(\boldsymbol{\omega}|\boldsymbol{\lambda}) = \prod_{i=1}^{N} p(\omega_i|\lambda_i) = \prod_{i=1}^{N} \frac{\lambda_i}{2} \exp(-\lambda_i \|\omega_i\|_1) \tag{11}$$

and Gamma prior is placed on $\boldsymbol{\lambda}$, $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_N]^T$ the function as follows.

$$p(\boldsymbol{\lambda}|c, d) = \prod_{i=1}^{N} \text{Gamma}(\lambda_i; c, d) \tag{12}$$

3.1.3. Dictionary model

From 2.2, we know that the grid points $\mathcal{G}$ is limited in the ROI $\mho$, so we model $\mathcal{G}$ as random variable with uniform prior in the region $\{\boldsymbol{g}_j \in \mho, j = 1, \cdots, N\}$. According to (5), to handle the offset problem, we should update the offset $\boldsymbol{\delta_g} = [\boldsymbol{\delta}_u, \boldsymbol{\delta}_v]$ iterative to obtain optimal grid points and get the source locations from the optimal grid points. So the offsets also are modeled as uniform prior as follows:

$$\boldsymbol{\delta}_u \sim U\left(\left[-\frac{\bar{d}}{2}, \frac{\bar{d}}{2}\right]^N\right), \boldsymbol{\delta}_v \sim U\left(\left[-\frac{\bar{d}}{2}, \frac{\bar{d}}{2}\right]^N\right) \tag{13}$$

where $\bar{d}$ is the neighborhood size between grid points.

And PLE η is limited in $[\underline{\eta}, \bar{\eta}]$, so we assume that the PLE is modeled as uniform prior.

$$\eta \sim U(\underline{\eta}, \bar{\eta}) \tag{14}$$

where $\underline{\eta}, \bar{\eta}$ represents as the lower and upper boundary of η respectively.

3.2. Variational approximation for Laplace prior

It is worth noting that the Bayesian inference is intractable, since the Laplace prior $\text{p}(\boldsymbol{\omega}|\boldsymbol{\lambda})$ is not conjugate to Gaussian likelihood $p(\mathbf{y}|\boldsymbol{\omega}, \beta)$ [42]. To tackle this problem, the variational method is employed to seek a Gaussian approximation for the Laplace prior so that the Bayesian analysis can be carried on.

The logarithm function of the Laplace prior in (11) can be expressed as

$$\ln p(\omega_i|\lambda_i) = \ln \lambda_i - \lambda_i \|\omega_i\|_1 - \ln 2, i \in \{1, \ldots, N\}. \tag{15}$$

Note that this is a convex function about $\|\omega_i\|_1^2$, which can be easily verified by taking the second derivative. Thus, a lower bound on $\ln p(\omega_i|\lambda_i)$ can be found through its conjugate function [35]

$$g(\varsigma) = \max_{\|\omega_i\|_1^2} \left\{\varsigma \|\omega_i\|_1^2 - \ln p(\omega_i|\lambda_i)\right\}. \tag{16}$$

According to the stationarity condition

$$0 = \varsigma - \frac{d(\|\omega_i\|_1)}{d\left(\|\omega_i\|_1^2\right)} \frac{d[\ln p(\omega_i|\lambda_i)]}{d(\|\omega_i\|_1)} = \varsigma + \frac{\lambda_i}{2\|\omega_i\|_1}, \tag{17}$$

we have

$$\varsigma(\varepsilon_i) = -\frac{\lambda_i}{2\|\varepsilon_i\|_1} \tag{18}$$

where, for the sake of analysis, the value of ω_i corresponding to the contact condition (17) is denoted by ε_i thereafter. Then, the conjugate function $g(\varsigma)$ can be rewritten as

$$g(\varsigma) = \varsigma(\varepsilon_i)\|\varepsilon_i\|_1^2 - \ln \lambda_i + \lambda_i \|\varepsilon_i\|_1 + \ln 2 \tag{19}$$

Based on (16) and (18), the bound on the logarithm function $\ln p(\omega_i|\lambda_i)$ can be obtained as

$$\begin{aligned} \ln p(\omega_i|\lambda_i) &\geq \varsigma(\varepsilon_i)\|\omega_i\|_1^2 - \varsigma(\varepsilon_i)\|\varepsilon_i\|_1^2 \\ &+ \ln \lambda_i - \lambda_i \|\varepsilon_i\|_1 - \ln 2 \triangleq h(\omega_i, \varepsilon_i) \end{aligned} \tag{20}$$

Clearly, the bound $h(\omega_i, \varepsilon_i)$ on $p(\omega_i|\lambda_i)$ has the form of the exponential of a quadratic function of $\|\omega_i\|_1$, making it feasible for seeking Gaussian approximations of the Laplace prior $p(\omega_i|\lambda_i)$. Based on $h(\omega_i, \varepsilon_i)$, a lower bound $h(\boldsymbol{\omega}, \boldsymbol{\varepsilon})$ on Laplace prior $p(\boldsymbol{\omega}|\boldsymbol{\lambda})$ is obtained as

$$\begin{aligned} h(\boldsymbol{\omega}, \boldsymbol{\varepsilon}) = \prod_{i=1}^{N} \exp[&\varsigma(\varepsilon_i)\|\omega_i\|_1^2 - \varsigma(\varepsilon_i)\|\varepsilon_i\|_1^2 \\ &+ \ln \lambda_i - \lambda_i \|\varepsilon_i\|_1 - \ln 2] \end{aligned} \tag{21}$$

and $h(\boldsymbol{\omega}, \boldsymbol{\varepsilon})$ has also the Gaussian form as an exponential of a quadratic function $\|\omega_i\|_1^2$. Based on such a Gaussian lower bound on the Laplace prior $p(\boldsymbol{\omega}|\boldsymbol{\lambda})$, we can carry on Bayesian inference procedure for parameter estimation.

Note that the parameter ε_i in (20)-(21) can be optimized by maximizing the bound on the logarithmic marginal likelihood $\ln p(\mathbf{y})$, i.e.,

$$\begin{aligned} \ln p(\mathbf{y}) &= \ln \int p(\mathbf{y}|\boldsymbol{\omega}) p(\boldsymbol{\omega}) d\boldsymbol{\omega} \\ &\geq \ln \int p(\mathbf{y}|\boldsymbol{\omega}) h(\boldsymbol{\omega}, \boldsymbol{\varepsilon}) d\boldsymbol{\omega} \triangleq \mathcal{L}(\boldsymbol{\varepsilon}) \end{aligned} \tag{22}$$

Take derivative of $\mathcal{L}(\boldsymbol{\varepsilon})$ with respect to $\|\varepsilon_i\|_1$ and set the corresponding result to zero. Then, we have

$$\begin{aligned} \|\varepsilon_i\|_1^2 =& \frac{1}{2}\left\{\left[\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} - 2\,\text{diag}(\boldsymbol{\varsigma})\right]^{-1}\right\}_{ii} \\ &+ \frac{1}{2}\left\{\left[\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} - 2\,\text{diag}(\boldsymbol{\varsigma})\right]^{-1} \beta \boldsymbol{\Phi}^T \mathbf{y} \beta \mathbf{y}^T \boldsymbol{\Phi}\right. \\ &\left. \times \left[\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} - 2\,\text{diag}(\boldsymbol{\varsigma})\right]^{-1}\right\}_{ii} \end{aligned} \tag{23}$$

where the related deduction employed the property that $\|\varepsilon_i\|_1 = |\varepsilon_i|$, $\|\omega_i\|_1 = |\omega_i|$ for the scalar RSS observations.

3.3. Sparse Bayesian multiple source localization

3.3.1. Bayesian inference

The type-II maximum likelihood procedure is utilized herein to estimate the coefficients $\boldsymbol{\omega}$, the hyper-parameters $\boldsymbol{\lambda}$, β, and the dictionary parameters η, $\mathcal{G}$.

$$\begin{aligned} (\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta, \eta, \mathcal{G}) &= \arg\max_{\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta, \eta, \mathcal{G}} \text{p}(\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta|\mathbf{y}; \eta, \mathcal{G}) \\ &= \arg\max_{\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta, \eta, \mathcal{G}} \text{p}(\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \beta; \eta, \mathcal{G}) \\ &= \arg\max_{\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta, \eta, \mathcal{G}} \ln \text{p}(\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \beta; \eta, \mathcal{G}) \end{aligned} \tag{24}$$

Maximizing the logarithm of the marginal likelihood $\ln \text{p}(\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \beta; \eta, \mathcal{G})$ w.r.t. $\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta, \eta, \mathcal{G}$ for the stationary point is intractable due to the Laplace prior $\text{p}(\boldsymbol{\omega}|\boldsymbol{\lambda})$, and the estimation cannot be computed in closed form. Therefore, approximation has to be made throughout iterative inference methods. In this section, variational approximations have been made to carry on the Bayesian inference of the proposed probabilistic model.

(1) Update of Sparse Coefficients and Model Parameters

Regard $\{\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta\}$ as the latent variables, and introduce a variational distribution $q(\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta)$. The related variational inference starts with the decomposition of the logarithmic marginal likelihood $\ln \mathbf{y}$:

$$\ln \mathbf{y} = \mathcal{L}(q) + \mathcal{KL}(q||p) \tag{25}$$

where $\mathcal{KL}(q||p)$ is the Kullback-Leibler divergence, and $\mathcal{L}(q)$ is the lower bound of $\ln \mathbf{y}$, given by

$$\mathcal{L}(q) = \iiint q(\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta) \ln \left\{ \frac{p(\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \beta; \eta, \mathcal{G})}{q(\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta)} \right\} d\boldsymbol{\omega} d\boldsymbol{\lambda} d\beta \tag{26}$$

Note that the lower bound $\mathcal{L}(q)$ is intractable due to the Laplace prior $p(\boldsymbol{\omega}|\boldsymbol{\lambda})$. As to this issue, it is feasible to place a bound on $\mathcal{L}(q)$ by employing the variational approximation (21) for the prior $p(\boldsymbol{\omega}|\boldsymbol{\lambda})$, leading to

$$\begin{aligned} \ln p(\mathbf{y}) &\geq \mathcal{L}(q) \geq \mathcal{L}(q, \boldsymbol{\varepsilon}) \\ &= \iiint q(\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta) \ln \left\{ \frac{\tilde{p}(\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \beta; \eta, \mathcal{G})}{q(\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta)} \right\} \\ &\quad \times d\boldsymbol{\omega} d\boldsymbol{\lambda} d\beta \end{aligned} \tag{27}$$

where

$$\tilde{p}(\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \beta; \eta, \mathcal{G}) = p(\mathbf{y}|\boldsymbol{\omega}, \beta; \eta, \mathcal{G}) h(\boldsymbol{\omega}, \boldsymbol{\varepsilon}) p(\boldsymbol{\lambda}) p(\beta) \tag{28}$$

Thus, maximizing the logarithmic marginal likelihood $\ln \mathbf{y}$ w.r.t. $\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta$ can be implemented by maximizing the lower bound $\mathcal{L}(q, \boldsymbol{\varepsilon})$ w.r.t. these parameters, respectively.

