

# Neural Enhanced Belief Propagation for Cooperative Localization

Mingchao Liang and Florian Meyer

University of California, San Diego, La Jolla, CA

Email: {m3liang, flmeyer}@ucsd.edu

## ABSTRACT

Location-aware networks will introduce innovative services and applications for modern convenience, applied ocean sciences, and public safety. In this paper, we establish a hybrid method for model-based and data-driven inference. We consider a cooperative localization (CL) scenario where the mobile agents in a wireless network aim to localize themselves by performing pairwise observations with other agents and by exchanging location information. A traditional method for distributed CL in large agent networks is belief propagation (BP) which is completely model-based and is known to suffer from providing inconsistent (overconfident) estimates. The proposed approach addresses these limitations by complementing BP with learned information provided by a graph neural network (GNN). We demonstrate numerically that our method can improve estimation accuracy and avoid overconfident beliefs, while its computational complexity remains comparable to BP. Notably, more consistent beliefs are obtained by not explicitly addressing overconfidence in the loss function used for training of the GNN.

**Index Terms**— Belief propagation, graph neural networks, cooperative localization, factor graph, agent networks

## 1. INTRODUCTION

Location awareness [1–8] is an important aspect in a variety of applications including autonomous navigation, applied ocean sciences, and public safety. Of particular interest are algorithmic solutions based on the framework of factor graphs and belief propagation (BP) [5–8] due to their ability to provide accurate results in high-dimensional nonlinear Bayesian estimation problems.

BP [9–11] is a message passing algorithm. It operates on the factor graph that represents the statistical model of an estimation problem. Given that the underlying factor graph is tree-structured, BP is guaranteed to provide the exact marginal posterior distributions or “beliefs” needed for optimal estimation. However, in cases where the factor graph has cycles or the statistical model represented by the factor graph does not accurately model the true data generating process, BP can only provide approximations of the marginal posterior distributions. In factor graphs with cycles, BP is typically faced by a lack of convergence guarantees. BP is also known to provide beliefs that are overconfident [10], i.e., the spread of the provided beliefs downplays the uncertainty of the estimates. This is particularly problematic in autonomous navigation applications where overconfidence can lead to catastrophic events [3]. We aim to improve the accuracy and reliability of BP-based localization and tracking algorithms by learning a refined model from data.

A graph neural network (GNN) [12, 13] is a type of neural network that implements a message passing mechanism similar to BP. It has been demonstrated that a learned GNN can outperform loop BP for Bayesian estimation if sufficient data is available [14]. Recently, [15] introduced neural enhanced belief propagation (NEBP) which pairs a factor graph with a GNN. The learned GNN messages

complement the corresponding BP messages to correct errors introduced by cycles and model mismatch. The resulting method combines the benefits of model-based and data-driven inference. NEBP can provide satisfactory estimation results when little data is available and leverages the performance advantages of GNNs in the large data regime. So far NEBP has only been considered for estimation problems with discrete random variables.

In this paper, we extend NEBP to estimation problems with continuous random variables and apply it to the cooperative localization (CL) problem [5–7, 16]. In particular, we represent the beliefs and messages related to continuous random variables by random samples or “particles” and update their weights by combining the BP message provided by the factor graph with the corresponding message provided by the GNN. Compared to BP-based CL, the proposed NEBP method has an improved estimation accuracy and can avoid overconfident beliefs.

The main contributions of this paper are as follows.

- We extend NEBP to continuous random variables and apply it to the CL problem.
- We demonstrate performance advantages compare to BP-based CL with a relative small amount of training data.

Our method preserves the advantages of BP-based methods for CL in wireless networks [5–7], i.e., it is fully distributed and requires little communication overhead, while its computational complexity only differs by a constant factor.

## 2. REVIEW OF PARTICLE-BASED BP FOR CL

We briefly review particle-based BP for CL which will be the basis for the development of the proposed NEBP method.

### 2.1. System Model and Problem Formulation

We consider a wireless network that consists of  $I$  mobile agents with indexes  $i \in \mathcal{I} \triangleq \{1, \dots, I\}$ . The topology of the agent network is described by the sets of neighbors  $\mathcal{N}_i \subseteq \mathcal{I} \setminus \{i\}$ . In particular, agent  $i$  is able to communicate and perform measurements with agents  $j \in \mathcal{N}_i$ . The state  $\mathbf{x}_{i,n} = [\mathbf{p}_{i,n}^T \mathbf{v}_{i,n}^T]^T$  of agent  $i \in \mathcal{I}$  at time  $n \in \{0, 1, \dots\}$  comprises the current position  $\mathbf{p}_{i,n} \in \mathbb{R}^d$  and other motion-related parameters  $\mathbf{v}_{i,n}$ . Agent motion is modeled by the mobility model

$$\mathbf{x}_{i,n} = f(\mathbf{x}_{i,n-1}, \mathbf{q}_{i,n}) \quad (1)$$

where  $\mathbf{q}_{i,n}$  is the driving noise with known probability density function (PDF)  $p(\mathbf{q}_{i,n})$  that is assumed to be statistically independent across  $i$  and  $n$ . From the state transition function (1) we can directly obtain the state-transition PDF  $p(\mathbf{x}_{i,n}|\mathbf{x}_{i,n-1})$ .

At time  $n$ , agent  $i$  exchanges information with neighboring agents  $j \in \mathcal{N}_i$  and performs a pairwise measurements that are modeled as

$$\mathbf{z}_{j \rightarrow i,n} = h(\mathbf{x}_{j,n}, \mathbf{x}_{i,n}, \mathbf{r}_{j \rightarrow i,n}) \quad (2)$$

where  $\mathbf{r}_{j \rightarrow i, n}$  is the measurement noise with known PDF  $p(\mathbf{r}_{j \rightarrow i, n})$  that is assumed to be statistically independent across edges  $(j, i)$ ,  $i \in \mathcal{I}$ ,  $j \in \mathcal{N}_i$  and time  $n$ . From the measurement model (2), we can directly obtain the likelihood function  $p(\mathbf{z}_{j \rightarrow i, n} | \mathbf{x}_{j, n}, \mathbf{x}_{i, n})$ .

