Robust Localization for Mixed LOS/NLOS Environments with Anchor Uncertainties

Yunfei Li, Shaodan Ma, Guanghua Yang and Kai-Kit Wong

Abstract

Localization is particularly challenging when the environment has mixed line-of-sight (LOS) and non-LOS paths and even more challenging if the anchors' positions are also uncertain. In the situations in which the parameters of the LOS-NLOS propagation error model and the channel states are unknown and uncertainties for the anchors exist, the likelihood function of a localizing node is computationally intractable. In this paper, assuming the knowledge of the prior distributions of the error model parameters and that of the channel states, we formulate the localization problem as the maximization problem of the posterior distribution of the localizing node. Then we apply variational distributions and importance sampling to approximate the true posterior distributions and estimate the target's location using an asymptotic minimum mean-square-error (MMSE) estimator. Furthermore, we analyze the convergence and complexity of the proposed variational Bayesian localization (VBL) algorithm. Computer simulation results demonstrate that the proposed algorithm can approach the performance of the Bayesian Cramer-Rao bound (BCRB) and outperforms conventional algorithms.

Index Terms

Bayesian Cramer-Rao bound (BCRB), variational Bayesian localization (VBL), mixed LOS/NLOS measurements, anchor node uncertainties, asymptotic minimum mean square error (MMSE).

Y. Li is the Department of Electrical and Computer Engineering, University of Macau, Taipa, Macao, China (email: yunfeili1989@gmail.com).

S. Ma is with the State Key Laboratory of Internet of Things for Smart City and the Department of Electrical and Computer Engineering, University of Macau, Taipa, Macao, China (e-mail: shaodanma@um.edu.mo).

G. Yang is with Institute of Physical Internet, Jinan University, Zhuhai Campus, Zhuhai 519070, China (e-mail: ghyang@jnu.edu.cn).

K. K. Wong is with the Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, United Kingdom (e-mail: kai-kit.wong@ucl.ac.uk).

I. Introduction

A. Motivation and Literature Review

Localization is essential for many applications such as smart transportation, search-and-rescue operations and etc. [1–3]. Localization problems in line-of-sight (LOS) only environments were relatively well understood, e.g., [4–10]. However, challenges arise in harsh environments such as indoor environments and urban areas [11, 12] and where the positions of the anchor nodes are uncertain [13–15]. In environments where there are many obstructions or scatterers, the paths between the anchor nodes and the localizing node may contain non-LOS (NLOS) paths [16]. As a result, localization needs to be achieved in a mixed LOS-NLOS environment and becomes much challenging. Another difficult situation is when the anchors' positions are only known coarsely [17–19]. Previous research tended to address these two issues in localization separately.

A mixture of LOS-NLOS paths in the measurements confuses localization and greatly degrades the localization accuracy [20–22]. This problem has been studied in the past and the methods can be classified into two approaches. One approach is to first identify the NLOS measurements and discard them when carrying out localization [21, 23, 24]. Another approach attempts to exploit the NLOS paths together with the LOS measurements to perform localization [25–30]. In the former approach, the NLOS-LOS identifications are usually based on hypothesis testing [21, 23, 24] and nonparametric machine learning [26–28]. In [21], the authors developed a NLOS detection method to discard NLOS measurements assuming that the Gauss-Newton method is unbiased in LOS environments. In [23], kurtosis was utilized to identify the LOS and NLOS measurements and then only the LOS measurements were used to complete estimation. In addition, the authors in [24] used a LOS Gaussian error distribution to design a confidence-region-based detector to distinguish the LOS pseudo-measured positions and select the LOS positions to complete localization.

Although the above-mentioned methods can achieve good localization accuracy, one would expect that the accuracy can be improved if the NLOS measurements are also employed in the localization instead of being discarded. To this end, [16] adopted the Gaussian mixture model (GMM) to represent the NLOS-LOS measurement errors. By using the extended Kalman filter (EKF), both NLOS and LOS measurements were combined to estimate the target's location. In [31], the localization problem was formulated as a reweighted least square problem given perfect identification knowledge of the NLOS/LOS measurements. In [32], Gaussian message passing-based algorithms were proposed for joint synchronization and localization in mixed LOS-NLOS environments with perfect path identification. It approximated the exponential NLOS bias

as Gaussian and linearized the non-linear range term so that the messages can be derived in closed-form for localization.

The assumption of perfect identification knowledge is unfortunately unrealistic especially in harsh environments, which has led to research to relax such assumption. In [29], the mixed LOS-NLOS localization problem was formulated as a maximum likelihood (ML) problem, which was approximately solved by two iterative algorithms based on expectation maximization (EM) and joint maximum a posterior and ML (JMAP-ML) criteria respectively with all measurements combined to localize the node using the quasi-Newton method. Moreover, in [26], root-mean-square (RMS) delay spread, mean excess delay and kurtosis were selected as the features to build the hypothesis and identify the LOS and NLOS measurements. The whole measurements were then used for localization using weighted least square (WLS) estimation. Furthermore, nonparametric methods such as support vector machine (SVM), Gaussian process (GP), relevance vector machine (RVM) have also been used to approximate the ranging errors in both LOS and NLOS measurements by selecting the similar features in [26] and then both types of measurements were used to estimate positions in [27, 28].

While much progress has been made to use both LOS and NLOS measurements for localization, one key assumption is that the anchors' locations are perfectly known. Uncertainties in anchors' locations result in intractable integrals of the likelihood function [13]. There have been efforts to address the issue of imperfect anchors' locations in LOS-only environments. In [14], the negative impact of anchor node uncertainty was analyzed using the Cramer-Rao lower bound (CRLB). Also, the anchor node uncertainty was considered with the measurements to formulate an ML problem and a semidefinite programming (SDP) method was proposed to solve it in [15]. In [17, 33], the uncertainties of anchor nodes were modelled as zero-mean Gaussian distributions and the estimation problem was solved by WLS and a Taylor expansion. In [34], the uncertainties of anchor nodes were taken as constraints in the optimization problem. In [35], the EM and Kullbak-Leibler divergence (KLD) were utilized to approximate the posterior distribution of uncertainty and estimate the target positions. Moreover, by taking the prior Gaussian distribution of node location uncertainties into consideration, a variational inference-based positioning algorithm was proposed based on maximum a posterior criterion in [36]. With respect to NLOS-only envoriments, anchor location uncertainty has also been considered [37]. However, similarly to LOS-only environments, no path identification is needed and the localization problem is relatively easier to solve than that in mixed LOS-NLOS environments.

TABLE I: Comparisons with the prior works

Works \ Assumptions	LOS-only	NLOS-only	Mixed LOS-NLOS	Anchor Node Location Uncertainty
[8], [9]	✓			
[14], [34]	✓			✓
[2], [12]		✓		
[37]		✓		✓
[21], [29]			✓	
Our Work			✓	✓

B. Contributions

In this paper, the uncertainties of the anchors' (or reference nodes') positions and the mixed LOS-NLOS measurements are both considered. This practical scenario widely occurs in harsh environments, e.g., urban areas with mobile anchor nodes. Moreover, the parameters of the LOS-NLOS propagation error model are assumed unknown. The estimation task aims to jointly estimate the error model parameters and the target node's position. The problem is extremely challenging due to the mixed measurements from LOS and NLOS paths, the nonlinear distance expression, and intractable integrals of the likelihood function with unknown error model parameters. To optimize the estimation performance, a variational Bayesian localization (VBL) algorithm using the variational Bayesian framework and importance sampling method is proposed. In summary, this paper has made the following contributions.

- We consider a challenging system model where the environment is composed of a mixture of NLOS and LOS paths and the anchors' positions are known with errors. This is significantly different from prior works as shown in **Table I**. Localization is also done while the error model parameters are unknown and jointly estimated.
- Using approximations to the true posterior distributions, we apply the variational framework to find the optimal variational distributions iteratively and propose a novel VBL algorithm[1].
- Also, we present complexity analysis of the proposed algorithm and the BCRB for the joint estimation problem.

[1]Although we also apply the variational framework for localization, our algorithm design is more challenging than that in [36] due to the existence of the mixed LOS/NLOS measurements.

C. Organization

The remainder of this paper is organized as follows. In Section II, the estimation problem is formulated as a maximum a posterior problem. Section III introduces the detailed derivations of variational distributions and the framework of VBL. In order to make specified interpretations of the proposed algorithm, Section VI gives the BCRB of the proposed algorithm. Finally, we present the simulation results in Section V and conclude the paper in Section VI.

II. Problem Formulation

Consider a wireless sensor network (WSN) with $M(\geq 3)$ anchors located at $\mathbf{x}_i$, for $i = 1, \ldots, M$, respectively. The anchors are either pre-deployed nodes or mobile nodes which join the WSN afterwards. When a new node joins into the WSN, it will estimate its location with the help of the existing anchors. Due to the presence of deployment errors and/or estimation errors, the locations of the anchors are imperfectly known with uncertainties. Denoting the coarse location of the ith anchor as $\bar{\mathbf{x}}_i$, its location uncertainty $\Delta\mathbf{x}_i = \mathbf{x}_i - \bar{\mathbf{x}}_i$ is usually assumed to be a zero-mean Gaussian random variable with covariance matrix of $\mathbf{\Sigma}_i$. In other words, the true location of the ith anchor follows Gaussian distribution as $\mathbf{x}_i \sim \mathcal{N}(\bar{\mathbf{x}}_i, \mathbf{\Sigma}_i)$. This Gaussian uncertainty model has been widely adopted in the literature and well justified by the result that the location estimation error can be well modeled as Gaussian random variable in [35, 36].

When a target node enters into the coverage area of the WSN, it can communicate with the anchor nodes and obtain range measurements based on time of arrival (TOA) [2] or received signal strength, etc.. Denoting the location of the target node as $\mathbf{x}$, the location $\mathbf{x}$ can be assumed to follow a Gaussian distribution $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{\Sigma})$ [36]. Meanwhile, the kth range measurement from the ith anchor $r_{i,k}$ can be written as

$$r_{i,k} = \underbrace{\|\mathbf{x} - \mathbf{x}_i\|_2}_{d_i} + \varepsilon_{i,k}, \ k = 1, \ldots, K, \tag{1}$$

where $\varepsilon_{i,k}$ denotes the ranging error. In WSNs, LOS path is not always possible between any two nodes. Thus, the range measurements may be obtained either from an LOS path or NLOS path, and the ranging error $\varepsilon_{i,k}$ can be modelled as a two-mode Gaussian random variable as [29]

$$\varepsilon_{i,k} \sim \begin{cases} \mathcal{N}(\varepsilon_{i,k}|\mu_1, \Lambda_1^{-1}) & \text{LOS error with probability of } \alpha_1, \\ \mathcal{N}(\varepsilon_{i,k}|\mu_2, \Lambda_2^{-1}) & \text{NLOS error with probability of } \alpha_2, \end{cases} \tag{2}$$

[2]For TOA-based ranging technique, time synchronization among the anchor and target nodes is essential and necessary. Thanks to the rich literature in synchronization [38, 39], it can be achieved by many well developed algorithms. In our paper, we thus assume the anchor and target nodes are perfectly synchronized.

with $\mu_2 \gg \mu_1$ incorporating extra traveling distance through the NLOS path and $\sum_{l=1}^{2} \alpha_l = 1$. To distinguish between the LOS and NLOS errors, a binary indicator vector $\mathbf{y}_{i,k} = [y_{i,k,1}, y_{i,k,2}]$ is introduced here. Specifically, the indicator vector is defined as

$$\mathbf{y}_{i,k} = \begin{cases} (1,0) & r_{i,k} \in \text{LOS}, \\ (0,1) & r_{i,k} \in \text{NLOS}. \end{cases} \tag{3}$$

Given the probability of $\boldsymbol{\alpha}$, the indicator vector $\mathbf{y} = \{\mathbf{y}_{1,1}, \mathbf{y}_{2,1}, \ldots, \mathbf{y}_{M,1}, \ldots, \mathbf{y}_{1,K}, \mathbf{y}_{2,K}, \ldots, \mathbf{y}_{M,K}\}$ then follows the distribution of

$$p(\mathbf{y}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} \prod_{k=1}^{K} \prod_{l=1}^{2} {\alpha_l}^{y_{i,k,l}}. \tag{4}$$

The probabilities α_l, $l = 1, 2$, can be regarded empirically as the ratios of the number of LOS/NLOS paths over the total number of paths, respectively. In practice, the range measurements could come from either LOS or NLOS paths and the ratios of the number of LOS/NLOS paths over the total number of paths vary from environment to environment and also from time to time. It is thus reasonable to model the probabilities/ratios as random and unknown with the constraints $\sum_{l=1}^{2} \alpha_l = 1$ and $0 \leq \alpha_l \leq 1$. Generally, to meet the physical constraints (i.e., $\sum_{l=1}^{2} \alpha_l = 1$ and $0 \leq \alpha_l \leq 1$) of the individual probabilities of NLOS and LOS measurements, the variable $\boldsymbol{\alpha}$ can be assumed to follow a Dirichlet distribution with order of 2 [40]. The prior distribution of $\boldsymbol{\alpha}$ is then given by

$$p(\boldsymbol{\alpha}) = Dir(2, \bar{\boldsymbol{\lambda}}), \tag{5}$$

where $\bar{\boldsymbol{\lambda}}$ is the parameter of Dirichelet distribution with elements $(\bar{\lambda}_1, \bar{\lambda}_2)$. This model is also aligned with the facts that the paths are indicated as LOS or NLOS paths by the associated variable $\mathbf{y}$ which follows a multinomial distribution and the Dirichlet distribution is a conjugate prior of a multinomial distribution. On the other hand, the statistical information of the LOS and NLOS ranging errors (i.e., μ_1, Λ_1, μ_2, and Λ_2) is also difficult to obtain in practice and is considered random and unknown. Since Wishart distribution is the conjugate prior of inverse covariance of a Gaussian distribution, it is widely adopted to model the distribution of the inverse covariance Λ_l of the ranging errors [40]. This Wishart model has been well justified by the experimental data in [36]. Meanwhile, as shown in [40], given the inverse covariance Λ_l, the mean μ_l of the ranging errors can be modeled using a Gaussian distribution. As a result, the prior distribution $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ can be written as

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{l=1}^{2} \mathcal{N}(\mu_l | \bar{m}_l, (\bar{\beta}_l \Lambda_l)^{-1}) \mathcal{W}(\Lambda_l | \bar{W}_l, \bar{v}_l), \tag{6}$$

where $\bar{\beta}_l$ is the correlation coefficient between μ_l and Λ_l, $\bar{m}_l$, $\bar{\beta}_l$ and Λ_l are the parameters of the Gaussian distribution $p(\mu_l|\Lambda_l) = \mathcal{N}(\mu_l|\bar{m}_l, (\bar{\beta}_l\Lambda_l)^{-1})$, $\bar{W}_l$ and $\bar{v}_l$ are the parameters of the Wishart distribution $p(\Lambda_l) = \mathcal{W}(\Lambda_l|\bar{W}_l, \bar{v}_l)$ with mean of $\mathbb{E}(\Lambda_l) = \bar{W}_l\bar{v}_l$.