Assume that the variational distribution $q(\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta)$ can be factorized, via the Mean-field variational inference [43], as follows

$$q(\boldsymbol{\omega}, \boldsymbol{\lambda}, \beta) = q(\boldsymbol{\omega}) q(\boldsymbol{\lambda}) q(\beta) \tag{29}$$

Based on the factorization (29) and the condition for establishment of the lower bound in (27), the approximate posterior distribution $q^*(\boldsymbol{\omega})$ of $\boldsymbol{\omega}$ can be computed according to

$$\begin{aligned} \ln q^*(\boldsymbol{\omega}) =& E_{\beta, \boldsymbol{\lambda}}[\ln p(\mathbf{y}|\boldsymbol{\omega}, \beta; \eta, \mathcal{G}) h(\boldsymbol{\omega}, \boldsymbol{\varepsilon})] + \text{const} \\ =& -\frac{1}{2} \beta \|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\omega}\|_2^2 \\ &+ \boldsymbol{\omega}^T \operatorname{diag}(\boldsymbol{\varsigma}) \boldsymbol{\omega} + \text{const} \end{aligned} \tag{30}$$

where $E_{\beta, \boldsymbol{\lambda}}\{\cdot\}$ denotes the expectation w.r.t. the posterior of β and $\boldsymbol{\lambda}$. Obviously, from the above function, we can know the approximate posterior $q^*(\boldsymbol{\omega})$ of $\boldsymbol{\omega}$ is Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ respectively given by:

$$\boldsymbol{\mu} = \boldsymbol{\Sigma} \beta \boldsymbol{\Phi}^T \mathbf{y} \tag{31}$$

$$\boldsymbol{\Sigma} = \left[\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} + diag(-2\boldsymbol{\varsigma}) \right]^{-1} \tag{32}$$

Note that these updating formulas for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ involve $\mathcal{O}(N^3)$ computations, which is rather computations demanding. Actually, the covariance matrix $\boldsymbol{\Sigma}$ can be rewritten, using the Woodbury matrix identity, as follows

$$\begin{aligned} \Sigma &= \Lambda^{-1} - \Lambda^{-1} \boldsymbol{\Phi}^T (\beta^{-1} I + \boldsymbol{\Phi} \Lambda^{-1} \boldsymbol{\Phi}^T)^{-1} \boldsymbol{\Phi} \Lambda^{-1} \\ &= \Lambda^{-1} - \Lambda^{-1} \boldsymbol{\Phi}^T \mathbf{C}^{-1} \boldsymbol{\Phi} \Lambda^{-1} \end{aligned} \tag{33}$$

where $\boldsymbol{\Lambda} = diag(-2\boldsymbol{\varsigma})$, and $\mathbf{C} = \beta^{-1}\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}^T$. Thus, the computations for $\boldsymbol{\Sigma}$ are reduced from $\mathcal{O}(N^3)$ to $\mathcal{O}(M^3)$, where $M \ll N$.

Similarly, the approximate posteriors $q^*(\boldsymbol{\lambda})$ and $q^*(\beta)$ for $\boldsymbol{\lambda}$ and β can be computed respectively according to

$$\ln q^*(\boldsymbol{\lambda}) = E_{\boldsymbol{\omega}}\{\ln h(\boldsymbol{\omega}, \boldsymbol{\varepsilon}) p(\boldsymbol{\lambda})\} + \text{const} \tag{34}$$

$$\ln q^*(\beta) = E_{\boldsymbol{\omega}}\{\ln p(\mathbf{y}|\boldsymbol{\omega}, \beta; \eta, \mathcal{G}) p(\beta|c, d)\} + \text{const} \tag{35}$$

where $E_{\boldsymbol{\omega}}\{\cdot\}$ denotes the expectation w.r.t. the posterior of $\boldsymbol{\omega}$.

Note that, to maximize the lower bound $\mathcal{L}(q, \boldsymbol{\varepsilon})$ w.r.t. the parameters $\boldsymbol{\lambda}$ and β are tantamount to maximize $E_{\boldsymbol{\omega}}\{\ln h(\boldsymbol{\omega}, \boldsymbol{\varepsilon}) p(\boldsymbol{\lambda})\}$ and $E_{\boldsymbol{\omega}}\{\ln p(\mathbf{y}|\boldsymbol{\omega}, \beta; \eta, \mathcal{G}) p(\beta|c, d)\}$ respectively, which leads to the estimation equations of $\boldsymbol{\lambda}$ and β as follows.

$$\lambda_i = \frac{2c + 2}{d + \|\varepsilon_i\|_1 + \frac{\langle \omega_i^2 \rangle}{\|\varepsilon_i\|_1}}, i = 1, 2, \ldots, N \tag{36}$$

$$\beta = \frac{M + 2a}{2b + \|\mathbf{y} - \boldsymbol{\Phi}\mu\|_2^2 + \operatorname{tr}(\Sigma \boldsymbol{\Phi}^T \boldsymbol{\Phi})} \tag{37}$$

where $\langle w_i^2 \rangle = \mu_i^2 + \Sigma_{ii}$.

(2) Update of Dictionary Parameters

Now consider the estimation of the dictionary parameters η and $\mathcal{G}$. From (28), we are able to estimate the dictionary parameters by maximizing $E\{\ln p(\mathbf{y}|\boldsymbol{\omega}, \beta; \eta, \mathcal{G})\}$, which is equal to minimize the following objective function [29]

$$\begin{aligned} &E\left\{ (\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\omega}))^T (\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\omega}) \right\} \\ &= (\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\mu})^T (\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\mu}) + \operatorname{tr}(\Sigma \boldsymbol{\Phi}^T \boldsymbol{\Phi}) \end{aligned} \tag{38}$$

By replacing $\boldsymbol{\Phi}$ in formula (38) with formula (8b), minimizing (38) w.r.t. $\eta, \mathcal{G}$ can be further converted to the following optimization problem with constraints, i.e.,

$$\arg\min \left\{ \begin{array}{l} \boldsymbol{\delta}_u^T \boldsymbol{M}_{uu} \boldsymbol{\delta}_u + \boldsymbol{\delta}_v^T \boldsymbol{M}_{vv} \boldsymbol{\delta}_v + \epsilon \delta_\eta^2 \\ +2\boldsymbol{\delta}_u^T \boldsymbol{M}_{uv} \boldsymbol{\delta}_v + 2\delta_\eta \boldsymbol{\nu}_{u\eta}^T \boldsymbol{\delta}_u + 2\delta_\eta \boldsymbol{\nu}_{v\eta}^T \boldsymbol{\delta}_v \\ +2\boldsymbol{\nu}_u^T \boldsymbol{\delta}_u + 2\boldsymbol{\nu}_v^T \boldsymbol{\delta}_v + 2\varrho \delta_\eta \end{array} \right\} \tag{39a}$$

$$\text{s.t.} \boldsymbol{\delta}_u \in [\boldsymbol{LB}_u, \boldsymbol{UB}_u], \boldsymbol{\delta}_{\boldsymbol{\nu}} \in [\boldsymbol{LB}_v, \boldsymbol{UB}_v], \delta_\eta \in \left[LB_\eta, UB_\eta\right] \tag{39b}$$

with

$$\boldsymbol{M}_{uu} = \boldsymbol{\Phi}_u'^T \boldsymbol{\Phi}_u' \circ \left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T \right) \tag{40a}$$

$$\boldsymbol{M}_{vv} = \boldsymbol{\Phi}_v'^T \boldsymbol{\Phi}_v' \circ \left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T \right) \tag{40b}$$

$$\boldsymbol{M}_{uv} = \boldsymbol{\Phi}_u'^T \boldsymbol{\Phi}_v' \circ \left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T \right) \tag{40c}$$

$$\boldsymbol{\nu}_{u\eta} = \left[\boldsymbol{\Phi}_u'^T \boldsymbol{\Phi}_\eta' \circ \left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T \right) \right] \cdot \mathbf{1}_N \tag{40d}$$

$$\boldsymbol{\nu}_v = \left[\boldsymbol{\Phi}_v'^T \boldsymbol{\Phi}_\eta' \circ \left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T \right) \right] \cdot \mathbf{1}_N \tag{40e}$$

$$\boldsymbol{\nu}_u = \operatorname{diag}\left(\boldsymbol{\Phi}_u'^T \boldsymbol{\Phi}_0 \boldsymbol{\Sigma} \right) - \operatorname{diag}(\boldsymbol{\mu}) \boldsymbol{\Phi}_u'^T (\mathbf{y} - \boldsymbol{\Phi}_0 \boldsymbol{\mu}) \tag{40f}$$

$$\boldsymbol{\nu}_v = \operatorname{diag}\left(\boldsymbol{\Phi}_v'^T \boldsymbol{\Phi}_0 \boldsymbol{\Sigma} \right) - \operatorname{diag}(\boldsymbol{\mu}) \boldsymbol{\Phi}_v'^T (\mathbf{y} - \boldsymbol{\Phi}_0 \boldsymbol{\mu}) \tag{40g}$$

$$\epsilon = \operatorname{tr}\left\{ \boldsymbol{\Phi}_\eta' \Sigma \boldsymbol{\Phi}_\eta'^T \right\} + \operatorname{tr}\{\boldsymbol{\mu}^T \boldsymbol{\Phi}_\eta'^T \boldsymbol{\Phi}_\eta' \boldsymbol{\mu}\} \tag{40h}$$

$$\varrho = \operatorname{tr}\left\{ \boldsymbol{\Phi}_0 \Sigma \boldsymbol{\Phi}_\eta'^T \right\} - \operatorname{tr}\left\{ (\mathbf{y} - \boldsymbol{\Phi}_0 \boldsymbol{\mu})^T \boldsymbol{\Phi}_\eta' \boldsymbol{\mu} \right\} \tag{40i}$$

Then, the goal of this optimization problem is to estimate the offsets $\boldsymbol{\delta}_g = [\boldsymbol{\delta}_u, \boldsymbol{\delta}_v]$ and δ_η. The key derivations are presented as follows [32]. Note that the goal function in (39a) is convex. The estimate of $\boldsymbol{\delta}_u$ can be obtained by taking partial derivate w.r.t. $\boldsymbol{\delta}_u$, and setting the result equal to zero

$$\frac{\partial f}{\partial \boldsymbol{\delta}_u} = 2\left(\boldsymbol{M}_{uu} \boldsymbol{\delta}_u + \boldsymbol{M}_{uv} \boldsymbol{\delta}_v + \delta_\eta \boldsymbol{\nu}_{u\eta} + \boldsymbol{\nu}_u \right) = 0 \tag{41}$$

where if $\boldsymbol{M}_{uu}$ is invertible, we can obtain the estimate of $\boldsymbol{\delta}_u$ directly, i.e.,

$$\check{\boldsymbol{\delta}}_u = -\boldsymbol{M}_{uu}^{-1} (\boldsymbol{M}_{uv} \boldsymbol{\delta}_v + \delta_\eta \boldsymbol{\nu}_{u\eta} + \boldsymbol{\nu}_u) \tag{42}$$

Otherwise, if $\boldsymbol{M}_{uu}$ is noninvertible, we can utilize other elements but $(\delta_u)_j$ to update $(\delta_u)_j$. To do this, we firstly rewrite the right-hand side of (41), and then set it equal to zero.

$$(\boldsymbol{M}_{uu}^j)_{-j} (\delta_u)_{-j} + (M_{uu})_{jj} (\delta_u)_j + (\boldsymbol{M}_{uv} \boldsymbol{\delta}_{\boldsymbol{\nu}} + \delta_\eta \boldsymbol{\nu}_{u\eta} + \boldsymbol{\nu}_u)_j = 0$$

(43)

After solving the above equation, the element of $\bar{\boldsymbol{\delta}_u}$ can be expressed as

$$\left(\bar{\delta}_u\right)_j = -\frac{\left(\boldsymbol{M}_{uu}^j\right)_{-j}(\boldsymbol{\delta}_u)_{-j} + \left(\boldsymbol{M}_{uv}\boldsymbol{\delta}_v + \delta_\eta \boldsymbol{\nu}_{u\eta} + \boldsymbol{\nu}_u\right)_j}{(M_{uu})_{jj}} \tag{44}$$

In total, based on (42) and (44), the update formula of $\boldsymbol{\delta}_u$ can be expressed as

$$\left(\delta_u^{new}\right)_j = \begin{cases} (LB_u)_j, & \text{if } \left(\tilde{\delta}_u\right)_j < (LB_u)_j \\ \left(\tilde{\delta}_u\right)_j, & \text{otherwise} \\ (UB_u)_j, & \text{if } \left(\tilde{\delta}_u\right)_j > (LB_u)_j \end{cases} \tag{45}$$

where

$$(\tilde{\delta_u})_j = \begin{cases} (\check{\delta}_u)_j, & \text{if } \boldsymbol{M}_{uu} \text{ is invertible;} \\ (\bar{\delta}_u)_j, & \text{otherwise} \end{cases} \tag{46}$$

Similarity, the update formula of $\boldsymbol{\delta}_v$ can be expressed as

$$\left(\delta_v^{new}\right)_j = \begin{cases} (LB_v)_j, & \text{if } \left(\tilde{\delta}_v\right)_j < (LB_v)_j \\ \left(\tilde{\delta}_v\right)_j, & \text{otherwise} \\ (UB_v)_j, & \text{if } \left(\tilde{\delta}_v\right)_j > (LB_v)_j \end{cases} \tag{47}$$

where

$$(\tilde{\delta_v})_j = \begin{cases} (\check{\delta}_v)_j, & \text{if } \boldsymbol{M}_{vv} \text{ is invertible;} \\ (\bar{\delta}_v)_j, & \text{otherwise} \end{cases} \tag{48}$$

with

$$(\check{\delta}_v)_j = \left(-\boldsymbol{M}_{vv}^{-1}\left(\boldsymbol{M}_{uv}^T\boldsymbol{\delta}_u + \delta_\eta \boldsymbol{\nu}_{v\eta} + \boldsymbol{\nu}_v\right)\right)_j \tag{49a}$$

$$\left(\bar{\delta}_v\right)_j = -\frac{\left(\boldsymbol{M}_{vv}^j\right)_{-j}(\boldsymbol{\delta}_v)_{-j} + \left(\boldsymbol{M}_{uv}^T\boldsymbol{\delta}_u + \delta_\eta \boldsymbol{\nu}_{v\eta} + \boldsymbol{\nu}_v\right)_j}{(M_{vv})_{jj}} \tag{49b}$$

Similarly, the PLE offset δ_η can be updated as follows:

$$\delta_\eta^{\text{new}} = -\frac{\nu_{u\eta}^T\boldsymbol{\delta}_u + \nu_{v\eta}^T\boldsymbol{\delta}_v + \varrho}{\epsilon} \tag{50}$$

where $\left(\delta_\eta^{\text{new}}\right) \in \left[\left(LB_\eta\right), \left(UB_\eta\right)\right]$.

Note that the sparse coefficients $\boldsymbol{\omega}$, hyper-parameters $\boldsymbol{\lambda}$ and β, and dictionary parameters η and $\mathcal{G}$ are iteratively estimated in the localization procedure until the convergence criterion meets, when these parameters will be further employed for sources localization.

3.3.2. Localization

When the convergence criterion meets, the estimated sparse vector $\hat{\boldsymbol{\omega}}$ and dictionary parameters $\hat{\mathcal{G}}, \hat{\eta}$ can be obtained. Specifically, the PLE is estimated by parameter $\hat{\eta}$, and the sources' locations are estimated from $\hat{\mathcal{G}}$. And among all grid points $\hat{\mathcal{G}}$, the grid points with the K largest spatial power spectrum $\hat{P}$ are the source positions, and the computation of $\hat{P}$ is as follows.

$$\hat{P} = E(\boldsymbol{\omega}) = \boldsymbol{\mu} = \hat{\boldsymbol{\omega}} \tag{51}$$

To summarize, the SBMSL method is depicted in Algorithm 1. At each iteration of the proposed algorithm, the sparse coefficients $\boldsymbol{\omega}$, hyper-parameters $\boldsymbol{\lambda}$ and β, and dictionary parameters η and $\mathcal{G}$ are re-estimated. In detail, the covariance parameter Σ of sparse coefficient parameter $\boldsymbol{\omega}$, hyper-parameters $\boldsymbol{\lambda}$ and β are obtained after iteration SSR converges, then by using these updated parameters, new dictionary parameters $\mathcal{G}^{(t+1)}, \eta^{(t+1)}$ are obtained when the iteration DL converges.

Algorithm 1 Sparse Bayesian Multiple Source Localization Method.

Input: Observations $\mathbf{y}$, propagation model $f\left(s_i, g_j, \eta\right)$, sensors localization $\mathbb{S}$, K, M, N.
Output: PLE $\hat{\eta}$ and sources localizations $\mathbb{L}$

```
Initialize t = 0, a, b, c, d, G^(0), η^(0), β^(0), λ^(0), ς^(0)
while convergence criterion not met do
    Build Φ_0, Φ'_u, Φ'_v, Φ'_η based on G^(t) and η^(t)
    Φ = Φ_0, β = β^(0), λ = λ^(0), ς = ς^(0)
    while SSR convergence criterion not met do
        Update ς(ε) and ε by (18) and (23)
        Update Σ and μ using (33), (31)
        Update λ using (36)
        Update β using (37)
    end while
    k = 0, G^(0) = G^(t), η^(0) = η^(t),
    while DL convergence criterion not met do
        Compute δ_g,δ_η by solving (39a)
        G^(k+1) = G^(k) + δ_g and η^(k+1) = η^k + δ_η
        Update Φ_0, Φ'_u, Φ'_v, Φ'_η according to G^(k+1) and η^(k+1)
        Update the iteration number k = k + 1
    end while
    G^(t+1) = G^(k), η^(t+1) = η^(k), t = t + 1
end while
ω̂ = μ, Ĝ = G, η̂ = η;
Estimate the sources localizations L from Ĝ with the K largest spatial power
    spectrum by (51)
return η̂, L
```

3.3.3. Complexity analysis of SBMSL

Based on the above analysis, part of the dictionary matrix $\boldsymbol{\Phi}$ can be considered where the components of $\boldsymbol{\omega}$ are nonzero and this means that the DL step can be constrained to these GPs where the element of $\boldsymbol{\omega}$ is non-zero. And the number of nonzero value of coefficient $\boldsymbol{\omega}$ is denoted as N_0. From Algorithm 1, we can know that: building dictionary at Step (3) requires $\mathcal{O}(MN)$ operations; updating $\boldsymbol{\varepsilon}$ at Step (6) requires $\mathcal{O}(N)$; updating $\boldsymbol{\mu}$ at Step (7) requires $\mathcal{O}(M^3 + MN + N^2M)$; updating $\boldsymbol{\lambda}$ at Step (8) requires $\mathcal{O}(N)$; updating β at Step (9) requires $\mathcal{O}(M + MN + N)$; computing the offsets at Step (13) needs $\mathcal{O}(N_0M^2)$; updating the dictionary at Step (15) requires $\mathcal{O}(N_0M)$. Note that $K \leq N_0 \ll N$ and $K < M \ll N$, thus the theoretical asymptotic complexity of SBMSL is $\mathcal{O}(N^2M)$, and it is the same as conventional methods such as PSBDL [38] and GEMTL [39]. This computational complexity makes SBMSL unsuitable for practical applications, particularly for large-scale problems [33]. Meanwhile, numerical errors [33] due to the underdetermined system (4) creates practical difficulties for SBMSL as well. To address the above-mentioned problems, a fast iterative strategy is developed for the model parameters estimations of SBMSL, yields the following fast algorithm, referred to as FastSBMSL.

3.4. FastSBMSL algorithm

In order to decrease the computational requirements and numerical errors, and concurrently promote sparsity, the FastSBMSL algorithm is presented. The related deduction is described as follows.