After collecting the range measurements from all the anchor nodes as $\mathbf{r} = vec[r_{i,k}]$, the likelihood probability can be written as

$$p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{x}) = \prod_{i=1}^{M}\prod_{k=1}^{K}\prod_{l=1}^{2}\left[\mathcal{N}(r_{i,k} - d_i|\mu_l, {\Lambda_l}^{-1})\right]^{y_{i,k,l}}, \tag{7}$$

where $\boldsymbol{\theta}$ denotes the unknown nuisance parameters as $\boldsymbol{\theta} = [\mathbf{x}_1, \ldots, \mathbf{x}_M, \mu_1, \mu_2, \Lambda_1, \Lambda_2, \boldsymbol{\alpha}, \mathbf{y}]$. Meanwhile, the posterior probability $p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{r})$ follows

$$p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{r}) \propto p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{x})p(\boldsymbol{\theta}, \mathbf{x}). \tag{8}$$

Location estimation can be conducted based on ML or the maximum a posterior (Bayesian) criterion. Bayesian estimation incorporates the prior information of the unknown parameters and generally leads to better performance. In this paper, we thus adopt the Bayesian criterion for location estimation.

Based on the Bayesian rules, the formulated posterior distribution in (8) can be factorized as follows:

$$p\left(\boldsymbol{\theta}, \mathbf{x}|\mathbf{r}\right) \propto p\left(\mathbf{r}|\boldsymbol{\theta}, \mathbf{x}\right) p\left(\boldsymbol{\theta}, \mathbf{x}\right) = p\left(\mathbf{r}|\boldsymbol{\theta}, \mathbf{x}\right) p\left(\mathbf{y}|\boldsymbol{\alpha}\right) p\left(\boldsymbol{\alpha}\right) \prod_{l=1}^{2} p\left(\mu_l, \Lambda_l\right) \prod_{i=1}^{M} p\left(\mathbf{x}_i\right) p\left(\mathbf{x}\right). \tag{9}$$

In the considered WSN, mixed LOS/NLOS range measurements are collected with anchor location uncertainties and unknown ranging error statistics. Moreover, nonlinear distance term $d_i = \|\mathbf{x} - \mathbf{x}_i\|_2$ is involved. The Bayesian posterior probability in (9) is a very complicated expression and direct maximization is intractable. To solve this problem, a VBL method will be proposed to estimate the target location based on an approximated posterior distribution and infer the nuisance parameters.

III. VBL

Since the Bayesian posterior probability (9) is too complicated for direct maximization, we propose to find an approximation of the Bayesian posterior distribution based on mean-field variational inference. Essentially, it is to find a variational distribution $q\left(\boldsymbol{\theta}, \mathbf{x}\right)$ to approximate the Bayesian posterior distribution $p\left(\boldsymbol{\theta}, \mathbf{x}|\mathbf{r}\right)$ such that the KLD between them is minimized. As defined in [41], the KLD between two distributions has the non-negative property as

$$\mathrm{KL}\left(q\left(\boldsymbol{\theta}, \mathbf{x}\right)||p\left(\boldsymbol{\theta}, \mathbf{x}|\mathbf{r}\right)\right) \triangleq -\int_{\boldsymbol{\theta}, \mathbf{x}} q\left(\boldsymbol{\theta}, \mathbf{x}\right) \ln \frac{p\left(\boldsymbol{\theta}, \mathbf{x}|\mathbf{r}\right)}{q\left(\boldsymbol{\theta}, \mathbf{x}\right)} d\boldsymbol{\theta} d\mathbf{x} \geq 0, \tag{10}$$

where the equality holds when $q(\boldsymbol{\theta}, \mathbf{x}) = p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{r})$. The minimization of the KLD therefore can lead to a good approximation of the Bayesian posterior distribution.

Given the Bayesian posterior probability in (9) and based on the mean-field theory [42], the variational distribution can be set to follow the form as

$$q(\boldsymbol{\theta}, \mathbf{x}) = \underbrace{q(\mathbf{y})\, q(\boldsymbol{\alpha})\;\, q(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \prod_{i=1}^{M} q(\mathbf{x}_i)}_{q(\boldsymbol{\theta})}\, q(\mathbf{x}), \tag{11}$$

where the individual variational distributions can be regarded as approximations to the corresponding posterior distributions, e.g., $q(\mathbf{x})$ is the approximation to the posterior distribution $p(\mathbf{x}|\mathbf{r})$ and $q(\boldsymbol{\Lambda}, \boldsymbol{\mu})$ is the approximation to the posterior distribution $q(\boldsymbol{\Lambda}, \boldsymbol{\mu}|\mathbf{r})$, etc.. Considering the prior distributions in (5) and (6) and the conjugate prior principle, the variational distributions of $\boldsymbol{\alpha}$ and $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ can also be set respectively as [40]

$$q(\boldsymbol{\alpha}) = Dir(\boldsymbol{\alpha}|2, \boldsymbol{\lambda}), \tag{12}$$

$$q(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{l=1}^{2} \mathcal{N}(\mu_l | m_l, (\beta_l \Lambda_l)^{-1}) \mathcal{W}(\Lambda_l | W_l, v_l), \tag{13}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, and all the parameters including $\{\boldsymbol{\lambda}, m_l, \beta_l, W_l, v_l\}$ are to be determined.

With the expressions in (11)–(13), variational Bayesian inference method is now introduced to solve the KLD minimization problem and infer the locations and the nuisance parameters. Specifically, the KLD can be rewritten as

$$\mathrm{KL}\left(q(\boldsymbol{\theta}, \mathbf{x}) \,||\, p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{r})\right) = -\int_{\boldsymbol{\theta}, \mathbf{x}} q(\boldsymbol{\theta}, \mathbf{x}) \ln \frac{p(\boldsymbol{\theta}, \mathbf{x}|\mathbf{r})}{q(\boldsymbol{\theta}, \mathbf{x})} d\boldsymbol{\theta} d\mathbf{x} = \underbrace{-\int_{\boldsymbol{\theta}, \mathbf{x}} q(\boldsymbol{\theta}, \mathbf{x}) \ln \frac{p(\boldsymbol{\theta}, \mathbf{x}, \mathbf{r})}{q(\boldsymbol{\theta}, \mathbf{x})} d\boldsymbol{\theta} d\mathbf{x}}_{\mathbb{F}(\boldsymbol{\theta}, \mathbf{x})} + \ln p(\mathbf{r}). \tag{14}$$

Clearly, the divergence minimization is equivalent to the maximization of the term $\mathbb{F}(\boldsymbol{\theta}, \mathbf{x})$, which is called as evidence lower bound (ELBO). Putting (11) into (14), we have

$$\mathbb{F}(\boldsymbol{\theta}, \mathbf{x}) = \int_{\boldsymbol{\theta}, \mathbf{x}} q(\mathbf{y})\, q(\boldsymbol{\alpha})\;\, q(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \prod_{i=1}^{M} q(\mathbf{x}_i) q(\mathbf{x}) \ln \frac{p(\boldsymbol{\theta}, \mathbf{x}, \mathbf{r})}{q(\mathbf{y})\, q(\boldsymbol{\alpha})\;\, q(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \prod_{i=1}^{M} q(\mathbf{x}_i) q(\mathbf{x})} d\boldsymbol{\theta} d\mathbf{x}. \tag{15}$$

Since multiple distributions to be determined are involved in (15), alternating optimization is adopted to solve the maximization of $\mathbb{F}(\boldsymbol{\theta}, \mathbf{x})$ iteratively. In each iteration, each of the distributions $q(\mathbf{y})$, $q(\boldsymbol{\alpha})$, $q(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, $q(\mathbf{x}_i)$ and $q(\mathbf{x})$ is sequentially optimized by fixing the other distributions. More specifically, in the ηth iteration, given the distribution $q^{(\eta)}(\boldsymbol{\theta})$, the optimization of $q^{(\eta)}(\mathbf{x})$ is formulated as

$$\begin{aligned}\max_{q^{(\eta)}(\mathbf{x})} \mathbb{F}^{(\eta)}(\boldsymbol{\theta},\mathbf{x}) &= \int_{\mathbf{x}} q^{(\eta)}(\mathbf{x})\, q^{(\eta)}(\boldsymbol{\theta}) \int_{\boldsymbol{\theta}} \ln \frac{p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})}{q^{(\eta)}(\boldsymbol{\theta})\, q^{(\eta)}(\mathbf{x})} d\boldsymbol{\theta} d\mathbf{x} \\ &= \int_{\mathbf{x}} q^{(\eta)}(\mathbf{x}) \int_{\boldsymbol{\theta}} q^{(\eta)}(\boldsymbol{\theta}) \ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\, d\boldsymbol{\theta} d\mathbf{x} - \int_{\mathbf{x}} q^{(\eta)}(\mathbf{x}) \ln q^{(\eta)}(\mathbf{x})\, d\mathbf{x} + const \\ &= \underbrace{-\int_{\mathbf{x}} q^{(\eta)}(\mathbf{x}) \ln \frac{q^{(\eta)}(\mathbf{x})}{\exp\left(\int_{\boldsymbol{\theta}} q^{(\eta)}(\boldsymbol{\theta}) \ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\, d\boldsymbol{\theta}\right)} d\mathbf{x}}_{-KL\left(q^{(\eta)}(\mathbf{x}) \| \exp\left(\int_{\boldsymbol{\theta}} q^{(\eta)}(\boldsymbol{\theta}) \ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r}) d\boldsymbol{\theta}\right)\right)} + const.\end{aligned} \tag{16}$$

Clearly, the ELBO in (16) is maximized when the KLD term regarding to $q^{(\eta)}(\mathbf{x})$ reaches zero, namely,

$$q^{(\eta)}(\mathbf{x}) = \tfrac{1}{\mathcal{Z}_{\mathbf{x}}} \exp\left(\int_{\boldsymbol{\theta}} q^{(\eta)}(\boldsymbol{\theta}) \ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\, d\boldsymbol{\theta}\right) = \tfrac{1}{\mathcal{Z}_{\mathbf{x}}} \exp\left(\mathbb{E}_{q^{(\eta)}(\boldsymbol{\theta})}\left[\ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\right]\right), \tag{17}$$

where $\mathbb{E}_{q^{(\eta)}(\boldsymbol{\theta})}$ denotes the expectation with respect to $q^{(\eta)}(\boldsymbol{\theta})$ and $\mathcal{Z}_{\mathbf{x}}$ is the associated normalization scalar given as

$$\mathcal{Z}_{\mathbf{x}} = \int_{\mathbf{x}} \exp\left(\mathbb{E}_{q^{(\eta)}(\boldsymbol{\theta})}\left[\ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\right]\right) d\mathbf{x}. \tag{18}$$

Similarly, given $q^{(\eta)}(\mathbf{x})$ and $q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus \mathbf{y}}\right)$, the optimal distribution $q^{(\eta)}(\mathbf{y})$ that maximizes the ELBO can be derived as

$$q^{(\eta)}(\mathbf{y}) = \frac{1}{\mathcal{Z}_{\mathbf{y}}} \exp\left(\mathbb{E}_{q^{(\eta)}(\mathbf{x}) q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus \mathbf{y}}\right)}\left[\ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\right]\right), \tag{19}$$

where $\boldsymbol{\theta}_{\setminus \mathbf{y}}$ denotes the parameter vector of $\boldsymbol{\theta}$ excluding $\mathbf{y}$. Other individual distributions can also be optimized respectively as

$$q^{(\eta)}(\boldsymbol{\mu},\boldsymbol{\Lambda}) = \frac{1}{\mathcal{Z}_{\boldsymbol{\mu},\boldsymbol{\Lambda}}} \exp\left(\mathbb{E}_{q^{(\eta)}(\mathbf{x}) q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus\{\boldsymbol{\mu},\boldsymbol{\Lambda}\}}\right)}\left[\ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\right]\right), \tag{20}$$

$$q^{(\eta)}(\boldsymbol{\alpha}) = \frac{1}{\mathcal{Z}_{\boldsymbol{\alpha}}} \exp\left(\mathbb{E}_{q^{(\eta)}(\mathbf{x}) q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus \boldsymbol{\alpha}}\right)}\left[\ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\right]\right), \tag{21}$$

$$q^{(\eta)}(\mathbf{x}_i) = \frac{1}{\mathcal{Z}_{\mathbf{x}_i}} \exp\left(\mathbb{E}_{q^{(\eta)}(\mathbf{x}) q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus \mathbf{x}_i}\right)}\left[\ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\right]\right). \tag{22}$$

Using the mean-field factorization and alternating optimization, an analytical approximation to the complicated joint posterior distribution can be obtained by the variational Bayesian inference method. Moreover, the approximation is a product of a number of variational distributions and each of them is an approximation of the corresponding posterior distribution of one unknown parameter. With the approximated individual posterior distributions, the locations and the nuisance parameters can then be estimated. Since the derivations of the variational distributions in (17) and (19)–(22) are not straightforward, the details are given in the following.

A. Variational Distribution $q^{(\eta+1)}(\mathbf{y})$

By taking logarithm on both sides of (19) and given $q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus \mathbf{y}}\right)$ and $q^{(\eta)}(\mathbf{x})$, the variational distribution $q^{(\eta+1)}(\mathbf{y})$ can be written as

$$\ln q^{(\eta+1)}(\mathbf{y}) \propto \int\limits_{\boldsymbol{\theta}_{\setminus \mathbf{y}}, \mathbf{x}} q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus \mathbf{y}}\right) q^{(\eta)}(\mathbf{x}) \ln p(\boldsymbol{\theta}, \mathbf{x}, \mathbf{r}) \, d\boldsymbol{\theta} d\mathbf{x}. \tag{23}$$

Substituting the likelihood probability (7) and the prior distribution (4) into (19) and putting other terms irrelevant to $\mathbf{y}$ as constant, we can obtain

$$\begin{aligned} \ln q^{(\eta+1)}(\mathbf{y}) \propto & \sum_{i=1}^{M} \sum_{k=1}^{K} \sum_{l=1}^{2} y_{i,k,l} \mathbb{E}_{q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus \mathbf{y}}, \mathbf{x}\right)}\left(\ln \mathcal{N}\left(r_{i,k} - d_i | \mu_l, {\Lambda_l}^{-1}\right)\right) + \mathbb{E}_{q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus y}, \mathbf{x}\right)}\left[\ln p(\boldsymbol{\theta}, \mathbf{x})\right] \\ = & \sum_{i=1}^{M} \sum_{k=1}^{K} \sum_{l=1}^{2} y_{i,k,l} \ln \Upsilon_{i,k,l}^{(\eta)} + const, \end{aligned} \tag{24}$$

where $\ln \Upsilon_{i,k,l}^{(\eta)}$ is given by

$$\ln \Upsilon_{i,k,l}^{(\eta)} = \mathbb{E}_{q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus \mathbf{y}}, \mathbf{x}\right)}\left(\ln \mathcal{N}\left(r_{i,k} - d_i | \mu_l, {\Lambda_l}^{-1}\right)\right) + \mathbb{E}_{q^{(\eta)}(\boldsymbol{\alpha})}\left[\ln \alpha_l\right]. \tag{25}$$

Putting (12) and (13) into (25) and taking expectations with respect to the variational distributions, it follows

$$\begin{aligned} \ln \Upsilon_{i,k,l}^{(\eta)} = & \frac{1}{2}\left(\ln 2 + \psi\left(\frac{1}{2} v_l^{(\eta)}\right) + \ln\left|W_l^{(\eta)}\right|\right) - \frac{1}{2} \ln 2\pi \\ & - \frac{1}{2}\left[[\beta_l^{(\eta)}]^{(-1)} + v_l^{(\eta)} W_l^{(\eta)} \mathbb{E}_{q^{(\eta)}(\mathbf{x}, \mathbf{x}_i)}\left(r_{i,k} - d_i - m_l^{(\eta)}\right)^2\right] + \psi\left(\lambda_l^{(\eta)}\right) - \psi\left(\sum_{l=1}^{2} \lambda_l^{(\eta)}\right), \end{aligned} \tag{26}$$

where $\psi(\cdot)$ is the digamma function defined as the logarithmic derivative of Gamma function.

Based on (24), the variational distribution $q^{(\eta+1)}(\mathbf{y})$ can then be reformulated as

$$q^{(\eta+1)}(\mathbf{y}) \propto \prod_{i=1}^{M} \prod_{k=1}^{K} \prod_{l=1}^{2} e^{y_{i,k,l} \ln \Upsilon_{i,k,l}^{(\eta)}} = \prod_{i=1}^{M} \prod_{k=1}^{K} \prod_{l=1}^{2} \left(\Upsilon_{i,k,l}^{(\eta)}\right)^{y_{i,k,l}}. \tag{27}$$

Clearly, $\Upsilon_{i,k,l}^{(\eta)}$ can roughly be regarded as an approximation of the posterior probability of the path indicator, i.e., $p(y_{i,k,l} | r_{i,k})$ [40]. In the calculation of $\Upsilon_{i,k,l}^{(\eta)}$ in (26), the expectation with respect to $q^{(\eta)}(\mathbf{x}, \mathbf{x}_i)$ is involved. Due to the presence of the nonlinear distance term $d_i = \|\mathbf{x} - \mathbf{x}_i\|_2$, it is very difficult to obtain the expectation in closed form. We will adopt the technique of importance sampling to solve it, as introduced later in Section III-E.

B. Variational Distribution $q^{(\eta+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$

The variational distribution $q^{(\eta+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ in (20) can be rewritten as

$$q^{(\eta+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \propto \exp\left(\mathbb{E}_{q^{(\eta)}(\boldsymbol{\theta}_{\setminus\{\boldsymbol{\mu},\boldsymbol{\Lambda}\}})q^{(\eta)}(\mathbf{x})}\left[\ln p(\boldsymbol{\theta}, \mathbf{x}, \mathbf{r})\right]\right). \tag{28}$$

By plugging the likelihood probability (7) and the prior distribution (6) into (28), it follows

$$\begin{aligned} q^{(\eta+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) &\propto \exp\left(\mathbb{E}_{q^{(\eta)}(\boldsymbol{\theta}_{\setminus\{\boldsymbol{\mu},\boldsymbol{\Lambda}\}})q^{(\eta)}(\mathbf{x})}\left[\ln p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{x})\right] + \ln p(\boldsymbol{\mu}, \boldsymbol{\Lambda})\right) \\ &\propto \exp\left(\sum_{j=1}^{M}\sum_{k=1}^{K}\sum_{l=1}^{2}\mathbb{E}_{q^{(\eta)}(\boldsymbol{\theta}_{\setminus\{\boldsymbol{\mu},\boldsymbol{\Lambda}\}})q^{(\eta)}(\mathbf{x})}\left[y_{i,k,l}\ln\mathcal{N}\left(r_{i,k} - d_i|\mu_l, \Lambda_l^{-1}\right)\right]\right) \times \exp\left(\sum_{l=1}^{2}\ln p(\mu_l, \Lambda_l)\right). \end{aligned} \tag{29}$$

With (27), the expectation of $y_{i,k,l}$ with respect to $q^{(\eta)}(\mathbf{y})$ is given by

$$\mathbb{E}_{q^{(\eta)}(\mathbf{y})}\left[y_{i,k,l}\right] = \Gamma_{i,k,l}^{(\eta)} = \frac{\Upsilon_{i,k,l}^{(\eta)}}{\sum_{l=1}^{2}\Upsilon_{i,k,l}^{(\eta)}}. \tag{30}$$

Defining

$$\gamma_l^{(\eta)} = \sum_{i=1}^{M}\sum_{k=1}^{K}\Gamma_{i,k,l}^{(\eta)}, \tag{31}$$

and after some tedious derivation as shown in Appendix A, $q^{(\eta+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ can be written as

$$q^{(\eta+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \propto \prod_{l=1}^{2}\mathcal{N}\left(\mu_l|m_l^{(\eta+1)}, \left(\beta_l^{(\eta+1)}\Lambda_l\right)^{-1}\right)\mathcal{W}\left(\Lambda_l|W_l^{(\eta+1)}, v_l^{(\eta+1)}\right), \tag{32}$$

where the parameters are, respectively, given by

$$m_l^{(\eta+1)} = \frac{\bar{\beta}_l\bar{m}_l + \gamma_l^{(\eta)}\chi_l^{(\eta)}}{\bar{\beta}_l + \gamma_l^{(\eta)}}, \tag{33}$$

$$v_l^{(\eta+1)} = \bar{v}_l + \gamma_l^{(\eta)}, \tag{34}$$

$$\left(W_l^{(\eta+1)}\right)^{-1} = \left(\bar{W}_l\right)^{-1} + \gamma_l^{(\eta)}\left(\varsigma_l^{(\eta)} + \Delta_l^{(\eta)}\right) + \frac{\bar{\beta}_l\gamma_l^{(\eta)}}{\bar{\beta}_l + \gamma_l^{(\eta)}}\left(\chi_l^{(\eta)} - \bar{m}_l\right)^2. \tag{35}$$

and $\beta_l^{(\eta+1)}$ is given in (67).

As aforementioned, $q^{(\eta+1)}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ is an approximation of the posterior distribution $q(\boldsymbol{\Lambda}, \boldsymbol{\mu}|\mathbf{r})$. With (32), the estimates of μ_l and Λ_l based on the minimum mean-square-error (MMSE) criterion can be obtained, respectively, as [36]

$$\mu_l^{(\eta+1)} = m_l^{(\eta+1)}, \tag{36}$$

$$\Lambda_l^{(\eta+1)} = W_l^{(\eta+1)}v_l^{(\eta+1)}. \tag{37}$$

C. Variational Distribution $q^{(\eta+1)}(\boldsymbol{\alpha})$

By substituting (4)–(7) into (21) and ignoring the terms irrelevant to $\boldsymbol{\alpha}$, the variational distribution $q^{(\eta+1)}(\boldsymbol{\alpha})$ can be written as

$$q^{(\eta+1)}(\boldsymbol{\alpha}) \propto \exp\left(\mathbb{E}_{q^{(\eta)}\left(\boldsymbol{\theta}_{\setminus\boldsymbol{\alpha}}\right)q^{(\eta)}(\mathbf{x})}\left[\ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\right]\right) \propto \exp\left(\mathbb{E}_{q^{(\eta)}(\mathbf{y})}\left[\ln p(\mathbf{y}|\boldsymbol{\alpha})\right] + \ln p(\boldsymbol{\alpha})\right). \quad (38)$$

By taking variational expectation and using (5), (30) and (31), it yields

$$\begin{aligned} q^{(\eta+1)}(\boldsymbol{\alpha}) \propto & \exp\left(\sum_{i=1}^{M}\sum_{k=1}^{K}\sum_{l=1}^{2}\mathbb{E}_{q^{(\eta)}(\mathbf{y})}\left[y_{i,k,l}\ln\alpha_l\right] - \ln\mathcal{U}(\bar{\lambda}) + \sum_{l=1}^{2}\left(\bar{\lambda}-1\right)\ln\alpha_l\right) \\ = & \exp\left(\sum_{l=1}^{2}\gamma_l^{(\eta)}\ln\alpha_l - \ln\mathcal{U}(\bar{\lambda}) + \sum_{l=1}^{2}\left(\bar{\lambda}-1\right)\ln\alpha_l\right) \\ = & \frac{1}{\mathcal{U}\left(\bar{\boldsymbol{\lambda}}\right)}\prod_{l=1}^{2}\alpha_l^{\left(\bar{\lambda}_l-1+\gamma_l^{(\eta)}\right)}, \end{aligned} \quad (39)$$

where $\mathcal{U}(\cdot)$ is the Beta function defined as $\mathcal{U}\left(\bar{\boldsymbol{\lambda}}\right) = \frac{\vartheta\left(\sum_{l=1}^{2}\bar{\lambda}_l\right)}{\vartheta\left(\bar{\lambda}_1\right)\vartheta\left(\bar{\lambda}_2\right)}$.

Similarly, with conjugate prior principle, the variational distribution $q^{(\eta+1)}(\boldsymbol{\alpha})$ follows the Dirichlet distribution in (12). We then have $\lambda_l^{(\eta+1)} = \bar{\lambda}_l + \gamma_l^{(\eta)}$ by mapping (39) into the form (12). The MMSE estimate of α_l can also be obtained as

$$\alpha_l^{(\eta+1)} = \frac{\lambda_l^{(\eta+1)}}{\sum_{l=1}^{l=2}\lambda_l^{(\eta+1)}}, \quad \text{for } l = 1,2. \quad (40)$$

D. Variational Distributions $q^{(\eta+1)}(\mathbf{x})$ and $q^{(\eta+1)}(\mathbf{x}_i)$

Substituting (7) into (17) leads to the variational distribution $q^{(\eta+1)}(\mathbf{x})$ as

$$\begin{aligned} \ln q^{(\eta+1)}(\mathbf{x}) \propto & \int_{\boldsymbol{\theta}} q^{(\eta)}(\boldsymbol{\theta})\ln p(\boldsymbol{\theta},\mathbf{x},\mathbf{r})\,d\boldsymbol{\theta} \\ = & \mathbb{E}_{q^{(\eta)}(\boldsymbol{\theta})}\left[\sum_{i=1}^{M}\sum_{k=1}^{K}\sum_{l=1}^{2}y_{i,k,l}\ln\mathcal{N}\left(r_{i,k}-d_i|\mu_l,\Lambda_l^{-1}\right)\right] + \ln p(\mathbf{x}) + const. \end{aligned} \quad (41)$$

By taking exponential operations on both sides and based on (30), we have

$$\begin{aligned} q^{(\eta+1)}(\mathbf{x}) \propto & \, p(\mathbf{x})\exp\left(\sum_{i=1}^{M}\sum_{k=1}^{K}\sum_{l=1}^{2}\mathbb{E}_{q^{(\eta)}(\boldsymbol{\theta})}\left[y_{i,k,l}\ln\mathcal{N}\left(r_{i,k}-d_i|\mu_l,\Lambda_l^{-1}\right)\right]\right) \\ = & \, p(\mathbf{x})\prod_{i=1}^{M}\prod_{k=1}^{K}\prod_{l=1}^{2}\exp\left(\Gamma_{i,k,l}^{(\eta)}\ln\widetilde{\Upsilon}_{i,k,l}^{(\eta)}\right), \end{aligned} \quad (42)$$

where $\ln \widetilde{\Upsilon}_{i,k,l}^{(\eta)}$ is given by

$$\begin{aligned}\ln \widetilde{\Upsilon}_{i,k,l}^{(\eta)} &= \mathbb{E}_{q^{(\eta)}(\boldsymbol{\theta})}\left[\ln \mathcal{N}\left(r_{i,k}-d_i|\mu_l,\Lambda_l^{-1}\right)\right] \\ &= \frac{1}{2}\left(\ln 2+\psi\left(v_l^{(\eta)}/2\right)+\ln\left|W_l^{(\eta)}\right|\right)-\frac{1}{2}\ln 2\pi \\ &\quad -\frac{1}{2}\left[[\beta_l^{(\eta)}]^{(-1)}+v_l^{(\eta)}W_l^{(\eta)}\mathbb{E}_{q^{(\eta)}(\mathbf{x}_i)}\left[\left(r_{i,k}-d_i-m_l^{(\eta)}\right)^2\right]\right].\end{aligned} \tag{43}$$

It then directly follows

$$q^{(\eta+1)}(\mathbf{x}) \propto p(\mathbf{x})\prod_{i=1}^{M}\prod_{k=1}^{K}\prod_{l=1}^{2}\exp\left(\Gamma_{i,k,l}^{(\eta)}\ln\widetilde{\Upsilon}_{i,k,l}^{(\eta)}\right)=p(\mathbf{x})\prod_{i=1}^{M}\prod_{k=1}^{K}\prod_{l=1}^{2}\left(\widetilde{\Upsilon}_{i,k,l}^{(\eta)}\right)^{\Gamma_{i,k,l}^{(\eta)}}. \tag{44}$$

Here $\widetilde{\Upsilon}_{i,k,l}^{(\eta)}$ can be roughly regarded as an approximation of the likelihood probability $p\left(r_{i,k}|y_{i,k,l}=1,\mathbf{x}\right)$.