3.4.1. Fast suboptimal solution for estimating model parameters

Inspired by the work of [33], a fast solution for estimating the parameters is derived in this subsection. To further reduce the computations of the SBMSL algorithm, only a single λ_i rather than the whole vector $\boldsymbol{\lambda}$ will be updated at each iteration. Since it is assumed that the vector $\boldsymbol{\omega}$ is sparse, most of its components are zero; thus, most of its hyper-parameter λ_i are equal to zero, leading to μ_i being equals to zero and the corresponding entry being pruned out from the model [44], and the covariance matrix $\boldsymbol{\Sigma}$ can be expressed with fewer dimensions. Based on such properties, a much more efficient solution for parameters estimation can be ob-

tained. The related derivation starts from the lower bound of the logarithmic marginal likelihood $p(\mathbf{y}, \boldsymbol{\lambda}, \beta)$, which is obtained from $p(\mathbf{y}, \boldsymbol{\omega}, \boldsymbol{\lambda}, \beta)$ by integrating out $\boldsymbol{\omega}$,

$$\begin{aligned}\mathcal{L} &= \ln \int p(\mathbf{y}|\boldsymbol{\omega}, \beta) h(\boldsymbol{\omega}, \boldsymbol{\varepsilon}) p(\boldsymbol{\lambda}) p(\beta) d\boldsymbol{\omega} \\ &= \sum_{i=1}^{N} \left[-\varsigma(\varepsilon_i) \|\varepsilon_i\|_1^2 + \ln \lambda_i - \lambda_i \|\varepsilon_i\|_1 \right] \\ &\quad - \frac{1}{2} \ln \left(-2\varsigma(\varepsilon_1) \cdots - 2\varsigma(\varepsilon_N) \right) \\ &\quad - \frac{1}{2} \ln |\mathbf{C}| - \frac{1}{2} \left(\mathbf{y}^T \mathbf{C}^{-1} \mathbf{y} \right) \\ &\quad + \sum_{i=1}^{N} [(c-1) \ln \lambda_i - d\lambda_i] + \text{const}\end{aligned} \tag{52}$$

where the related deduction employed the property that $\|\varepsilon_i\|_1 = |\varepsilon_i|$, $\|\omega_i\|_1 = |\omega_i|$ for the scalar RSS observations, and the matrix $\mathbf{C}$ can be rewritten as

$$\begin{aligned}\mathbf{C} &= \beta^{-1} I + \boldsymbol{\Phi} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}^T \\ &= \beta^{-1} I + \sum_i \gamma_i \phi_i \phi_i^T \\ &= \beta^{-1} I + \sum_{j \neq i} \gamma_j \phi_j \phi_j^T + \gamma_i \phi_i \phi_i^T \\ &= \mathbf{C}_{-i} + \gamma_i \phi_i \phi_i^T\end{aligned} \tag{53}$$

where ϕ_i is the i-th column vector of matrix $\boldsymbol{\Phi}$, and

$$\gamma_i = \frac{1}{-2\varsigma(\varepsilon_i)} = \frac{\|\varepsilon_i\|_1}{\lambda_i}, i = 1, 2, \ldots, N \tag{54}$$

Note that, for convenience, the parameters γ_i and $\|\varepsilon_i\|_1$ instead of λ_i are iteratively estimated in the fast algorithm. Apply the matrix determinant and inversion lemmas to Eq. (53), respectively, yielding

$$|\mathbf{C}| = |\mathbf{C}_{-i}| |1 + \gamma_i \phi_i^T \mathbf{C}_{-i}^{-1} \phi_i| \tag{55}$$

$$\mathbf{C}^{-1} = \mathbf{C}_{-i}^{-1} - \frac{\mathbf{C}_{-i}^{-1} \phi_i \phi_i^T \mathbf{C}_{-i}^{-1}}{1/\gamma_i + \phi_i^T \mathbf{C}_{-i}^{-1} \phi_i} \tag{56}$$

Substitute (54), (55) and (56) into (52), and treat $\mathcal{L}$ as the function of $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$. The logarithmic posterior $\mathcal{L}$ can be rewritten as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\varepsilon}) = &-\frac{1}{2} \Big[\log |\mathbf{C}_{-i}| + \mathbf{y}^T \mathbf{C}_{-i}^{-1} \mathbf{y} - \sum_{j \neq i} \ln \gamma_j \Big] \\ &+ \sum_{j \neq i} \left[-\varsigma(\varepsilon_j) \|\varepsilon_j\|_1^2 + \ln \|\varepsilon_j\|_1 - \ln \gamma_j - \|\varepsilon_j\|_1^2 / \gamma_j \right] \\ &+ \sum_{j \neq i} [(c-1) \ln(\|\varepsilon_j\|_1 / \gamma_j) - d(\|\varepsilon_j\|_1 / \gamma_j)] \\ &+ \frac{1}{2} \left[\ln \frac{1}{1 + \gamma_i \bar{s}_i} + \frac{q_i^2 \gamma_i}{1 + \gamma_i \bar{s}_i} + \gamma_i \right] \\ &+ \left[-\varsigma(\|\varepsilon_i\|_1) \|\varepsilon_i\|_1^2 + \ln(\|\varepsilon_i\|_1 / \gamma_i) - (|\varepsilon_i|^2 / \gamma_i) \right] \\ &+ \left[(c-1) \ln(\|\varepsilon_i\|_1 / \gamma_i) - d(\|\varepsilon_i\|_1 / \gamma_i) \right] \\ = &\mathcal{L}(\boldsymbol{\gamma}_{-i}, \boldsymbol{\varepsilon}_{-i}) + l(\gamma_i, \varepsilon_i)\end{aligned} \tag{57}$$

where $\mathcal{L}(\boldsymbol{\gamma}_{-i}, \boldsymbol{\varepsilon}_{-i})$ denotes the first three terms of the right-hand equation, which is independent of γ_i, $l(\gamma_i, \varepsilon_i)$ denotes the last three terms of the right-hand equation, $\bar{s}_i$ and q_i are respectively defined as

$$\bar{s}_i = \phi_i^T \mathbf{C}_{-i}^{-1} \phi_i \tag{58}$$

$$q_i = \phi_i^T \mathbf{C}_{-i}^{-1} \mathbf{y}. \tag{59}$$

Now, in Eq. (57), the terms associated with a single hyperparameter γ_i are separated from others. Thus, a solution for maximizing $\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\varepsilon})$, when only the i-th component is changed, can be obtained by taking the derivative of the goal function $\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\varepsilon})$ with respect to γ_i and set the corresponding result equal to zero

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\varepsilon})}{\partial \gamma_i} &= \frac{\partial l(\gamma_i, \varepsilon_i)}{\partial \gamma_i} \\ &= \frac{\|\varepsilon_i\|_1^2}{2\gamma_i^2} - \frac{1}{2} \frac{\bar{s}_i}{1 + \gamma_i \bar{s}_i} + \frac{1}{2} \frac{q_i^2}{(1 + \gamma_i \bar{s}_i)^2} \\ &\quad - \frac{(c - \frac{1}{2})}{\gamma_i} + \frac{d\|\varepsilon_i\|_1}{\gamma_i^2} = 0\end{aligned} \tag{60}$$

For purpose of analysis and avoiding the complexity of the zero-finding of polynomials, we assume that $\gamma_i \bar{s}_i \gg 1$. Then, an approximate solution for γ_i can be obtained as

$$\gamma_i = \begin{cases} \frac{(\|\varepsilon_i\|_1^2 + 2d\|\varepsilon_i\|_1)\bar{s}_i^2 + q_i^2}{2c\bar{s}_i^2}, & \text{if } h_i > 20c\bar{s}_i \\ 0, & \text{otherwise,} \end{cases} \tag{61}$$

where we define $h_i = (\|\varepsilon_i\|_1^2 + 2d\|\varepsilon_i\|_1)\bar{s}_i^2 + q_i^2$ for simplification of forthcoming expressions. Similarly, the hyperparameters ε_i can be estimated by taking the derivative of $\mathcal{L}$ with respect to corresponding parameters, and then setting them to zero, yielding the following update equations

$$\|\varepsilon_i\|_1 = \frac{-d + \sqrt{d^2 + 4c\gamma_i}}{2} \tag{62}$$

Similarly to [44], the parameters s_i and q_i can be calculated for all vectors ϕ_is efficiently according to

$$\bar{s}_i = \frac{S_i}{1 - \gamma_i S_i} \tag{63}$$

$$q_i = \frac{Q_i}{1 - \gamma_i S_i} \tag{64}$$

with

$$S_i = \beta \phi_i^T \phi_i - \beta^2 \phi_i^T \boldsymbol{\Phi} \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \phi_i \tag{65}$$

$$Q_i = \beta \phi_i^T \mathbf{y} - \beta^2 \phi_i^T \boldsymbol{\Phi} \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{y} \tag{66}$$

To summarize, the procedure for FastSBMSL can be summarized as Algorithm 2. In the algorithm, we can only update a single parameter γ_i instead of the whole vector $\boldsymbol{\gamma}$ in an iterative way, and then update $\|\varepsilon_i\|_1$ by (62), finally update $\Sigma, \boldsymbol{\mu}, q_i, \bar{s}_i$ efficiently (see [[44], Appendix A] for more details). Moreover, from Algorithm 2, we can find that μ_i equals to zero when $\gamma_i = 0$, thus the corresponding vector should be pruned out from the sparse model. As more γ_i become zero, the coefficient vector $\boldsymbol{\omega}$ becomes more sparse, whose dimension changes from $N \times 1$ to $N_0 \times 1$, where $N \geqslant N_0$. Therefore, part of the dictionary matrix $\boldsymbol{\Phi}$ cannot be considered where the components of $\boldsymbol{\omega}$ are zero. This means that the dimension of dictionary matrix $\boldsymbol{\Phi}$ can be reduced to $M \times N_0$. Owing to the reduced dimensions of the coefficient vector and dictionary matrix, less computing can be expected.

Algorithm 2 Fast Sparse Bayesian Multiple Source Localization Method.

Input: K, M, N, Observation $\mathbf{y}$, propagation model $f(s_i, g_j, \eta)$, sensors localization $\mathbb{S}$.
Output: PLE $\hat{\eta}$ and sources localizations $\mathbb{L}$

```
Initialize t = 0, ε^(0), G^(0), η^(0), all γ_i = 0
while convergence criterion not met do
    Compute Φ_0, Φ'_u, Φ'_v, Φ'_η according to G^t and η^t
    Φ = Φ_0, ε = ε^(0)
    while convergence criterion not met do
        Choose a γ_i
        if h_i > 20cs̄_i and γ_i = 0 then
            Add γ_i to the model
        else
            if h_i > 20cs̄_i and γ_i > 0 then
                Re-estimate γ_i;
            else {h_i < 20cs̄_i}
                Prune i from the model;
            end if
        end if
        Update Σ and μ
        Update s̄_i, q_i
        Update ‖ε_i‖_1 by (62)
    end while
    Update β using (37)
    k = 0, G^(0) = G^(t), η^(0) = η^(t),
    while DL convergence criterion not met do
        Compute δ_g,δ_η by solving (39a)
        G^(k+1) = G^(k) + δ_g and η^(k+1) = η^k + δ_η
        Update Φ_0, Φ'_u, Φ'_v, Φ'_η according to G^(k+1) and η^(k+1)
        Update the iteration number k = k + 1
    end while
    G^(t+1) = G^(k), η^(t+1) = η^(k), t = t + 1
end while
ω̂ = μ, Ĝ = G, η̂ = η;
Estimate the sources localizations L from Ĝ with the K largest spatial power spectrum by (51)
return η̂, L
```

3.4.2. Complexity analysis of FastSBMSL

From Algorithm 2, we can know that: building dictionary at Step (3) requires $\mathcal{O}(MN)$ operations; updating $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ requires $\mathcal{O}(MN)$; updating β at Step (20) requires $\mathcal{O}(MN_0)$; computing the offsets at Step (23) needs $\mathcal{O}(N_0M^2)$; updating the dictionary at Step (25) requires $\mathcal{O}(N_0M)$. Note that $K \leq N_0 \ll N$ and $K < M \ll N$, thus the theoretical asymptotic complexity of FastSBMSL is $\mathcal{O}(NM)$.