Similarly, the variational distribution $q^{(\eta+1)}(\mathbf{x}_i)$ in (22) can be derived as

$$q^{(\eta+1)}(\mathbf{x}_i) \propto p(\mathbf{x}_i)\prod_{k=1}^{K}\prod_{l=1}^{2}\exp\left(\Gamma_{i,k,l}^{(\eta)}\ln\bar{\Upsilon}_{i,k,l}^{(\eta)}\right), \tag{45}$$

with $\bar{\Upsilon}_{i,k,l}^{(\eta)}$ given by

$$\begin{aligned}\ln \bar{\Upsilon}_{i,k,l}^{(\eta)} &= \frac{1}{2}\left(\ln 2+\psi\left(v_l^{(\eta)}/2\right)+\ln\left|W_l^{(\eta)}\right|\right)-\frac{1}{2}\ln 2\pi \\ &\quad -\frac{1}{2}\left[[\beta_l^{(\eta)}]^{(-1)}+v_l^{(\eta)}W_l^{(\eta)}\mathbb{E}_{q^{(\eta)}(\mathbf{x})}\left[\left(r_{i,k}-d_i-m_l^{(\eta)}\right)^2\right]\right].\end{aligned} \tag{46}$$

E. Importance Sampling-Based Expectation Calculation

It is clearly seen in (26), (43) and (46) that the expectations with respect to $q^{(\eta)}(\mathbf{x},\mathbf{x}_i)$, $q^{(\eta)}(\mathbf{x}_i)$ and $q^{(\eta)}(\mathbf{x})$ respectively are necessary to be calculated in the iterative algorithm. Due to the presence of the nonlinear distance term $d_i = \|\mathbf{x}-\mathbf{x}_i\|_2$ and the complicated expressions of the variational distributions $q^{(\eta)}(\mathbf{x}_i)$ and $q^{(\eta)}(\mathbf{x})$, direct calculation of these expectations is very difficult. Here we resort to the importance sampling technique [43] for the expectation derivations. According to importance sampling theory [44] and with (44) and (45), the variational distributions $q^{(\eta)}(\mathbf{x})$ and $q^{(\eta)}(\mathbf{x}_i)$ can be approximated respectively based on variational particle sets $\left\{\mathbf{x}_\tau,\omega_\tau^{(\eta)}|\forall\tau=1:N_p\right\}$ and $\left\{\mathbf{x}_{i,\iota},\omega_{i,\iota}^{(\eta)}|\forall\iota=1:N_p\right\}$ as

$$q^{(\eta)}(\mathbf{x}) \approx \sum_{\tau=1:N_p}\omega_\tau^{(\eta)}\delta(\mathbf{x}-\mathbf{x}_\tau), \tag{47}$$

$$q^{(\eta)}(\mathbf{x}_i) \approx \sum_{\iota=1:N_p}\omega_{i,\iota}^{(\eta)}\delta(\mathbf{x}_i-\mathbf{x}_{i,\iota}), \tag{48}$$

where N_p is the number of particles, $\delta(\cdot)$ is the Dirac delta function, $\mathbf{x}_\tau$ and $\mathbf{x}_{i,\iota}$ are the support points of particles drawn from the priori distributions $p(\mathbf{x})$ and $p(\mathbf{x}_i)$, respectively, i.e.,

$$\mathbf{x}_\tau \sim p(\mathbf{x}), \quad \mathbf{x}_{i,\iota} \sim p(\mathbf{x}_i), \tag{49}$$

and $\omega_\tau^{(\eta)}$ and $\omega_{i,\iota}^{(\eta)}$ denote the corresponding weights, respectively, given by

$$\omega_\tau^{(\eta)} \propto \prod_{i=1}^{M}\prod_{k=1}^{K}\prod_{l=1}^{2} e^{\Gamma_{i,k,l}^{(\eta-1)} \ln \widetilde{\Upsilon}_{i,k,l,\tau}^{(\eta-1)}}, \tag{50}$$

$$\omega_{i,\iota}^{(\eta)} \propto \prod_{k=1}^{K}\prod_{l=1}^{2} e^{\Gamma_{i,k,l}^{(\eta-1)} \ln \bar{\Upsilon}_{i,k,l,\iota}^{(\eta-1)}}. \tag{51}$$

In (50) and (51), $\ln \widetilde{\Upsilon}_{i,k,l,\tau}^{(\eta)}$ and $\ln \bar{\Upsilon}_{i,k,l,\iota}^{(\eta)}$ are defined based on (43) and (46) for given particles $\mathbf{x}_\tau$ and $\mathbf{x}_{i,\iota}$, respectively, namely,

$$\begin{aligned}\ln \widetilde{\Upsilon}_{i,k,l,\tau}^{(\eta)} =& \frac{1}{2}\left(\ln 2 + \psi\left(v_l^{(\eta)}/2\right) + \ln\left|W_l^{(\eta)}\right|\right) - \frac{1}{2}\ln 2\pi \\ & - \frac{1}{2}\left[[\beta_l^{(\eta)}]^{(-1)} + v_l^{(\eta)} W_l^{(\eta)} \sum_{\iota=1:N_p} \omega_{i,\iota}^{(\eta)} \left[\left(r_{i,k} - \left\|\mathbf{x}_\tau - \mathbf{x}_{i,\iota}\right\|_2 - m_l^{(\eta)}\right)^2\right]\right],\end{aligned} \tag{52}$$

$$\begin{aligned}\ln \bar{\Upsilon}_{i,k,l,\iota}^{(\eta)} =& \frac{1}{2}\left(\ln 2 + \psi\left(v_l^{(\eta)}/2\right) + \ln\left|W_l^{(\eta)}\right|\right) - \frac{1}{2}\ln 2\pi \\ & - \frac{1}{2}\left[[\beta_l^{(\eta)}]^{(-1)} + v_l^{(\eta)} W_l^{(\eta)} \sum_{\tau=1:N_p} \omega_{\tau}^{(\eta)} \left[\left(r_{i,k} - \left\|\mathbf{x}_\tau - \mathbf{x}_{i,\iota}\right\|_2 - m_l^{(\eta)}\right)^2\right]\right].\end{aligned} \tag{53}$$

With the particle-based approximate distributions in (47) and (48), the expectation with respect to $q^{(\eta)}(\mathbf{x}, \mathbf{x}_i)$ in (26) can be calculated and $\ln \Upsilon_{i,k,l}^{(\eta)}$ can be derived as

$$\begin{aligned}\ln \Upsilon_{i,k,l}^{(\eta)} \approx & \frac{1}{2}\left(\ln 2 + \psi\left(v_l^{(\eta)}/2\right) + \ln\left|W_l^{(\eta)}\right|\right) - \frac{1}{2}\ln 2\pi \\ & - \frac{1}{2}\left[[\beta_l^{(\eta)}]^{(-1)} + v_l^{(\eta)} W_l^{(\eta)} \sum_{\iota=1:N_p}\sum_{\tau=1:N_p} \omega_{i,\iota}^{(\eta)}\omega_{\tau}^{(\eta)} \left[\left(r_{i,k} - \left\|\mathbf{x}_\tau - \mathbf{x}_{i,\iota}\right\|_2 - m_l^{(\eta)}\right)^2\right]\right].\end{aligned} \tag{54}$$

Moreover, the MMSE estimates of the locations $\mathbf{x}$ and $\mathbf{x}_i$ directly follow from (47) and (48), respectively, as

$$\mathbf{x}^{(\eta)} = \sum_{\tau=1:N_p} \omega_\tau^{(\eta)} \mathbf{x}_\tau, \tag{55}$$

$$\mathbf{x}_i^{(\eta)} = \sum_{\iota=1:N_p} \omega_{i,\iota}^{(\eta)} \mathbf{x}_{i,\iota}. \tag{56}$$

F. Summary and Analysis

Our VBL algorithm is an iterative algorithm developed based on mean-field factorization, KLD minimization and alternating optimization. Specifically, in each iteration, each of the variational distributions $q(\mathbf{y})$, $q(\boldsymbol{\alpha})$, $q(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, $q(\mathbf{x}_i)$ and $q(\mathbf{x})$ is sequentially optimized by fixing the other distributions. As shown in (16)–(22), each step in every iteration leads to a higher ELBO, thus, a lower KLD. Therefore, the KLD is reduced in each iteration, and as a consequence, the proposed VBL algorithm is guaranteed to converge.

Based on the above derivations, the VBL algorithm is summarized as **Algorithm 1**. Clearly, the complexity is dominated by the updates of $\ln \Upsilon_{i,k,l}^{(\eta)}$ in (54) and the weightings $\omega_{\tau}^{(\eta)}$ and $\omega_{i,\iota}^{(\eta)}$ in (50) and (51). Each of them involves a complexity linear to KMN_p^2. Therefore, the overall complexity of the VBL algorithm can be given as $\mathcal{O}(\mathrm{KM}{N_p}^2\eta)$, where η is the number of iterations in the algorithm. Notice that the factor of N_p^2 in the complexity is introduced due to the implementation of the importance sampling technique. Although the complexity is higher than that of the JMAP-ML and EM algorithms in [29], i.e., $\mathcal{O}(\mathrm{KM}\eta)$, and that of the particle swarm optimization (PSO) in [45, 46], i.e., $\mathcal{O}(2\mathrm{KM}N_p\eta)$, our algorithm is robust against location uncertainties of anchor nodes, which will be demonstrated later in the simulation section.

As shown in **Algorithm 1**, the proposed VBL algorithm needs the statistical information of the anchor location uncertainty, target location, and Dirichlet and Wishart distributions, as the input. The prior information of the anchor location uncertainty can be obtained from empirical data or through analysis of the location estimator, i.e., the inverse Fisher information matrix of the Bayesian estimator. Meanwhile, the prior information of target node can be obtained from a coarse estimator, e.g., a simple triangulation localization algorithm. With respect to the statistical parameters in the prior distributions, e.g., Dirichlet and Wishart distributions, they can be obtained from measurements and empirical data. Since statistical information usually can be estimated with high accuracy, the initialization is not a key issue in our proposed algorithm.

In this paper, only one target node is considered. When multiple target nodes appear, our proposed algorithm is still applicable. Similarly to [4, 11], cooperative localization can be conducted by utilizing the information from other target nodes. Specifically, the target nodes only need to exchange their coarse locations and its covariance matrices with their neighbors. Then each node can collect range measurements and estimate its location in parallel using our proposed VBL algorithm.

Algorithm 1 VBL Algorithm

Input the prior distributions $p(\mathbf{x})$, $p(\mathbf{x}_i)$ and (5)-(6) with parameters $\bar{\lambda}_l$,$\bar{m}_l$,$\bar{v}_l$,$\bar{W}_l$,$\bar{\beta}_l$;
Collect the range measurements $\mathbf{r}$;
Generate initial particle sets according to (49)-(51);
Set $\eta = 1$;
while ELBO does not converge or η less than maximum iteration number **do**
    Update $\ln \Upsilon_{i,k,l}^{(\eta)}$ according to (54) and the variational distribution $q^{(\eta)}(\mathbf{y})$ in (27);
    Update $m_l^{(\eta)}$, $v_l^{(\eta)}$, $W_l^{(\eta)}$ and $\beta_l^{(\eta)}$ according to (33)-(35) and (67) respectively, and the variational distribution $q^{(\eta)}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ in (32);
    Update $\lambda_l^{(\eta)}$ and the variational distribution $q^{(\eta)}(\boldsymbol{\alpha})$ in (39);
    Update the weighting $\omega_{i,\iota}^{(\eta)}$ according to (51) and (53) and the variational distribution $q^{(\eta)}(\mathbf{x}_i)$ in (48);
    Update the weighting $\omega_{\tau}^{(\eta)}$ according to (50) and (52) and the variational distribution $q^{(\eta)}(\mathbf{x})$ in (47);
    $\eta = \eta + 1$;
end while
Output the location estimate $\mathbf{x}^{(\eta)}$ according to (55) as well as the estimates of the nuisance parameters $\mathbf{x}_i^{(\eta)}$, $\alpha_l^{(\eta)}$, $\Lambda_l^{(\eta)}$ and $\mu_l^{(\eta)}$ according to (56), (40), (37) and (36), respectively.

IV. BCRB

It is well known that BCRB is an effective bound for Bayesian estimators [47]. It is adopted here as a benchmark for evaluating the performance of our VBL algorithm. Considering the complete variable vector $\boldsymbol{\xi} = \{\mathbf{x}, \boldsymbol{\theta}\}^T$, the BCRB is derived based on the Bayesian Fisher information matrix $J(\boldsymbol{\xi})$ defined as follows [48]

$$J(\boldsymbol{\xi}) = \underbrace{-\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p(\mathbf{r}|\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T}\right]}_{\mathcal{I}_{\mathrm{M}}} \underbrace{-\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p(\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T}\right]}_{\mathcal{I}_{\mathrm{P}}}, \tag{57}$$

where $\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}[\cdot]$ and $\mathbb{E}_{\boldsymbol{\xi}}[\cdot]$ denote the expectations with respect to the joint distribution $p(\mathbf{r}, \boldsymbol{\xi})$ and prior distribution $p(\boldsymbol{\xi})$, respectively.

With the definition $\boldsymbol{\xi}=\{\mathbf{x},\boldsymbol{\theta}\}^T$, the Bayesian Fisher information matrix can be partitioned as

$$J\left(\boldsymbol{\xi}\right)=\left[\begin{array}{cc}\underbrace{-\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2\ln p\left(\mathbf{r}|\boldsymbol{\xi}\right)}{\partial\mathbf{x}\partial\mathbf{x}^T}\right]-\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2\ln p\left(\boldsymbol{\xi}\right)}{\partial\mathbf{x}\partial\mathbf{x}^T}\right]}_{\boldsymbol{\Delta}} & \underbrace{-\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2\ln p\left(\mathbf{r}|\boldsymbol{\xi}\right)}{\partial\mathbf{x}\partial\boldsymbol{\theta}^T}\right]-\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2\ln p\left(\boldsymbol{\xi}\right)}{\partial\mathbf{x}\partial\boldsymbol{\theta}^T}\right]}_{\boldsymbol{\Psi}^T}\\ \underbrace{-\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2\ln p\left(\mathbf{r}|\boldsymbol{\xi}\right)}{\partial\boldsymbol{\theta}\partial\mathbf{x}^T}\right]-\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2\ln p\left(\boldsymbol{\xi}\right)}{\partial\boldsymbol{\theta}\partial\mathbf{x}^T}\right]}_{\boldsymbol{\Psi}} & \underbrace{-\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2\ln p\left(\mathbf{r}|\boldsymbol{\xi}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]-\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2\ln p\left(\boldsymbol{\xi}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]}_{\boldsymbol{\Omega}}\end{array}\right]. \quad (58)$$

Through some tedious derivations as shown in Appendices B–D, the submatrices $\boldsymbol{\Delta}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$ can be derived as

$$\boldsymbol{\Delta}=\sum_{i=1}^{M}\sum_{k=1}^{K}\sum_{l=1}^{2}\frac{\bar{\alpha}_l\bar{v}_l\bar{W}_l}{d_i^2}\left[\bar{\mathbf{x}}\bar{\mathbf{x}}^T+\boldsymbol{\Sigma}-\bar{\mathbf{x}}\bar{\mathbf{x}}_i^T-\bar{\mathbf{x}}_i\bar{\mathbf{x}}^T+\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T+\boldsymbol{\Sigma}_i\right]+\boldsymbol{\Sigma}^{-1}, \quad (59)$$

$$\boldsymbol{\Psi}=\left[\begin{array}{cc}\underbrace{\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{x}_1},\ldots,\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{x}_M},\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\boldsymbol{\mu}}}_{\boldsymbol{\Psi}_1} & \mathbf{0}\end{array}\right]^T. \quad (60)$$

$$\boldsymbol{\Omega}=\left[\begin{array}{cc}\boldsymbol{\Omega}_1 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Omega}_2\end{array}\right]. \quad (61)$$

where $\bar{\alpha}_l=\frac{\bar{\lambda}_l}{\sum_{l=1}^{2}\bar{\lambda}_l}$,

$$\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{x}_i}=-\sum_{k=1}^{K}\sum_{l=1}^{2}\frac{\bar{\alpha}_l\bar{v}_l\bar{W}_l}{d_i^2}\left[\bar{\mathbf{x}}\bar{\mathbf{x}}^T+\boldsymbol{\Sigma}-\bar{\mathbf{x}}\bar{\mathbf{x}}_i^T-\bar{\mathbf{x}}_i\bar{\mathbf{x}}^T+\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T+\boldsymbol{\Sigma}_i\right], \quad (62)$$

$$\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mu_l}=\sum_{i=1}^{M}\sum_{k=1}^{K}\frac{\bar{\alpha}_l\bar{v}_l\bar{W}_l}{d_i}(\bar{\mathbf{x}}-\bar{\mathbf{x}}_i), \quad (63)$$

and $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are defined in (83) with elements given in (84)–(88).

Based on the Bayesian Fisher information matrix $J\left(\boldsymbol{\xi}\right)$, the BCRB corresponding to the estimate of the target node location is given as $\mathrm{BCRB}_{\mathbf{x}}=\left[J(\boldsymbol{\xi})^{-1}\right]_{1,1}+\left[J(\boldsymbol{\xi})^{-1}\right]_{2,2}$ in [48]. With (59)–(61) and Using Schur's complement [1, 49], it follows

$$\mathrm{BCRB}_{\mathbf{x}}=\left[J(\boldsymbol{\xi})^{-1}\right]_{1,1}+\left[J(\boldsymbol{\xi})^{-1}\right]_{2,2}=tr[(\boldsymbol{\Delta}-\boldsymbol{\Psi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi})^{-1}]=tr[(\boldsymbol{\Delta}-\boldsymbol{\Psi}_1\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Psi}_1^T)^{-1}]. \quad (64)$$

Intuitively, the term $\boldsymbol{\Psi}_1\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Psi}_1^T$ quantifies the information loss from the imperfectly known parameters. This BCRB will be used as a lower bound for the mean-square-error (MSE) of the target location estimate in our paper.

Notice that our BCRB analysis is more challenging than that in [50, 51]. Specifically, the CRLBs in [50, 51] were derived for the localization problems in LOS-only environments with anchor location uncertainty. However, our BCRB is derived for the localization problem in mixed LOS-NLOS environments with anchor location uncertainty. Moreover, the LOS/NLOS ratios

are unknown and assumed following certain prior distribution, e.g., Dirichlet distribution. From mathematical point of view, our BCRB derivation can be extended from that in [50, 51] by introducing additional expectation calculation with respect to the prior distributions. But the extension is by no means straightforward.

V. Simulations and Results

A. Simulation Setup

In this section, we show the convergence property and localization performance of the proposed VBL algorithm in different scenarios. We consider the 2D localization of a target and five anchor nodes are both deployed in an area $[-150, 150] \times [-150, 200]$ meter (m) in Fig. 1. The hyper parameters are respectively given as $\bar{m}_1 = 0\ m$, $\bar{m}_2 = 120\ m$, $\delta_1 = (\bar{v}_1\bar{W}_1)^{-1/2} = 20\ m$, $\delta_2 = (\bar{v}_2\bar{W}_2)^{-1/2} = 60\ m$ and $\bar{\beta}_1 = \bar{\beta}_2 = 0.1$. The maximum iteration number is set to be 30. The number of particles is $N_p = 20$. By setting $\bar{\lambda}_1 = 60$ and $\bar{\lambda}_2 = 40$, the expectation of NLOS ratio is set to be $\mathbb{E}(\alpha_2) = 0.4$. The number of measurements is set to be $K = 10$. The covariance matrix of Gaussian distribution of the target node is set to be $\mathbf{\Sigma} = 100\mathbf{I}$. The uncertainties of anchor nodes are assumed to be identical to each other and are given by $\mathbf{\Sigma}_i = u\mathbf{I}$. The mentioned settings keep unaltered otherwise stated differently. The proposed algorithm VBL is compared to the following algorithms:

- The EM and joint maximum a posterior and ML proposed for localization problem in the mixed LOS/NLOS environment in [29].
- The PSO optimization algorithm in [45, 46] and the cost function in [29]. We consider the perfect situation with error model parameters $\boldsymbol{\mu}$ and $\mathbf{\Lambda}$ and identification information are all perfectly known. We also consider the imperfect situation PSO with only 5% identification error, which means $0.05MK$ measurements are misidentified.

B. Convergence Properties

We aim at the numerical analysis of the convergence properties of the proposed VBL method in different scenarios. In Fig. 2, the number of measurements is set to be $K = 10$ and the number of particles is set to be $N_p = 20$. The convergence property is validated in the following cases:

- The NLOS ratio $\mathbb{E}(\alpha_2) = 0.4$ and $u = 0$ correspond to the case with the mixed NLOS/LOS measurements and accurate anchor node positions.

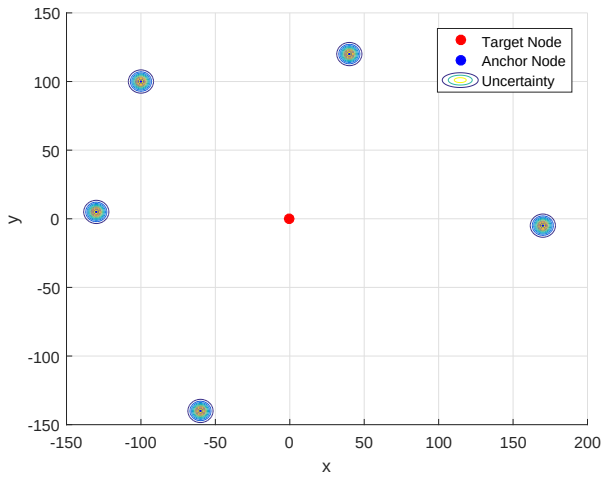

Fig. 1: Geometry of the target node and the anchor nodes with uncertainties.

- The NLOS ratio $\mathbb{E}(\alpha_2)$ =0.4 and $u = 25$ correspond to the scenario with the mixed N-LOS/LOS measurements and anchor node uncertainties of $\mathbf{\Sigma}_i = 25\mathbf{I}$.

From the results in Fig. 2, the root MSE (RMSE) of the proposed VBL algorithm decreases monotonically over iterations and this means the variational distribution $q(\mathbf{x})$ approaches the true posterior distribution gradually, which follows our theoretical derivations. The simulation results show that the fast convergence property of the VBL algorithm and the robust localization against the uncertainties of anchor nodes. The RMSE difference in two cases is mainly resulted form the different level of anchor node uncertainty. Moreover, comparing the results in Fig. 2, the gaps between the RMSEs and corresponding BCRBs show the significant negative impact of uncertainties of anchor nodes.

C. Performance Comparisons

In the first experiment, we show the localization performance of the proposed algorithm and algorithms JMAP-ML and EM in mixed NLOS/LOS environments with different anchor node uncertainties. The NLOS ratio expectation is set to be $\mathbb{E}(\alpha_2) = 0.4$. The uncertainty parameter u is set to be from 0 to 200 and the variance of LOS error model δ_1 is set to be 20.

In Fig. 3, the increase of anchor node uncertainties will degrade the localization accuracy and the VBL algorithm outperforms other comparison algorithms. For EM and JMAP-ML algorithms, the 'hard decision' in JMAP-ML and 'soft decision' in EM do not incorporate the anchor node uncertainties and it brings more errors in identification of NLOS/LOS measurements. For the

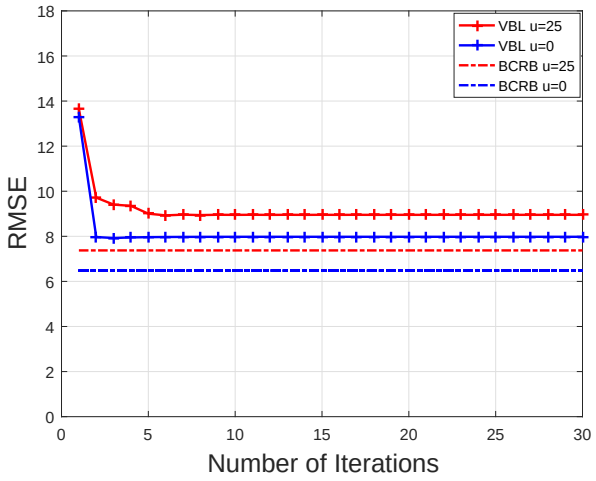

Fig. 2: The RMSE versus number of iterations in different cases with $\mathbb{E}(\alpha_2) = 0.4, K = 10$ and $N_p = 20$.

imperfect PSO and perfect PSO, the minor identification errors (5%) can result in a significant increase in localization errors. For the proposed VBL algorithm, the anchor node positions can be refined by using MMSE estimation, which means that the uncertainty will decrease and brings benefits to estimate the target. Meanwhile, the increasing RMSEs of JMAP-ML, EM and PSO over the increase of uncertainty demonstrate the high negative impact of anchor node uncertainty, especially in large uncertainty cases such as bad preceding estimation of anchor nodes in urban areas or in the network of unmaned aerial vehicles (UAVs). Essentially, our proposed VBL algorithm exploits the approximations to posterior distributions to estimate the target node location, which can achieve better performance than the approximations to likelihood functions in the EM, JMAP-ML and PSO algorithms.

In the second experiment, we investigate the localization performance of the mentioned algorithms in different level of NLOS measurement ratios $\mathbb{E}(\alpha_2)$ with anchor node uncertainties. In this experiment, we set $\bar{\lambda}_1$ to be fixed at $\bar{\lambda}_1 = 60$ and $\bar{\lambda}_2$ ranges in $[15, 240]$, which makes the NLOS ratio $\mathbb{E}(\alpha_2)$ ranges in $[0.2, 0.8]$ and covers the slight and severe NLOS measurement contaminations. The anchor node uncertainty is set to be $u = 25$.