4. Numerical simulations

In this section, numerical simulations are carried out to evaluate the performance of proposed SBMSL and FastSBMSL.

4.1. Evaluation metrics

The localization performance metric is widely used the root-mean-square error (RMSE) [32]:

$$\mathrm{RMSE} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left\|\boldsymbol{l}_k - \hat{\boldsymbol{l}}_k\right\|_2^2} \tag{67}$$

where $\boldsymbol{l}_k$ and $\hat{\boldsymbol{l}}_k$ denote the actual and estimated location of the k-th target, respectively.

The PLE performance metric is defined as the absolute error

$$error(\mathrm{PLE}) = \left|\eta - \hat{\eta}\right| \tag{68}$$

4.2. Simulations and result discussions

In the simulations, we set $K = 4$ source targets in a 25 m×25 m region comprising M uniformly deployed sensors. The region of interest is divided into N grid points to constitute a candidate GPs. In the RSS function, the transmitted power P_k and the reference distance d_0 are set as 100 and 1 m respectively. The value range of γ is [2, 4], and the initial value of η is $\eta = 2$. And the signal-to-noise ratio (SNR) is defined as [32]

$$\mathrm{SNR} = \frac{\sum_{k=1}^{K} P_k}{K\sigma^2} \tag{69}$$

To evaluate the proposed SBMSL and FastSBMSL, comparisons were made with other state-of-the-art MSL methods: the conventional sparse Bayesian learning [45]-based MSL method (referred to as SBL), the GEMTL method [38], the MSL method [39] developed based on Gaussian prior for the sparse coefficients (referred to as PSBDL), and the theoretical limits CRLB (derived in Appendix D), under the same conditions for various scenarios, including different signal-noise-ratio (SNR) conditions, different numbers of sensors, and different grid granularities. Specifically, SBL uses hierarchical sparse prior, a fixed localization dictionary based on initialized parameters and a fixed PLE, GEMTL adopts hierarchical sparse prior, an adaptive localization dictionary and a fixed PLE. The simulation results are averaged over 500 randomized trials carried out in MATLAB R2019a on a PC with Windows 10 OS and an Intel i7-7700 CPU.

4.2.1. Impacts of observation noise

To evaluate the impacts of observation noise on the performance of the proposed methods, simulations were performed under different SNR levels while M = 50 and N = 15×15, where the SNR is changed from 0 dB to 40 dB at a step size of 5 dB.

Fig. 1 illustrates the localization performance of respective methods in terms of RMSE. When SNR = 0 dB, the observations are almost masked by noises, yielding remarkable performance deterioration of all methods. As SNR gets larger, the RMSE of each method witnesses a monotonous decrease, which means that fewer noises disturb the observations as SNR gets larger, and more accurate localization can be obtained. When SNR > 25 dB, the RMSE of all methods tends to be stable, indicating that the performance improvement will be very small when SNR is large enough. Meanwhile, the results show that the PSBDL, the proposed SBMSL and FastSBMSL perform better than SBL and GEMTL and their results are closer to CRLB, which may be attributed to the adaptive inference of PLE and the proper grid points $\mathcal{G}$ for the former three methods. Compared with PSBDL, the proposed SBMSL and FastSBMSL performs better when SNR >10 dB, which implies that by adopting Laplacian prior for the sparse coefficients, the proposed methods can provide better sparse recovery and then obtain better localization accuracy. Yet, the PSBDL performs better than the proposed methods when 5 dB ⩽SNR ⩽ 10 dB. This may because that, to tackle the problem that Laplace prior is not conjugate to the Gaussian likelihood, a variational approximation is employed for the objective function in the update of the sparse coefficients and model parameters, where this approximation may yield adverse effects for the parameters estimation under lower SNR conditions and finally degrade the localization performance of the proposed methods. Furthermore, it can be seen that the FastSBMSL and SBMSL obtain comparable estimation accuracy. Compared with SBMSL, the FastSBMSL performs even better although it is suboptimal in theory, indicating that the FastSBMSL not only reduced computation complexity but also obtained improved estimation accuracy. Actually, the system in (4) is underdetermined with $M \ll N$. Compared with FastSBMSL, the higher computational cost and increased numerical errors [33] of SBMSL may degrade its performance in practice.

Additionally, Table 2 reports the average PLE errors of respective methods versus SNR. Noticeably, the PLE results are different from

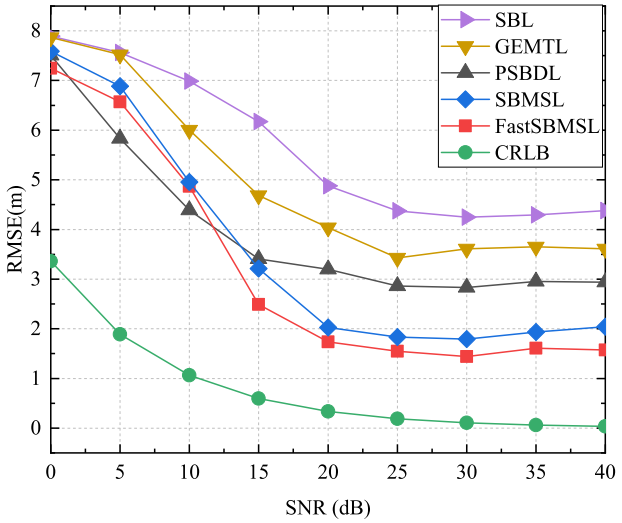

Fig. 1. The RMSE of different approaches versus SNR.

Table 2
PLE Error Results Versus SNR.

SNR (dB)	0	5	10	15	20
PSBDL	0.9872	0.9419	0.8858	0.8285	0.6344
SBMSL	0.9874	0.9794	0.9541	0.7233	0.4306
FastSBMSL	0.9574	0.8944	0.8147	0.5724	0.3114
CRLB	0.9253	0.7894	0.5012	0.2892	0.1626
SNR (dB)	25	30	35	40	
PSBDL	0.5659	0.5891	0.5387	0.5388	
SBMSL	0.2365	0.1766	0.1557	0.1670	
FastSBMSL	0.2128	0.1811	0.2041	0.2055	
CRLB	0.0914	0.0514	0.0286	0.0162	

the RMSE results of Fig. 1. The PLE results of SBMSL are better than those of PSBDL for SNR $>$ 10 dB, whereas they are worse for lower SNR, whose tendency is similar with that of Fig. 1. By comparison, the PLE results of FastSBMSL are better than those of PSBDL even for lower SNR, which is different from Fig. 1. The performance differences of RMSE and PLE under lower SNR for FastSBMSL imply that adverse effects under low SNR conditions for the PLE estimation are not as bad as those of RMSE. Explore of FastSBMSL upon the parameters estimation under lower SNR will be our future work.

4.2.2. Impacts of the number of sensors

Herein, we studied the effects of the number of sensors, varying from 30 to 50 at a step size of 5 for SNR = 20 dB and SNR = 10 dB respectively, while $N = 15\times15$. The RMSE results are depicted in Fig. 2 and Fig. 3.

From Fig. 2, it can be seen that the RMSE of all the methods gradually decreases as the number of sensors gets larger. Obviously, the number of observations employed for localization increases as the number of sensors increases, which probably implies more accurate localization. The results also demonstrate that the proposed SBMSL and FastSBMSL achieve better localization estimation precision than the other three methods with the same sensor number, and their results are closer to CRLB. Meanwhile, it can be observed that the FastSBMSL performs better than SBMSL, whereas the gap is not large. Fig. 3 shows that the performance of the proposed methods is better than SBL and GEMTL, whereas they are inferior to PSBDL when SNR = 10 dB. These results are consistent with those in Fig. 1.

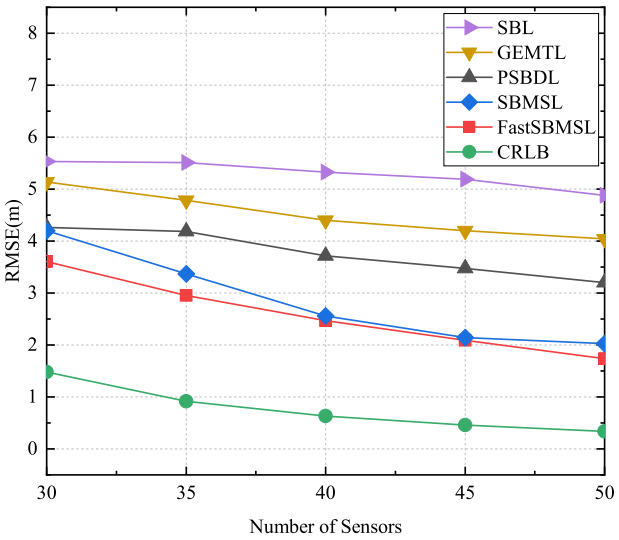

Fig. 2. The RMSE of different methods versus the number of sensors when SNR = 20 dB.

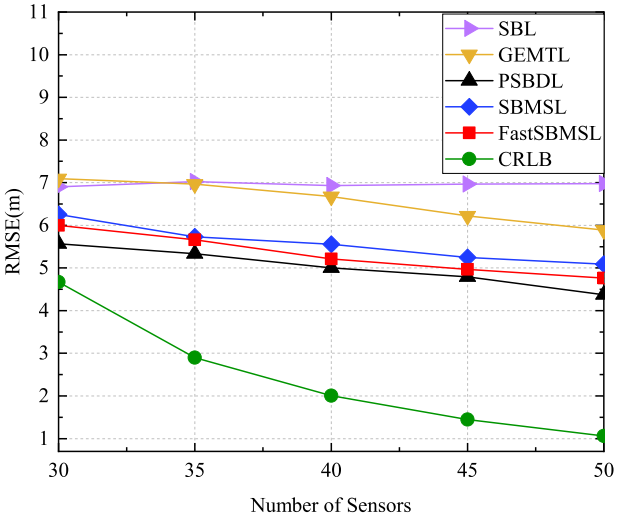

Fig. 3. The RMSE of different methods versus the number of sensors when SNR = 10 dB.