In the results in Fig. 4, the 'soft decision' mechanism of EM algorithm can only provide the conditional probability of NLOS/LOS identifications not the exact classification of NLOS/LOS measurements. Hence, the increase of NLOS ratio will inevitably enlarge the identification errors and the localization performance will degrade. For the JMAP-ML algorithm, the 'hard decision'

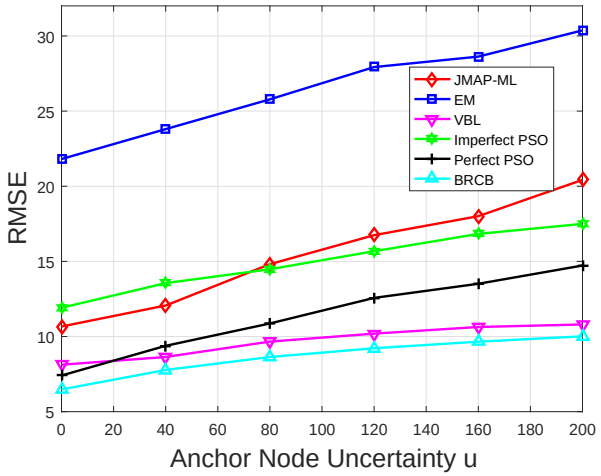

Fig. 3: RMSE versus different anchor node uncertainties with $\mathbb{E}(\alpha_2) = 0.4, K = 10$ and $N_p = 20$.

mechanism can effectively estimate the NLOS/LOS indicators in a binary way. The localization accuracy of the proposed VBL algorithm slightly degrades with the increase of NLOS ratio and can achieve better performance than the other algorithms, which is resulted from the advantage of the statistical information of NLOS ratio and it is better than the perfect PSO.

From the results in Figs. 3-4, we can conclude that our proposed VBL algorithm is more robust than the other algorithms when the environments become harsher, e.g., larger anchor location uncertainty and/or higher NLOS measurement contaminations, etc.. The robustness mainly comes from the adoption of maximum a posterior criterion to fully exploit the prior information of the anchor locations, range errors and LOS/NLOS ratios for localization. Notice that the JMAP-ML and EM algorithms follow maximum likelihood criterion for localization and have stronger assumptions on the localization environments, e.g., no anchor location uncertainty.

D. Parameter Impacts

In this subsection, we study the impacts of level of variances δ_l, for $l = 1, 2$, the number of measurements K and the number of particles N_p. In the third experiment, we change the value of $\bar{W}_1$ and fix the value $\bar{v}_1 = 30$ so that δ_1 varies in the range $[20, 60]$. The NLOS path ratio is set to be $\mathbb{E}(\alpha_2) = 0.4$. The anchor node uncertainty is set to be $u = 25$. The number of measurements is set to be $K = 10$ and the number of particles is set to be $N_p = 20$. In the fourth experiment, we change the value of $\bar{W}_2$ and fix the value $\bar{v}_2 = 40$ and δ_2 varies in the range $[20, 60]$.

Combining the results of two experiments in Figs. 5 and 6, the proposed VBL algorithm

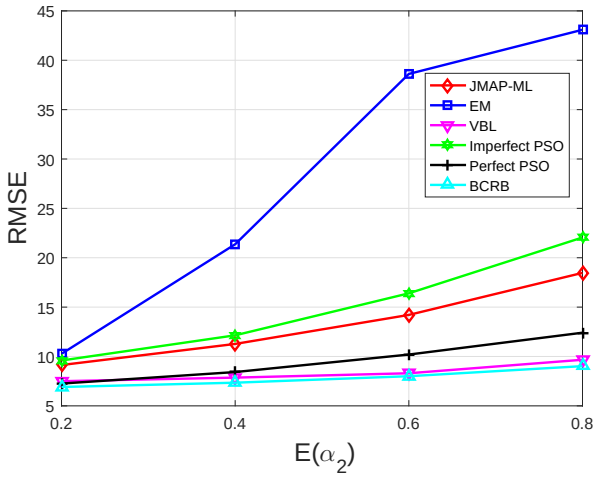

Fig. 4: The RMSE versus different NLOS ratios with $K = 10, N_p = 20$ and $u = 25$.

has better performance than the other algorithms. In Fig. 6, the BCRB is scaled by the term $\left(K\sum_{l=1}^{2}\bar{\alpha}_l\bar{v}_l\bar{W}_l\right)^{-1} = \left(K\left(\frac{\bar{\alpha}_1}{\delta_1^2}+\frac{\bar{\alpha}_2}{\delta_2^2}\right)\right)^{-1}$ in (64) and the scale parameter is approximately dominated by $\left(K\left(\frac{\bar{\alpha}_1}{\delta_1^2}\right)\right)^{-1}$ when $\delta_2 \gg \delta_1$. Thus, the BCRB is increasing more and more slowly over the NLOS variance δ_2 in Fig. 6. For the results of EM and JMAP-ML in Figs. 5 and 6, the increase of NLOS and LOS variances have negative impacts over the identification of the mixed measurements and will degrade the localization accuracy.

In the fifth experiment, the number of measurements K is set to range in $[5, 50]$. The level of NLOS measurement is set to be $\mathbb{E}(\alpha_2) = 0.4$. The anchor node uncertainty is set to be $u = 25$. The number of particles is set to be $N_p = 20$. As shown in Fig. 7, the RMSE is decreasing over the increase of number of measurements K. As we mentioned above, the EM and JMAP-ML algorithms have different decision mechanisms, and JMAP-ML can achieve superior identification that that of EM. While the VBL employs the posterior distribution of indicator variable $y_{i,k,l}$ to identify the NLOS/LOS mixed measurements. Thus, VBL can achieve more accurate identification results and localization accuracy. Meanwhile, the RMSE difference between perfect and imperfect PSO algorithms demonstrates the vulnerability of PSO in identification errors.

In the sixth experiment, the number of particles N_p is set to range in $[5, 30]$. The NLOS paths are set to be $\mathbb{E}(\alpha_2) = 0.4$. The other parameters keep unaltered. In Fig. 8, the RMSE decreases with the increases of the number of particles, which follows the theoretical derivations. More particles can capture and approximate the statistical characteristics of the uncertainties for both the target node and anchor nodes. The choice of the number of particles depends on the tradeoff

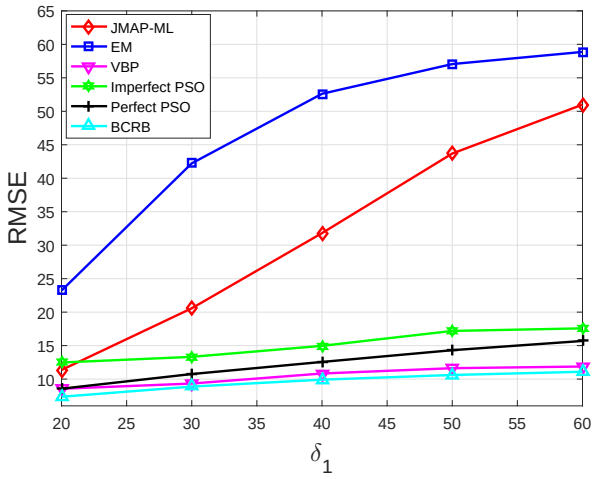

Fig. 5: The RMSE versus different LOS variance δ_1 with $\mathbb{E}(\alpha_2) = 0.4, K = 10$ and $u = 25$.

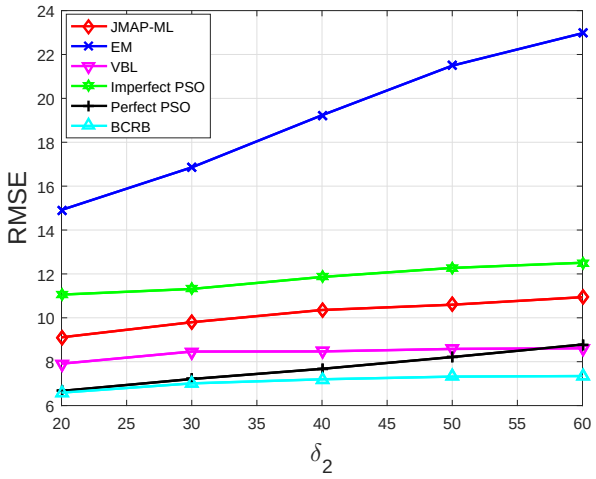

Fig. 6: The RMSE versus different NLOS variance δ_2 with $\mathbb{E}(\alpha_2) = 0.4, K = 10$ and $u = 25$.

between the computational complexity and the performance accuracy.

VI. Conclusion

In this paper, we considered a challenging localization problem with node uncertainties and unknown parameters of error model in mixed NLOS/LOS environment. To this end, we proposed a VBL algorithm for robust localization. In the proposed algorithm, the underlying measurement errors are modeled as a mixture Gaussian distribution and the statistical information of uncertainties has also been considered. Due to the intractable direct position estimation of likelihood function associated with the target node, the posterior distributions were approximated by variational

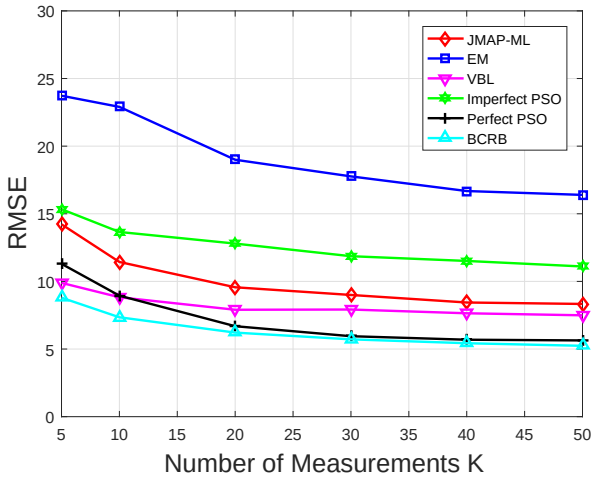

Fig. 7: The RMSE versus different number of measurements with $\mathbb{E}(\alpha_2) = 0.4, N_p = 20$ and $u = 25$.

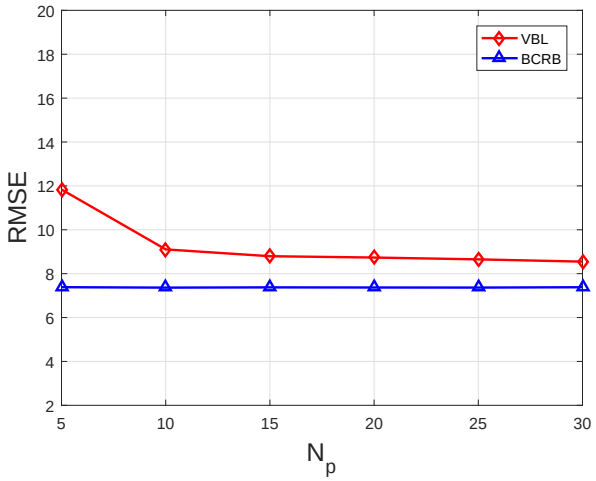

Fig. 8: The RMSE versus different number of particles with $\mathbb{E}(\alpha_2) = 0.4, K = 10$ and $u = 25$.

distributions through minimizing the KLD. Moreover, importance sampling was employed to handle the intractable term of nonlinearity and uncertainties of nodes. As an iterative algorithm, the convergence and complexity of the proposed VBL algorithm were investigated. Simulation results have shown the superiority of VBL in convergence, estimation and localization accuracy through various examples.

APPENDIX A

A. Derivation of (32)

Expanding the exponential terms in (29) and using the result (30), we can obtain

$$\begin{aligned} q^{(\eta+1)}\left(\boldsymbol{\mu},\boldsymbol{\Lambda}\right) \propto & \exp\left(\sum_{i=1}^{M}\sum_{k=1}^{K}\sum_{l=1}^{2}\Gamma_{i,k,l}^{(\eta)}\mathbb{E}_{q^{(\eta)}(\mathbf{x},\mathbf{x}_i)}\left(\frac{1}{2}\ln\left|\Lambda_l\right|-\frac{1}{2}\ln 2\pi-\frac{1}{2}\left(\left(r_{i,k}-d_i-\mu_l\right)^2\Lambda_l\right)\right)\right)\times \\ & \exp\left\{\sum_{l=1}^{2}\left[\left(-\frac{1}{2}\ln 2\pi+\frac{1}{2}\ln\left|\bar{\beta}_l\Lambda_l\right|-\frac{1}{2}(\mu_l-\bar{m}_l)^2\left(\bar{\beta}_l\Lambda_l\right)\right)\right]\right\}\times \\ & \exp\left\{\sum_{l=1}^{2}\ln\mathcal{B}\left(\bar{W}_l,\bar{v}_l\right)+\frac{\bar{v}_l-2}{2}\ln\left|\Lambda_l\right|-\frac{1}{2}\left(\bar{W}_l\right)^{-1}\Lambda_l\right\}, \end{aligned} \tag{65}$$

where $\mathcal{B}(\cdot)$ is defined as $\mathcal{B}\left(\bar{W}_l,\bar{v}_l\right)=\left|\bar{W}_l\right|^{\frac{\bar{v}_l}{2}}\left(2^{\frac{\bar{v}_l}{2}}\vartheta\left(\frac{\bar{v}_l}{2}\right)\right)^{-1}$ with $\vartheta(\cdot)$ as Gamma function.

Combining the terms with μ_l and Λ_l respectively and with the definition in (31), the variational distribution in (65) can be reformulated as

$$\begin{aligned} q^{(\eta+1)}\left(\boldsymbol{\mu},\boldsymbol{\Lambda}\right) \propto & \exp\left(\sum_{l=1}^{2}\left(\frac{\bar{v}_l-2+\gamma_l^{(\eta)}}{2}\ln\left|\Lambda_l\right|+\ln\mathcal{B}\left(\bar{W}_l,\bar{v}_l\right)\right)\right)\times \\ & \exp\left(\sum_{l=1}^{2}\left(-\frac{1}{2}tr\left(\left(\left(\bar{W}_l\right)^{-1}+\gamma_l^{(\eta)}\left(\varsigma_l^{(\eta)}+\Delta_l^{(\eta)}\right)+\frac{\bar{\beta}_l\gamma_l^{(\eta)}}{\bar{\beta}_l+\gamma_l^{(\eta)}}\left(\chi_l^{(\eta)}-\bar{m}_l\right)^2\right)\Lambda_l\right)\right)\right)\times \\ & \exp\left(\sum_{l=1}^{2}\left(-\frac{1}{2}\left(\mu_l-\frac{\left(\bar{\beta}_l\bar{m}_l+\gamma_l^{(\eta)}\chi_l^{(\eta)}\right)}{\bar{\beta}_l+\gamma_l^{(\eta)}}\right)^2\beta_l^{(\eta+1)}\Lambda_l-\frac{1}{2}\ln 2\pi-\frac{\gamma_l^{(\eta)}}{2}\ln 2\pi+\frac{1}{2}\ln\left|\beta_l^{(\eta+1)}\Lambda_l\right|\right)\right), \end{aligned} \tag{66}$$

where$\beta_l^{(\eta+1)}$, $\chi_l^{(\eta)}$, $\varsigma_l^{(\eta)}$ and $\Delta_l^{(\eta)}$ are, respectively, given by

$$\beta_l^{(\eta+1)}=\bar{\beta}_l+\gamma_l^{(\eta)}, \tag{67}$$

$$\chi_l^{(\eta)}=\frac{1}{\gamma_l^{(\eta)}}\sum_{i=1}^{M}\sum_{k=1}^{K}\Gamma_{i,k,l}^{(\eta)}\mathbb{E}_{q^{(\eta)}(\mathbf{x},\mathbf{x}_i)}\left(r_{i,k}-d_i\right), \tag{68}$$

$$\varsigma_l^{(\eta)}=\frac{1}{\gamma_l^{(\eta)}}\sum_{i=1}^{M}\sum_{k=1}^{K}\Gamma_{i,k,l}^{(\eta)}\left[\mathbb{E}_{q^{(\eta)}(\mathbf{x},\mathbf{x}_i)}[r_{i,k}-d_i]-\chi_l^{(\eta)}\right]^2, \tag{69}$$

$$\Delta_l^{(\eta)}=\frac{1}{\gamma_l^{(\eta)}}\sum_{i=1}^{M}\sum_{k=1}^{K}\Gamma_{i,k,l}^{(\eta)}\Xi_{\mathbf{x},\mathbf{x}_i}^{(\eta)}, \tag{70}$$

with $\Xi_{\mathbf{x},\mathbf{x}_i}^{(\eta)}$ denoting the variance of the ranging error $(r_{i,k}-d_i)$ with respect to $q^{(\eta)}(\mathbf{x},\mathbf{x}_i)$ as

$$\Xi_{\mathbf{x},\mathbf{x}_i}^{(\eta)}=\mathbb{E}_{q^{(\eta)}(\mathbf{x},\mathbf{x}_i)}\left[r_{i,k}-d_i-\mathbb{E}_{q^{(\eta)}(\mathbf{x},\mathbf{x}_i)}\left[r_{i,k}-d_i\right]\right]^2. \tag{71}$$

Denoting the product of the first two terms in (66) as $f\left(\Lambda_l, W_l^{(\eta+1)}, v_l^{(\eta+1)}\right)$, and with the definitions of $v_l^{(\eta+1)}$ and $W_l^{(\eta+1)}$ in (34) and (35) respectively, we have

$$\begin{aligned} & f\left(\Lambda_l, W_l^{(\eta+1)}, v_l^{(\eta+1)}\right) \\ & = \exp\left(\sum_{l=1}^{2}\left(\frac{v_l^{(\eta+1)}-2}{2}\ln|\Lambda_l| + \ln\mathcal{B}\left(\bar{W}_l, \bar{v}_l\right)\right)\right)\exp\left(\sum_{l=1}^{2}\left(-\frac{1}{2}tr\left(W_l^{(\eta+1)}\Lambda_l\right)\right)\right) \\ & \propto \prod_{l=1}^{2}\mathcal{W}\left(\Lambda_l | W_l^{(\eta+1)}, v_l^{(\eta+1)}\right). \end{aligned} \tag{72}$$

Meanwhile, the third term in (66) can be rewritten as

$$\begin{aligned} & \exp\left(\sum_{l=1}^{2}\left(-\frac{1}{2}\left(\mu_l - \underbrace{\frac{\left(\bar{\beta}_l\bar{m}_l + \gamma_l^{(\eta)}\chi_l^{(\eta)}\right)}{\bar{\beta}_l + \gamma_l^{(\eta)}}}_{\triangleq m_l^{(\eta+1)}}\right)^2 \beta_l^{(\eta+1)}\Lambda_l - \frac{1}{2}\ln 2\pi - \frac{\gamma_l^{(\eta)}}{2}\ln 2\pi + \frac{1}{2}\ln\left|\beta_l^{(\eta+1)}\Lambda_l\right|\right)\right) \\ & \propto \prod_{l=1}^{2}\mathcal{N}\left(\mu_l | m_l^{(\eta+1)}, \left(\beta_l^{(\eta+1)}\Lambda_l\right)^{-1}\right). \end{aligned} \tag{73}$$

Putting (72) and (73) into (66), we can get (32).

B. Submatrix $\boldsymbol{\Delta}$

As shown in (58), the submatrix $\boldsymbol{\Delta}$ is defined as

$$\boldsymbol{\Delta} = \underbrace{-\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p\left(\mathbf{r}|\boldsymbol{\xi}\right)}{\partial\mathbf{x}\partial\mathbf{x}^T}\right]}_{\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{x}}} \underbrace{-\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p\left(\boldsymbol{\xi}\right)}{\partial\mathbf{x}\partial\mathbf{x}^T}\right]}_{\mathcal{I}_{\mathrm{P}}^{\mathbf{x},\mathbf{x}}}. \tag{74}$$

Based on the prior and likelihood distributions, the term $\mathcal{I}_M^{\mathbf{x},\mathbf{x}}$ can be written and derived as

$$\mathcal{I}_{\mathbf{M}}^{\mathbf{x},\mathbf{x}} = -\sum_{i=1}^{M}\sum_{k=1}^{K}\mathbb{E}_{\boldsymbol{\xi},\mathbf{r}}\left[\frac{\partial^2 \ln p\left(r_{i,k}|\boldsymbol{\xi}\right)}{\partial\mathbf{x}\partial\mathbf{x}^T}\right] = \sum_{i=1}^{M}\sum_{k=1}^{K}\sum_{l=1}^{2}\bar{\alpha}_l\bar{v}_l\bar{W}_l\mathbb{E}_{\mathbf{x},\mathbf{x}_i}\left[\mathbf{D}\mathbf{D}^T\right], \tag{75}$$

where $\bar{\alpha}_l = \frac{\bar{\lambda}_l}{\sum_{l=1}^{2}\bar{\lambda}_l}$, $\mathbf{D} = \frac{\mathbf{x}-\mathbf{x}_i}{d_i}$, and $\mathbb{E}_{\mathbf{x},\mathbf{x}_i}\left[\mathbf{D}\mathbf{D}^T\right] = \frac{1}{d_i^2}\left[\bar{\mathbf{x}}\bar{\mathbf{x}}^T + \boldsymbol{\Sigma} - \bar{\mathbf{x}}\bar{\mathbf{x}}_i^T - \bar{\mathbf{x}}_i\bar{\mathbf{x}}^T + \bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T + \boldsymbol{\Sigma}_i\right]$.

On the other hand, the term $\mathcal{I}_{\mathrm{P}}^{\mathbf{x},\mathbf{x}}$ can be derived as

$$\mathcal{I}_{\mathrm{P}}^{\mathbf{x},\mathbf{x}} = -\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p(\boldsymbol{\xi})}{\partial\mathbf{x}\partial\mathbf{x}^T}\right] = \boldsymbol{\Sigma}^{-1}. \tag{76}$$

The submatrix $\boldsymbol{\Delta}$ then follows from (74), (75) and (76) as

$$\boldsymbol{\Delta} = \sum_{i=1}^{M}\sum_{k=1}^{K}\sum_{l=1}^{2}\frac{\bar{\alpha}_l\bar{v}_l\bar{W}_l}{d_i^2}\left[\bar{\mathbf{x}}\bar{\mathbf{x}}^T + \boldsymbol{\Sigma} - \bar{\mathbf{x}}\bar{\mathbf{x}}_i^T - \bar{\mathbf{x}}_i\bar{\mathbf{x}}^T + \bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T + \boldsymbol{\Sigma}_i\right] + \boldsymbol{\Sigma}^{-1}. \tag{77}$$

C. Submatrix $\mathbf{\Psi}$

The submatrix $\mathbf{\Psi}$ is defined as shown in (58) as

$$\mathbf{\Psi} = -\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p\left(\mathbf{r}|\boldsymbol{\xi}\right)}{\partial \boldsymbol{\theta} \partial \mathbf{x}^T}\right] - \mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p\left(\boldsymbol{\xi}\right)}{\partial \boldsymbol{\theta} \partial \mathbf{x}^T}\right]. \tag{78}$$

With the independence of the elements in $\boldsymbol{\theta}$ with the target location $\mathbf{x}$, the second term $\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p(\boldsymbol{\xi})}{\partial \boldsymbol{\theta} \partial \mathbf{x}^T}\right]$ directly follows as $\mathbf{0}$. With respect to the first term, it can be written as

$$-\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p\left(\mathbf{r}|\boldsymbol{\xi}\right)}{\partial \boldsymbol{\theta} \partial \mathbf{x}^T}\right] = \left[\underbrace{\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{x}_1}, ..., \mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{x}_{\mathrm{M}}}, \mathcal{I}_{\mathrm{M}}^{\mathbf{x},\boldsymbol{\mu}}}_{\mathbf{\Psi}_1}, \mathcal{I}_{\mathrm{M}}^{\mathbf{x},\boldsymbol{\alpha}}, \mathcal{I}_{\mathrm{M}}^{\mathbf{x},\boldsymbol{\Lambda}}, \mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{y}}\right]^T \tag{79}$$

with the element given as $\mathcal{I}_{\mathrm{M}}^{\mathbf{a},\mathbf{b}} = -\mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p(\mathbf{r}|\boldsymbol{\xi})}{\partial \mathbf{a} \partial \mathbf{b}^T}\right]$.

Following the similar derivation in (75), we have

$$\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{x}_i} = -\sum_{k=1}^{K}\sum_{l=1}^{2} \bar{\alpha}_l \bar{v}_l \bar{W}_l \mathbb{E}_{\mathbf{x},\mathbf{x}_i}\left[\mathbf{D}\mathbf{D}^T\right], \tag{80}$$

$$\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mu_l} = -\sum_{i=1}^{M}\sum_{k=1}^{K} \mathbb{E}_{\mathbf{r},\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p\left(r_{i,k}|\boldsymbol{\xi}\right)}{\partial \mathbf{x} \partial \mu_l}\right] = \sum_{i=1}^{M}\sum_{k=1}^{K} \bar{\alpha}_l \bar{v}_l \bar{W}_l \mathbb{E}_{\mathbf{x},\mathbf{x}_i}\left[\mathbf{D}\right], \tag{81}$$

$$\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\boldsymbol{\alpha}} = \mathbf{0}, \quad \mathcal{I}_{\mathrm{M}}^{\mathbf{x},\boldsymbol{\Lambda}} = \mathbf{0} \quad \text{and} \quad \mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{y}} = \mathbf{0}, \tag{82}$$

where $\mathbb{E}_{\mathbf{x},\mathbf{x}_i}\left[\mathbf{D}\right] = \frac{1}{d_i}(\bar{\mathbf{x}} - \bar{\mathbf{x}}_i)$. Then it follows that $\mathbf{\Psi} = \left[\underbrace{\mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{x}_1}, ..., \mathcal{I}_{\mathrm{M}}^{\mathbf{x},\mathbf{x}_{\mathrm{M}}}, \mathcal{I}_{\mathrm{M}}^{\mathbf{x},\boldsymbol{\mu}}}_{\mathbf{\Psi}_1} \ \mathbf{0}\right]^T$.