Table 3 illustrates the PLE estimate errors versus the number of sensors for two cases of SNR = 20 dB and SNR = 10 dB. The results of SNR = 20 dB show that the PLE estimation error decreases as the number of sensors increases. Actually, the more sensors, the more observations employed for the PLE estimation, probably implying better estimation accuracy. Yet, the improvements by increasing the number of sensors are not large. Moreover, it can be observed that the proposed methods obtain better PLE estimation than PSBDL and their results are closer to CRLB. Compared with the results of SNR = 20 dB, the PLE results for SNR = 10 dB are worse, indicating the performance degradation of all methods for low SNR.

4.2.3. Impacts of different grid granularities

It is crucial to explore the localization performance under different grid granularities since the grid granularity determines the complexity of the algorithm and the accuracy of the localization. Herein, the experiments were executed in the situation that the grid granularity (defined as $\sqrt{N}$) varies from 5 to 25 at a step size of 5 for SNR = 20 dB and SNR = 10 dB, respectively, when $M = 50$. The RMSE results are illustrated in Fig. 4 and Fig. 5.

Table 3
PLE Error Results Versus The Number Of Sensors.

	Sensors Number	30	35	40	45	50
SNR = 20 dB	PSBDL	0.7240	0.6735	0.6762	0.6619	0.6344
	SBMSL	0.6424	0.6376	0.5276	0.4874	0.4306
	FastSBMSL	0.4767	0.4671	0.3778	0.3335	0.3114
	CRLB	0.4626	0.3444	0.2556	0.2019	0.1626
SNR = 10 dB	PSBDL	0.8984	0.9020	0.8922	0.8706	0.8858
	SBMSL	0.9614	0.9548	0.9548	0.9537	0.9541
	FastSBMSL	0.7779	0.7696	0.8006	0.8046	0.8147
	CRLB	0.9145	0.7999	0.6893	0.5946	0.5012

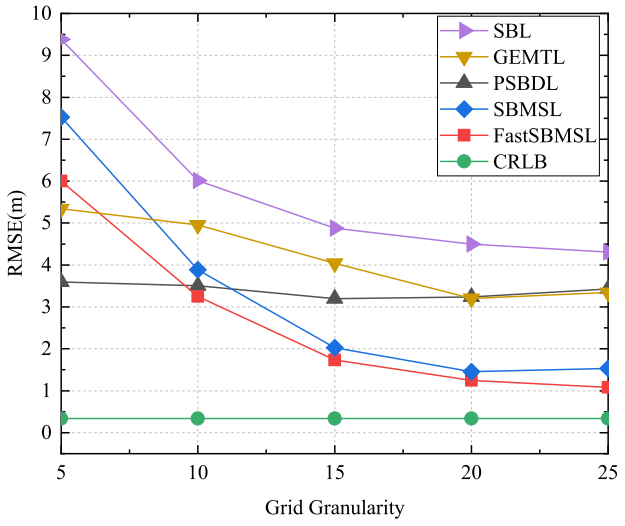

Fig. 4. The RMSE of different methods versus the grid granularity when SNR = 20 dB.

Fig. 4 show that the localization performance gradually improved as the grid granularity increased. This is because the finer the grid density, the higher possibility the grid set covers the sources locations, implying a better localization estimation. Meanwhile, we can also notice that the proposed SBMSL and FastSBMSL achieves better estimation accuracies than the other three methods and their results are closer to CRLB for the grid granularity larger than 10, indicating their performance superiority for fine grid granularity. The performance superiority of the proposed methods over the others becomes more remarkable as the grid granularity gets larger. However, when the grid is less than 10, the performance advantage of the proposed methods becomes weaker, and sometimes they perform worse, e.g., they obtain larger RMSEs than PSBDL and GEMTL for a grid granularity of 5. This may be because it is difficult for the grid set to cover the source locations for the coarse grid granularity in the proposed methods, yielding bad estimation accuracy.

Influenced by a low SNR level, the RMSE results in Fig. 5 are slightly different from those in Fig. 4. Although the proposed SBMSL and FastSBMSL perform better than SBL and GEMTL for the grid granularity larger than 10, their RMSE results are larger than PSBDL, which is consistent with the results of Fig. 1.

Table 4 shows the PLE estimation errors of all the methods for different grid granularities. It is clear that the PLE estimation errors of PSBDL, SBMSL and FastSBMSL decrease with gradually increased grid granularity.

4.2.4. Computational complexity

The theoretical asymptotic complexities of SBL, GEMTL, PSBDL and SBMSL are all $\mathcal{O}(N^2M)$, and the related value of FastSBMSL is $\mathcal{O}(NM)$. To evaluate the computational complexity of all methods

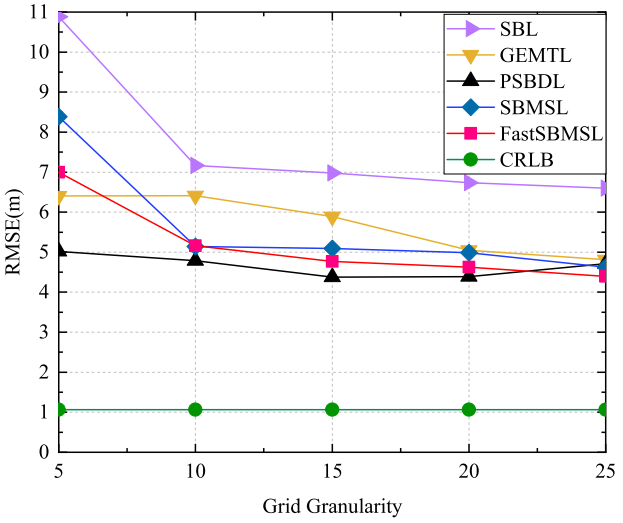

Fig. 5. The RMSE of different methods versus the grid granularity when SNR = 10 dB.

Table 4
PLE Error Results Versus The Grid Granularity.

	Grid Granularity	5	10	15	20	25
SNR = 20 dB	PSBDL	0.8055	0.5042	0.6344	0.6630	0.6418
	SBMSL	0.7496	0.5751	0.4306	0.3699	0.3356
	FastSBMSL	0.6630	0.3880	0.3114	0.2918	0.2985
	CRLB	0.1263	0.1263	0.1263	0.1263	0.1263
SNR = 10 dB	PSBDL	0.9760	0.9009	0.8858	0.9040	0.9047
	SBMSL	0.9575	0.9532	0.9541	0.9550	0.9552
	FastSBMSL	0.8562	0.7703	0.8147	0.8268	0.8268
	CRLB	0.2920	0.2920	0.2920	0.2920	0.2920

Table 5
CPU Running Times Versus Grid Granularity (Unit: Second).

Grid granularity	5	10	15	20	25
SBL	0.0273	0.0271	0.0402	0.0505	0.0646
GEMTL	0.9195	5.2652	11.1166	23.8898	57.3191
PSBDL	1.9162	4.2979	8.1071	20.2676	51.0358
SBMSL	3.1674	5.4797	7.3248	21.2270	44.7666
FastSBMSL	3.7538	4.2779	4.7291	4.5473	5.7858

clearly, the CPU running times w.r.t. different grid granularities are also given in Table 5, where SNR = 20 dB and $M = 50$. From the table, it can be seen that the running time of FastSBMSL is faster than GEMTL, PSBDL and SBMSL, showing its computational efficiency. Meanwhile, it can be also found that the running time of SBL is slower than FastSBMSL, although its theoretical asymptotic complexity is higher. This may because that the localization dictionary and PLE of SBL are fixed and require no inferences; thus, its running time can be faster. For a large-scale scenario, the computational efficiency of FastSBMSL would be demonstrated.

5. Conclusion

In this paper, the RSS-based MSL problem is addressed within the sparse Bayesian learning framework. To achieve improved sparsity, Laplace prior is employed to model the sparse coefficients. To tackle the problem that the Laplace prior is not conjugate to the Gaussian distribution and in order to estimate the related parameters via Bayesian inference, variational approximation for the Laplace prior is employed. Based on this approximation, the SBMSL algorithm is developed for MSL. Moreover, to reduce the com-

putational complexity, a fast algorithm, referred to as FastSBMSL, is further developed. In both SBMSL and FastSBMSL, all variables including sources localizations, sparse coefficients, and model parameters are estimated only from the RSS observations. Comparative experiments were made with the state-of-the-art RSS-based MSL methods. Numerical results demonstrated the validation of the proposed SBMSL and FastSBMSL.

CRediT authorship contribution statement

Roubing Tang: Methodology, Simulation Data curation, Writing-Original draft preparation. **Qiaoling Zhang:** Conceptualization, Investigation. **Weiwei Zhang:** Visualization, Validation. **Han Ma:** Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61806178), Zhejiang Provincial Natural Science Foundation of China (LY21F010015), Postdoctoral Science Foundation of China (2020M680932), and the Key Research and Development Program of Zhejiang Province (2021C01047).

Appendix A. Proof of (23)

To obtain (23), first we have

$$\begin{aligned}\mathcal{L}(\boldsymbol{\varepsilon}) &= \ln \int p(\mathbf{y}|\boldsymbol{\omega}) h(\boldsymbol{\omega}, \boldsymbol{\varepsilon}) d\boldsymbol{\omega} \\ &= \sum_{i=1}^{N} (-\varsigma(\varepsilon_i)\|\varepsilon_i\|_1^2) + \sum_{i=1}^{N} (-\lambda_i \|\varepsilon_i\|_1) \\ &- \frac{1}{2} \ln |\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \operatorname{diag}(-2\boldsymbol{\varsigma})| \\ &+ \frac{1}{2} \beta \boldsymbol{\Phi}^T \mathbf{y} [\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \operatorname{diag}(-2\boldsymbol{\varsigma})]^{-1} \beta \mathbf{y}^T \boldsymbol{\Phi} + \text{const}\end{aligned} \tag{A.1}$$

where const being the item constant to $\|\varepsilon_i\|_1$. To obtain the update formulation of $\|\varepsilon_i\|_1$, we set $\partial \mathcal{L}(\boldsymbol{\varepsilon})/\partial \|\varepsilon_i\|_1 = 0$, then we have

$$\begin{aligned}\|\varepsilon_i\|_1^2 =& \frac{1}{2} \left\{ \left[\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \operatorname{diag}(-2\boldsymbol{\varsigma}) \right]^{-1} \right\}_{ii} \\ &+ \frac{1}{2} \left\{ \left[\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \operatorname{diag}(-2\boldsymbol{\varsigma}) \right]^{-1} \beta \boldsymbol{\Phi}^T \mathbf{y} \beta \mathbf{y}^T \boldsymbol{\Phi} \right. \\ &\left. \times \left[\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \operatorname{diag}(-2\boldsymbol{\varsigma}) \right]^{-1} \right\}_{ii}\end{aligned} \tag{A.2}$$

Appendix B. Proof of (36)

First we have

$$\begin{aligned}\ln q^*(\boldsymbol{\lambda}) &= E_{\boldsymbol{\omega}} \{ \ln h(\boldsymbol{\omega}, \boldsymbol{\varepsilon}) p(\boldsymbol{\lambda}) \} + \text{const} \\ &= \sum_{i=1}^{N} \{ (-\frac{\|\varepsilon_i\|_1}{2} - d - \frac{\langle w_i^2 \rangle}{2\|\varepsilon_i\|_1}) \lambda_i \\ &+ c \ln \lambda_i \} + \text{const}\end{aligned} \tag{B.1}$$

It is seen that the posterior $q^*(\boldsymbol{\lambda})$ is the product of N independent Gamma distributions distributed

$$q^*(\lambda) = \prod_{i=1}^{N} \text{Gamma}\left(\lambda_i | \widetilde{c}, \widetilde{d}\right) \tag{B.2}$$

So we have

$$\lambda_i = \frac{2c+2}{d + \|\varepsilon_i\|_1 + \frac{\langle w_i^2 \rangle}{\|\varepsilon_i\|_1}} \tag{B.3}$$

Appendix C. Proof of (37)

First we have

$$\begin{aligned}\ln q^*(\beta) &= E_{\boldsymbol{\omega}} \{ \ln p(\mathbf{y}|\boldsymbol{\omega}, \beta) p(\beta | c, d) \} + \text{const} \\ &= \frac{N}{2} \log \beta - \frac{\beta}{2} \left\langle \|\mathbf{y} - \Phi \mathbf{w}\|^2 \right\rangle \\ &+ (a-1) \log \beta - b\beta + \text{ const} \\ &= \left(a + \frac{M}{2} - 1 \right) \log \beta \\ &- \left(\frac{1}{2} \left\langle \|\mathbf{y} - \Phi \mathbf{w}\|^2 \right\rangle + b \right) \beta + \text{ const}\end{aligned} \tag{C.1}$$

Obviously, the posterior $q^*(\beta)$ is Gamma distribution

$$q^*(\beta) = \text{Gamma}(\beta | \widetilde{a}, \widetilde{b}) \tag{C.2}$$

So, we have

$$\beta = \frac{M + 2a}{2b + \|\mathbf{y} - \boldsymbol{\Phi}\mu\|_2^2 + \operatorname{Tr}(\Sigma \boldsymbol{\Phi}^T \boldsymbol{\Phi})} \tag{C.3}$$

Appendix D. Cramer-Rao lower bound analysis

Denote $\vartheta = \left[u_1, v_1, P_1, \cdots, u_K, v_K, P_K, \eta, \sigma^2\right]^T$ as the unknown parameter vector. The Cramer-Rao lower bound (CRLB) for ϑ is derived herein. In estimation theory, the CRLB provides a theoretical performance limit for any unbiased estimator of the source locations $[u_k, v_k]$, the transmitted powers P_k and PLE η in the presence of unknown Gaussian noise variance β, given the observation $\mathbf{y}$.

The CRLB gives a lower bound for the error covariance matrix $\boldsymbol{F}$ of ϑ, which can be solved by inequality sign defined in the positive-semidefinite (PSD) sense.

$$\boldsymbol{F} = E\left\{ (\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})^T \right\} \geq \boldsymbol{J}^{-1} \tag{D.1}$$

where $\boldsymbol{J}$ is the Fisher information matrix (FIM) defined as follows.

$$\boldsymbol{J} = E\left\{ -\Delta_{\vartheta}^{\vartheta} \ln p(\mathbf{y}; \vartheta) \right\} \tag{D.2}$$

with $\Delta_{\vartheta}^{\vartheta} = \nabla_{\vartheta} \nabla_{\vartheta}^T$ being the second derivative (Hessian) operator, and ∇_{ϑ} being the gradient operator w.r.t. ϑ. The PDF of $\mathbf{y}$ is $p(\boldsymbol{y}; \boldsymbol{\vartheta}) = \mathcal{N}(\boldsymbol{y} \mid \tilde{\boldsymbol{\mu}}(\boldsymbol{\vartheta}), \boldsymbol{C}(\boldsymbol{\vartheta}))$ with $\tilde{\mu}_i = \sum_{k=1}^{K} P_k f(s_i, l_k, \eta), i = 1, \ldots, M$ and $\boldsymbol{C} = \sigma^2 \boldsymbol{I}_M$. For the Gaussian observation vector $\mathbf{y}$, the (i,j)-th element of the FIM $\boldsymbol{J}$ can be computed as [46]

$$[\boldsymbol{J}]_{i,j} = \frac{\partial \tilde{\boldsymbol{\mu}}^T}{\partial \vartheta_i} C^{-1} \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \vartheta_j} + \frac{1}{2} \operatorname{tr}\left(C^{-1} \frac{\partial \boldsymbol{C}}{\partial \vartheta_i} C^{-1} \frac{\partial \boldsymbol{C}}{\partial \vartheta_j} \right) \tag{D.3}$$

with

$$\frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \vartheta_i} = \left[\frac{\partial \tilde{\mu}_1}{\partial \vartheta_i}, \ldots, \frac{\partial \tilde{\mu}_M}{\partial \vartheta_i} \right]^T \tag{D.4a}$$

$$\frac{\partial \boldsymbol{C}}{\partial \vartheta_i} = \frac{\partial \sigma^2}{\partial \vartheta_i} \boldsymbol{I}_M \tag{D.4b}$$

For the unknown parameters $\vartheta_k, k = 1, \ldots, K$ in $\boldsymbol{\vartheta}$, the corresponding partial derivative terms in (D.4) are

$$\frac{\partial \tilde{\mu}_i}{\partial u_k} = P_k \frac{\partial f(s_i, l_k, \eta)}{\partial u_k} = -\eta P_k \frac{u_k - u_i^s}{d_{ik}^{\eta+2}}, \tag{D.5a}$$

$$\frac{\partial \tilde{\mu}_i}{\partial v_k} = P_k \frac{\partial f(s_i, l_k, \eta)}{\partial v_k} = -\eta P_k \frac{v_k - v_i^s}{d_{ik}^{\eta+2}}, \tag{D.5b}$$

$$\frac{\partial \tilde{\mu}_i}{\partial P_k} = f(s_i, l_k, \eta) = \frac{1}{d_{ik}^{\eta}}, \tag{D.5c}$$

$$\frac{\partial \tilde{\mu}_i}{\partial \eta} = \frac{\partial \sum_{k=1}^{K} P_k f(s_i, l_k, \eta)}{\partial \eta} = -\sum_{k=1}^{K} P_k \frac{\ln d_{ik}}{d_{ik}^{\eta}}, \tag{D.5d}$$

$$\frac{\partial \tilde{\mu}_i}{\partial \sigma^2} = 0, \tag{D.5e}$$

$$\frac{\partial \sigma^2}{\partial u_k} = 0, \frac{\partial \sigma^2}{\partial v_k} = 0, \frac{\partial \sigma^2}{\partial P_k} = 0, \frac{\partial \sigma^2}{\partial \eta} = 0, \frac{\partial \sigma^2}{\partial \sigma^2} = 1. \tag{D.5f}$$

Then, from (D.3) and (D.5), the FIM J can be computed as

$$\boldsymbol{J} = \sigma^2 \begin{bmatrix} \frac{\partial \tilde{\boldsymbol{\mu}}^T}{\partial \boldsymbol{\vartheta}_{1:3K+1}} \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\vartheta}_{1:3K+1}^T} & 0 \\ 0 & \frac{M}{2\sigma^2} \end{bmatrix} \tag{D.6}$$

with

$$\frac{\partial \tilde{\boldsymbol{\mu}}^T}{\partial \boldsymbol{\vartheta}_{1:3K+1}} = \left[\frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\vartheta}_{1:3K+1}^T}\right]^T = \begin{bmatrix} \frac{\partial \tilde{\mu}_1}{\partial \vartheta_1} & \cdots & \frac{\partial \tilde{\mu}_M}{\partial \vartheta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{\mu}_1}{\partial \vartheta_{3K+1}} & \cdots & \frac{\partial \tilde{\mu}_M}{\partial \vartheta_{3K+1}} \end{bmatrix} \tag{D.7}$$

Consequently, the error covariance matrix $\boldsymbol{F}$ can be summarized as

$$\boldsymbol{F} = \sigma^2 \begin{bmatrix} \left(\frac{\partial \tilde{\boldsymbol{\mu}}^T}{\partial \boldsymbol{\vartheta}_{1:3K+1}} \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\vartheta}_{1:3K+1}^T}\right)^{-1} & 0 \\ 0 & \frac{2\sigma^2}{M} \end{bmatrix} \tag{D.8}$$

Based on the above equation, the CRLB for the location estimate $\hat{l}_k$ and PLE estimate $\hat{\eta}$ can be expressed as

$$\mathrm{CRLB}\left(\hat{\boldsymbol{l}}_k\right) = \sqrt{[\boldsymbol{F}]_{3k-2} + [\boldsymbol{F}]_{3k-1}}, \quad \text{for } k = 1, \ldots, K \tag{D.9}$$

$$\mathrm{CRLB}(\hat{\eta}) = \sqrt{[\boldsymbol{F}]_{3K+1}}. \tag{D.10}$$

References

[1] M.A. Alsheikh, S. Lin, D. Niyato, H.-P. Tan, Machine learning in wireless sensor networks: algorithms, strategies, and applications, IEEE Commun. Surv. Tutor. 16 (4) (2014) 1996–2018, https://doi.org/10.1109/COMST.2014.2320099.

[2] D. Bao, C. Wang, J. Cai, Minimization of arc tangent function penalty for off-grid multi-source passive localization by using a moving array, Digit. Signal Process. 112 (2021) 103010.

[3] C. Shi, V. Andino-Pavlovsky, S.A. Lee, T. Costa, J. Elloian, E.E. Konofagou, K.L. Shepard, Application of a sub–0.1-mm3 implantable mote for in vivo real-time wireless temperature sensing, Sci. Adv. 7 (19) (2021) eabf6312.