D. Submatrix $\mathbf{\Omega}$

According to the definition in (58), the submatrix $\mathbf{\Omega}$ is formulated as

$$\begin{aligned}
\mathbf{\Omega} &= \left[\begin{array}{ccc|ccc}
\mathcal{I}_{\mathrm{M}}^{\mathbf{x}_1,\mathbf{x}_1} + \mathcal{I}_{\mathrm{P}}^{\mathbf{x}_1,\mathbf{x}_1} & \cdots & \mathcal{I}_{\mathrm{M}}^{\mathbf{x}_1,\mathbf{x}_M} & \mathcal{I}_{\mathrm{M}}^{\mathbf{x}_1,\boldsymbol{\mu}} & \mathcal{I}_{\mathrm{M}}^{\mathbf{x}_1,\boldsymbol{\alpha}} & \mathcal{I}_{\mathrm{M}}^{\mathbf{x}_1,\boldsymbol{\Lambda}} & \mathcal{I}_{\mathrm{M}}^{\mathbf{x}_1,\mathbf{y}} \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathcal{I}_{\mathrm{M}}^{\mathbf{x}_M,\mathbf{x}_1} & \cdots & \mathcal{I}_{M}^{\mathbf{x}_M,\mathbf{x}_M} + \mathcal{I}_{\mathrm{P}}^{\mathbf{x}_M,\mathbf{x}_M} & \mathcal{I}_{\mathrm{M}}^{\mathbf{x}_M,\boldsymbol{\mu}} & \mathcal{I}_{\mathrm{M}}^{\mathbf{x}_M,\boldsymbol{\alpha}} & \mathcal{I}_{\mathrm{M}}^{\mathbf{x}_M,\boldsymbol{\Lambda}} & \mathcal{I}_{\mathrm{M}}^{\mathbf{x}_M,\mathbf{y}} \\
\mathcal{I}_{\mathrm{M}}^{\boldsymbol{\mu},\mathbf{x}_1} & \cdots & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\mu},\mathbf{x}_M} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\mu},\boldsymbol{\mu}} + \mathcal{I}_{\mathrm{P}}^{\boldsymbol{\mu},\boldsymbol{\mu}} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\mu},\boldsymbol{\alpha}} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\mu},\boldsymbol{\Lambda}} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\mu},\mathbf{y}} \\
\hline
\mathcal{I}_{\mathrm{M}}^{\boldsymbol{\alpha},\mathbf{x}_1} & \cdots & \mathcal{I}_{M}^{\boldsymbol{\alpha},\mathbf{x}_M} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\alpha},\boldsymbol{\mu}} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\alpha},\boldsymbol{\alpha}} + \mathcal{I}_{\mathrm{P}}^{\boldsymbol{\alpha},\boldsymbol{\alpha}} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\alpha},\boldsymbol{\Lambda}} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\alpha},\mathbf{y}} \\
\mathcal{I}_{\mathrm{M}}^{\boldsymbol{\Lambda},\mathbf{x}_1} & \cdots & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\Lambda},\mathbf{x}_M} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\Lambda},\boldsymbol{\mu}} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\Lambda},\boldsymbol{\alpha}} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\Lambda},\boldsymbol{\Lambda}} + \mathcal{I}_{\mathrm{P}}^{\boldsymbol{\Lambda},\boldsymbol{\Lambda}} & \mathcal{I}_{\mathrm{M}}^{\boldsymbol{\Lambda},\mathbf{y}} \\
\mathcal{I}_{\mathrm{M}}^{\mathbf{y},\mathbf{x}_1} & \cdots & \mathcal{I}_{\mathrm{M}}^{\mathbf{y},\mathbf{x}_M} & \mathcal{I}_{\mathrm{M}}^{\mathbf{y},\boldsymbol{\mu}} & \mathcal{I}_{\mathrm{M}}^{\mathbf{y},\boldsymbol{\alpha}} & \mathcal{I}_{M}^{\mathbf{y},\boldsymbol{\Lambda}} & \mathcal{I}_{\mathrm{M}}^{\mathbf{y},\mathbf{y}} + \mathcal{I}_{\mathrm{P}}^{\mathbf{y},\mathbf{y}}
\end{array}\right] \\
&= \left[\begin{array}{c|c} \mathbf{\Omega}_1 & \mathbf{\Omega}_3^T \\ \hline \mathbf{\Omega}_3 & \mathbf{\Omega}_2 \end{array}\right],
\end{aligned} \tag{83}$$

where $\mathcal{I}_{\mathrm{P}}^{\mathbf{a},\mathbf{b}} = -\mathbb{E}_{\boldsymbol{\xi}}\left[\frac{\partial^2 \ln p(\boldsymbol{\xi})}{\partial \mathbf{a} \partial \mathbf{b}^T}\right]$.

Similarly, the elements in $\mathbf{\Omega}_1$ can be derived respectively as

$$\mathcal{I}_{\mathrm{M}}^{\mathbf{x}_i,\mathbf{x}_i} = \sum_{k=1}^{K}\sum_{l=1}^{2} \bar{\alpha}_l \bar{v}_l \bar{W}_l \mathbb{E}_{\mathbf{x},\mathbf{x}_i}\left[\mathbf{D}\mathbf{D}^T\right], \tag{84}$$

$$\mathcal{I}_{\mathrm{P}}^{\mathbf{x}_i,\mathbf{x}_i} = \mathbf{\Sigma}_i^{-1}, \tag{85}$$

$$\mathcal{I}_{\mathrm{M}}^{\mu_l,\mathbf{x}_i} = -\sum_{k=1}^{K} \bar{\alpha}_l \bar{v}_l \bar{W}_l \mathbb{E}_{\mathbf{x},\mathbf{x}_i}\left[\mathbf{D}^T\right], \tag{86}$$

$$\mathcal{I}_{\mathrm{M}}^{\mathbf{x}_i,\mathbf{x}_j} = \mathbf{0}, i \neq j, \tag{87}$$

while $\mathcal{I}_{\mathrm{M}}^{\boldsymbol{\mu},\boldsymbol{\mu}} + \mathcal{I}_{\mathrm{P}}^{\boldsymbol{\mu},\boldsymbol{\mu}}$ is a 2×2 diagonal matrix with diagonal elements,

$$\Pi_l = \mathcal{I}_{\mathrm{M}}^{\mu_l,\mu_l} + \mathcal{I}_{\mathrm{P}}^{\mu_l,\mu_l} = \sum_{i=1}^{M}\sum_{k=1}^{K} \bar{\alpha}_l \bar{v}_l \bar{W}_l + \bar{\beta}_l \bar{W}_l \bar{v}_l. \tag{88}$$

Moreover, $\mathbf{\Omega}_3 = \mathbf{0}$ follows from (82).

Finally, the submatrix $\mathbf{\Omega}$ can be written as

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_2 \end{bmatrix}. \tag{89}$$

REFERENCES

[1] Y. Shen and M. Z. Win, "Fundamental limits of wideband localization-part i: A general framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4956–4980, Oct 2010.

[2] C. Geng, X. Yuan, and H. Huang, "Exploiting channel correlations for nlos toa localization with multivariate gaussian mixture models," *IEEE Wireless Communications Letters*, vol. 9, no. 1, pp. 70–73, Jan 2020.

[3] Y. Wang, Y. Wu, and Y. Shen, "Joint spatiotemporal multipath mitigation in large-scale array localization," *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 783–797, Feb 2019.

[4] Z. Wang, H. Zhang, T. Lu, and T. A. Gulliver, "Cooperative rss-based localization in wireless sensor networks using relative error estimation and semidefinite programming," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 483–497, Jan 2019.

[5] Y. Sun, K. C. Ho, and Q. Wan, "Solution and analysis of tdoa localization of a near or distant source in closed form," *IEEE Transactions on Signal Processing*, vol. 67, no. 2, pp. 320–335, Jan 2019.

[6] M. Hamaoui, "Non-iterative mds method for collaborative network localization with sparse range and pointing measurements," *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 568–578, Feb 2019.

[7] M. Z. Win, F. Meyer, Z. Liu, W. Dai, S. Bartoletti, and A. Conti, "Efficient multisensor localization for the internet of things: Exploring a new class of scalable localization algorithms," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 153–167, Sept 2018.

[8] B. Xue, L. Zhang, Y. Yu, and W. Zhu, "Locating the nodes from incomplete euclidean distance matrix using bayesian learning," *IEEE Access*, vol. 7, pp. 37 406–37 413, 2019.

[9] H. Kim, S. W. Choi, and S. Kim, "Connectivity information-aided belief propagation for cooperative localization," *IEEE Wireless Communications Letters*, vol. 7, no. 6, pp. 1010–1013, Dec 2018.

[10] B. Li, N. Wu, H. Wang, P.-H. Tseng, and J. Kuang, "Gaussian message passing-based cooperative localization on factor graph in wireless networks," *Signal Processing*, vol. 111, pp. 1–12, 2015.

[11] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, Feb 2009.

[12] S. Zhang, S. Gao, G. Wang, and Y. Li, "Robust nlos error mitigation method for toa-based localization via second-order cone relaxation," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2210–2213, Dec 2015.

[13] L. Gavrilovska, V. Atanasovski, V. Rakovic, D. Denkovski, and M. Angjelicinoski, "Rem-enabled transmitter localization for ad hoc scenarios," in *MILCOM 2013 - 2013 IEEE Military Communications Conference*, Nov 2013, pp. 731–736.

[14] Z. Ma and K. C. Ho, "A study on the effects of sensor position error and the placement of calibration emitter for source localization," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5440–5452, Oct 2014.

[15] Y. Zou and Q. Wan, "Asynchronous time-of-arrival-based source localization with sensor position uncertainties," *IEEE Communications Letters*, vol. 20, no. 9, pp. 1860–1863, Sept 2016.

[16] W. Cui, B. Li, L. Zhang, and W. Meng, "Robust mobile location estimation in nlos environment using gmm, imm, and ekf," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3490–3500, Sep. 2019.

[17] K. C. Ho and L. Yang, "On the use of a calibration emitter for source localization in the presence of sensor position uncertainty," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5758–5772, Dec 2008.

[18] H. Lohrasbipeydeh, T. A. Gulliver, and H. Amindavar, "Unknown transmit power rssd based

source localization with sensor position uncertainty," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1784–1797, May 2015.

[19] D. Munoz, F. B. Lara, C. Vargas, and R. Enriquez-Caldera, *Position location techniques and applications*. Academic Press, 2009.

[20] R. Zekavat and R. M. Buehrer, *Handbook of position location: Theory, practice and advances*. John Wiley & Sons, 2011, vol. 27.

[21] U. Hammes and A. M. Zoubir, "Robust mobile terminal tracking in nlos environments based on data association," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5872–5882, Nov 2010.

[22] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich, and M. Coulon, "Direct localization for massive mimo," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2475–2487, May 2017.

[23] J. Zhang, J. Salmi, and E. Lohan, "Analysis of kurtosis-based los/nlos identification using indoor mimo channel measurement," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 6, pp. 2871–2874, July 2013.

[24] L. Yi, S. G. Razul, Z. Lin, and C. M. See, "Target tracking in mixed los/nlos environments based on individual measurement estimation and los detection," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 99–111, January 2014.

[25] X. Shi, G. Mao, Z. Yang, and J. Chen, "Localization algorithm design and performance analysis in probabilistic los/nlos environment," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[26] I. Guvenc, C. Chong, and F. Watanabe, "Nlos identification and mitigation for uwb localization systems," in *2007 IEEE Wireless Communications and Networking Conference*, March 2007, pp. 1571–1576.

[27] H. Wymeersch, S. Marano, W. M. Gifford, and M. Z. Win, "A machine learning approach to ranging error mitigation for uwb localization," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1719–1728, June 2012.

[28] T. V. Nguyen, Y. Jeong, H. Shin, and M. Z. Win, "Machine learning for wideband localization," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1357–1380, July 2015.

[29] F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "Em- and jmap-ml based joint estimation algorithms for robust wireless geolocation in mixed los/nlos environments," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 168–182, Jan 2014.

[30] C. Park and J. Chang, "Wls localization using skipped filter, hampel filter, bootstrapping and

gaussian mixture em in los/nlos conditions," *IEEE Access*, vol. 7, pp. 35 919–35 928, 2019.

[31] W. Wang, G. Wang, J. Zhang, and Y. Li, "Robust weighted least squares method for toa-based localization under mixed los/nlos conditions," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2226–2229, Oct 2017.

[32] W. Yuan, N. Wu, B. Etzlinger, H. Wang, and J. Kuang, "Cooperative joint localization and clock synchronization based on gaussian message passing in asynchronous wireless networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7258–7273, Sept 2016.

[33] K. C. Ho, X. Lu, and L. Kovavisaruch, "Source localization using tdoa and fdoa measurements in the presence of receiver location errors: Analysis and solution," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 684–696, Feb 2007.

[34] G. Naddafzadeh-Shirazi, M. B. Shenouda, and L. Lampe, "Second order cone programming for sensor network localization with anchor position uncertainty," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 749–763, February 2014.

[35] B. Li, N. Wu, H. Wang, and J. Kuang, "Expectation-maximisation-based localisation using anchors with uncertainties in wireless sensor networks," *IET Communications*, vol. 8, no. 11, pp. 1977–1987, July 2014.

[36] B. Zhou, Q. Chen, H. Wymeersch, P. Xiao, and L. Zhao, "Variational inference-based positioning with nondeterministic measurement accuracies and reference location errors," *IEEE Trans. Mob. Comput.*, vol. 16, no. 10, pp. 2955–2969, Oct 2017.

[37] S. Yousefi, R. Monir Vaghefi, X. Chang, B. Champagne, and R. M. Buehrer, "Sensor localization in nlos environments with anchor uncertainty and unknown clock parameters," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 742–747.

[38] Y. Xiong, N. Wu, Y. Shen, and M. Z. Win, "Cooperative network synchronization: Asymptotic analysis," *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 757–772, Feb 2018.

[39] N. M. Freris, H. Kowshik, and P. R. Kumar, "Fundamentals of large sensor networks: Connectivity, capacity, clocks, and computation," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1828–1846, Nov 2010.

[40] N. Nasios and A. G. Bors, "Variational learning for gaussian mixture models," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 4, pp. 849–862, 2006.

[41] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge university

press, 2003.

[42] M. J. Wainwright, M. I. Jordan *et al.*, "Graphical models, exponential families, and variational inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[43] M. Vemula, M. F. Bugallo, and P. M. Djurić, "Sensor self-localization with beacon position uncertainty," *Signal Process.*, vol. 89, no. 6, pp. 1144–1154, 2009.

[44] J. S. Liu, *Monte Carlo strategies in scientific computing*. Springer Science & Business Media, 2008.

[45] S. Kiranyaz, T. Ince, A. Yildirim, and M. Gabbouj, "Fractional particle swarm optimization in multidimensional search space," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 2, pp. 298–319, April 2010.

[46] K. K. Soo, Y. M. Siu, W. S. Chan, L. Yang, and R. S. Chen, "Particle-swarm-optimization-based multiuser detector for cdma communications," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 3006–3013, Sep. 2007.

[47] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*. Prentice Hall, 1993.

[48] Y. Wang, S. Ma, and C. L. P. Chen, "Toa-based passive localization in quasi-synchronous networks," *IEEE Communications Letters*, vol. 18, no. 4, pp. 592–595, April 2014.

[49] R. Mendrzik, H. Wymeersch, G. Bauch, and Z. Abu-Shaban, "Harnessing nlos components for position and orientation estimation in 5g millimeter wave mimo," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 93–107, Jan 2019.

[50] B. Zhou, Q. Chen, and P. Xiao, "The error propagation analysis of the received signal strength-based simultaneous localization and tracking in wireless sensor networks," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3983–4007, June 2017.

[51] B. Zhou, Q. Chen, P. Xiao, and L. Zhao, "On the spatial error propagation characteristics of cooperative localization in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1647–1658, Feb 2017.