[4] P. Saini, R.P. Singh, A. Sinha, Survey of radio propagation models for acoustic applications in underwater wireless sensor network, Int. J. Sensors Wireless Commun. Control 11 (1) (2021) 42–53.

[5] S. Thouhid, et al., Analysis of Water Contamination and Reporting Using Wireless Sensors Network, vol. 1964, IOP Publishing, 2021, 062030.

[6] Z. Feng, O. Shimizu, H. Sumiya, S. Nagai, H. Fujimoto, M. Sato, Influence of contamination between receiver coil and embedded transmitter coil for dynamic wireless power transfer system, in: 2021 IEEE PELS Workshop on Emerging Technologies: Wireless Power Transfer (WoW), IEEE, 2021, pp. 1–6.

[7] G. Patra, L. Goswami, Forest protection using wireless sensor network and iot, Materials Today: Proceedings (2021), https://doi.org/10.1016/j.matpr.2021.03.742.

[8] V. Rampa, G.G. Gentili, S. Savazzi, M. D'Amico, Electromagnetic models for passive detection and localization of multiple bodies, arXiv preprint, arXiv:2104.07354.

[9] S. Manickam, S.C. Swar, D.W. Casbeer, S.G. Manyam, Multi-unmanned aerial vehicle multi acoustic source localization, Proc. Inst. Mech. Eng., G J. Aerosp. Eng. 235 (3) (2021) 273–294.

[10] M.P. Michaelides, C. Laoudias, C.G. Panayiotou, Fault tolerant localization and tracking of multiple sources in WSNs using binary data, IEEE Trans. Mob. Comput. 13 (6) (2014) 1213–1227.

[11] J. Yang, Y. Yang, J. Lu, A variational Bayesian strategy for solving the DOA estimation problem in sparse array, Digit. Signal Process. 90 (2019) 28–35.

[12] Y. Zhang, Y.I. Wu, Multiple sources localization by the WSN using the direction-of-arrivals classified by the genetic algorithm, IEEE Access 7 (2019) 173626–173635, https://doi.org/10.1109/ACCESS.2019.2956825.

[13] Y. Hu, T.D. Abhayapala, P.N. Samarasinghe, Multiple source direction of arrival estimations using relative sound pressure based music, IEEE/ACM Trans. Audio Speech Lang. Process. 29 (2021) 253–264, https://doi.org/10.1109/TASLP.2020.3039569.

[14] H. Shen, Z. Ding, S. Dasgupta, C. Zhao, Multiple source localization in wireless sensor networks based on time of arrival measurement, IEEE Trans. Signal Process. 62 (8) (2014) 1938–1949, https://doi.org/10.1109/TSP.2014.2304433.

[15] C. Jia, D. Wang, J. Yin, X. Chen, L. Zhang, Joint multiple sources localization using TOA measurements based on Lagrange programming neural network, IEEE Access 7 (2019) 3247–3263, https://doi.org/10.1109/ACCESS.2018.2886909.

[16] T.-K. Le, K.C. Ho, Joint source and sensor localization by angles of arrival, IEEE Trans. Signal Process. 68 (2020) 6521–6534, https://doi.org/10.1109/TSP.2020.3037412.

[17] P. Qian, Y. Guo, N. Li, D. Fang, Compressive sensing based multiple source localization in the presence of sensor position uncertainty and nonuniform noise, IEEE Access 6 (2018) 36571–36583, https://doi.org/10.1109/ACCESS.2018.2852296.

[18] Y. Chu, K. You, W. Guo, Multiple sources localization with sparse recovery under log-normal shadow fading, arXiv:2105.15097, 2021.

[19] W. Meng, W. Xiao, Energy-based acoustic source localization methods: a survey, Sensors 17 (2) (2017) 376.

[20] L. Lin, H.-C. So, F.K. Chan, Multidimensional scaling approach for node localization using received signal strength measurements, Digit. Signal Process. 34 (2014) 39–47.

[21] X. Sheng, Y.-H. Hu, Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks, IEEE Trans. Signal Process. 53 (1) (2004) 44–53.

[22] X. Chen, U. Tureli, Underwater source localization based on energy measurement with randomly distributed sensor array, in: Unattended Ground, Sea, and Air Sensor Technologies and Applications IX, vol. 6562, International Society for Optics and Photonics, 2007, p. 65620L.

[23] W. Meng, W. Xiao, L. Xie, An efficient em algorithm for energy-based multisource localization in wireless sensor networks, IEEE Trans. Instrum. Meas. 60 (3) (2011) 1017–1027.

[24] L. Lu, H. Zhang, H.-C. Wu, Novel energy-based localization technique for multiple sources, IEEE Syst. J. 8 (1) (2013) 142–150.

[25] X. Sheng, Y.-H. Hu, P. Ramanathan, Distributed particle filter with gmm approximation for multiple targets localization and tracking in wireless sensor network, in: IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005, IEEE, 2005, pp. 181–188.

[26] J. Chen, U. Mitra, Underwater acoustic source localization using unimodal-constrained matrix factorization, in: 2017 51st Asilomar Conference on Signals, Systems, and Computers, IEEE, 2017, pp. 2002–2006.

[27] J. Chen, U. Mitra, Unimodality-constrained matrix factorization for non-parametric source localization, IEEE Trans. Signal Process. 67 (9) (2019) 2371–2386.

[28] D. Model, M. Zibulevsky, Signal reconstruction in sensor arrays using sparse representations, Signal Process. 86 (3) (2006) 624–638.

[29] B. Sun, Y. Guo, N. Li, D. Fang, Multiple target counting and localization using variational Bayesian EM algorithm in wireless sensor networks, IEEE Trans. Commun. 65 (7) (2017) 2985–2998, https://doi.org/10.1109/TCOMM.2017.2695198.

[30] Y. Guo, B. Sun, N. Li, D. Fang, Variational Bayesian inference-based counting and localization for off-grid targets with faulty prior information in wireless sensor networks, IEEE Trans. Commun. 66 (3) (2018) 1273–1283, https://doi.org/10.1109/TCOMM.2017.2770139.

[31] P. Qian, Y. Guo, N. Li, Z. Xu, Block variational Bayesian algorithm for multiple target localization with unknown and time-varying transmit powers in WSNs, IEEE Access 7 (2019) 54796–54808, https://doi.org/10.1109/ACCESS.2019.2913369.

[32] K. You, W. Guo, T. Peng, Y. Liu, P. Zuo, W. Wang, Parametric sparse Bayesian dictionary learning for multiple sources localization with propagation parameters uncertainty, IEEE Trans. Signal Process. 68 (2020) 4194–4209, https://doi.org/10.1109/TSP.2020.3009875.
[33] S.D. Babacan, R. Molina, A.K. Katsaggelos, Bayesian compressive sensing using Laplace priors, IEEE Trans. Image Process. 19 (1) (2009) 53–63.
[34] Z.-M. Liu, Z.-T. Huang, Y.-Y. Zhou, An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning, IEEE Trans. Wirel. Commun. 11 (10) (2012) 1–11, https://doi.org/10.1109/TWC.2012.090312.111912.
[35] J. Yang, Y. Yang, Sparse Bayesian DOA estimation using hierarchical synthesis lasso priors for off-grid signals, IEEE Trans. Signal Process. 68 (2020) 872–884, https://doi.org/10.1109/TSP.2020.2967665.
[36] D. Wang, Z. Zhang, Variational Bayesian inference based robust multiple measurement sparse signal recovery, Digit. Signal Process. 89 (2019) 131–144.
[37] T. Clouqueur, K.K. Saluja, P. Ramanathan, Fault tolerance in collaborative sensor networks for target detection, IEEE Trans. Comput. 53 (3) (2004) 320–333.
[38] K. You, W. Guo, Y. Liu, W. Wang, Z. Sun, Grid evolution: joint dictionary learning and sparse Bayesian recovery for multiple off-grid targets localization, IEEE Commun. Lett. 22 (10) (2018) 2068–2071.
[39] K. You, W. Guo, P. Zuo, Y. Liu, W. Wang, Sparse Bayesian learning for multiple sources localization with unknown propagation parameters, in: 2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2019, pp. 1–6.
[40] M.W. Seeger, H. Nickisch, Compressed sensing and Bayesian experimental design, in: Proceedings of the 25th International Conference on Machine Learning, 2008, pp. 912–919.
[41] J. Palmer, D. Wipf, K. Kreutz-Delgado, B. Rao, Variational EM algorithms for non-Gaussian latent variable models, Adv. Neural Inf. Process. Syst. 18 (2006) 1059.
[42] T.S. Jaakkola, M.I. Jordan, Bayesian parameter estimation via variational methods, Stat. Comput. 10 (1) (2000) 25–37.
[43] C. Wang, D.M. Blei, Variational inference in nonconjugate models, J. Mach. Learn. Res. 14 (4) (2013) 1005–1031.
[44] M.E. Tipping, A.C. Faul, Fast marginal likelihood maximisation for sparse Bayesian models, in: International Workshop on Artificial Intelligence and Statistics, in: PMLR, 2003, pp. 276–283.
[45] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Trans. Signal Process. 56 (6) (2008) 2346–2356.
[46] S.M. Kay, Fundamentals of Statistical Signal Processing: Practical Algorithm Development, vol. 3, Pearson Education, 2013.

Roubing Tang received the B.S. degree form Zhejiang Sci-Tech University, Hangzhou, China, in 2019, where she is currently pursuing the M.S. degree with the School of Electronics and Communication Engineering. Her current research interests include sound source localization, deep learning, Bayesian inference and speech enhancement.

Qiaoling Zhang received her Ph.D degrees in signal and information processing in the School of Information and Communication Engineering, Dalian University of Technology (DUT), Dalian, China, in 2016. She joined the School of Informatics and Electronics, Zhejiang Sci-Tech University, Hangzhou, China, as a Lecturer in 2017. Her research interests include speech processing, automatic speech recognition, object localization and tracking.

Weiwei Zhang received her Ph.D. degree in Signal and Information Processing in the School of Information and Communication Engineering, Dalian University of Technology (DUT), Dalian, China, in 2019. She is currently an associate professor in the Information Science and Technology College, Dalian Maritime University. Her research interests include music information retrieval, sound source localization and speech processing.

Han Ma received the B.S. degree form Zhejiang Sci-Tech University, Hangzhou, China, in 2018, where she is currently pursuing the M.S. degree with the School of Information and CommunicationEngineering. Her current research interests include automatic speech recognition, deep learning.