Let  $\mathbf{x}_n = [\mathbf{x}_{i, n}]_{i \in \mathcal{I}}$  and  $\mathbf{z}_n = [\mathbf{z}_{j \rightarrow i, n}]_{i \in \mathcal{I}, j \in \mathcal{N}_i}$  be the joint state and measurement vectors at time  $n$ . Furthermore, we introduce  $\mathbf{x}_{0:n} = [\mathbf{x}_0^T \cdots \mathbf{x}_n^T]^T$  and  $\mathbf{z}_{1:n} = [\mathbf{z}_1^T \cdots \mathbf{z}_n^T]^T$ . The goal of CL is to estimate the states of the agents  $\mathbf{x}_{i, n}$ ,  $i \in \mathcal{I}$  from the joint measurement vector  $\mathbf{z}_{1:n}$  by using, e.g., the minimum mean square error (MMSE) estimator  $\hat{\mathbf{x}}_{i, n}^{\text{MMSE}} = \int \mathbf{x}_{i, n} p(\mathbf{x}_{i, n} | \mathbf{z}_{1:n}) d\mathbf{x}_{i, n}$ . Estimation relies on the marginal posterior distributions  $p(\mathbf{x}_{i, n} | \mathbf{z}_{1:n}) = \int p(\mathbf{x}_{0:n} | \mathbf{z}_{1:n}) \mathbf{x}_{i, n} d\mathbf{x}_{i, n}$ . However, direct marginalization from  $p(\mathbf{x}_{0:n} | \mathbf{z}_{1:n})$  is infeasible as its computation complexity grows exponentially with the number of time steps and the number of agents.

## 2.2. BP for CL

BP [9] aims to calculate marginal posteriors  $p(\mathbf{x}_{i, n} | \mathbf{z}_{1:n})$  efficiently by passing messages on the edges of the factor graph that represents the joint PDF of an estimation problem. In tree-structured graphs, the beliefs provided by BP are guaranteed to converge to the true marginal posterior distributions  $p(\mathbf{x}_{i, n} | \mathbf{z}_{1:n})$ . In graphs with cycles, (i) there are typically no convergence guarantees but BP can nevertheless often provide an accurate approximation of  $p(\mathbf{x}_{i, n} | \mathbf{z}_{1:n})$  [9, 10]; and (ii) there are many possible orders in which messages are computed (also known as the message schedules), and different orders may lead to different beliefs.

BP for CL can provide accurate approximations of marginal distributions at a computational complexity that scales linearly with the number of time steps and the agent network size. In particular, by assuming that at time  $n = 0$  the agent states  $\mathbf{x}_{i, 0}$ ,  $i \in \mathcal{I}$  are statistically independent and by using Bayes rule, the joint posterior distribution  $p(\mathbf{x}_{0:n} | \mathbf{z}_{1:n})$  factorizes according to

$$p(\mathbf{x}_{0:n} | \mathbf{z}_{1:n}) \propto \prod_{i=1}^I p(\mathbf{x}_{i, 0}) \prod_{n'=1}^n p(\mathbf{x}_{i, n'} | \mathbf{x}_{i, n'-1}) \times \prod_{j \in \mathcal{N}_i} p(\mathbf{z}_{j \rightarrow i, n'} | \mathbf{x}_{j, n'}, \mathbf{x}_{i, n'}).$$

A single time step of the corresponding cyclic factor graph is shown in Fig. 1(a).

This factor graph provides the basis for BP for CL where a specific message schedule makes it possible to perform message passing in real time and facilitates a distributed implementation. In particular, messages are sent only forward in time and  $t \in \{1, \dots, T\}$  message passing iterations are performed at each time step  $n$  individually. At each message passing iteration  $t$ , messages are passed only in one direction over every edge [6, 7]. The resulting BP algorithm consists of prediction and update steps that are executed for each agent  $i \in \mathcal{I}$  in parallel.

- **Prediction Step:** Based on the belief  $b_i^{(T)}(\mathbf{x}_{i, n-1})$  calculated at the previous time step  $n - 1$  and the state transition PDF  $p(\mathbf{x}_{i, n} | \mathbf{x}_{i, n-1})$ , the “prediction message”  $\mu_{i, \rightarrow n}(\mathbf{x}_{i, n})$  is obtained as

$$\mu_{i, \rightarrow n}(\mathbf{x}_{i, n}) \propto \int p(\mathbf{x}_{i, n} | \mathbf{x}_{i, n-1}) b_i^{(T)}(\mathbf{x}_{i, n-1}) d\mathbf{x}_{i, n-1}. \quad (3)$$

At time step  $n = 0$ , message passing is initialized by setting  $b_i^{(T)}(\mathbf{x}_{i, 0}) \triangleq p(\mathbf{x}_{i, 0})$ .

- **Update Step:** At message passing iteration  $t \in \{1, \dots, T\}$ , beliefs  $b_j^{(t-1)}(\mathbf{x}_{j, n})$  are received from neighboring agents  $j \in \mathcal{N}_i$ .

Next, corresponding “measurement messages”  $\mu_{j \rightarrow i, n}^{(t)}(\mathbf{x}_{i, n})$  are calculated based on the likelihood function  $p(\mathbf{z}_{j \rightarrow i, n} | \mathbf{x}_{j, n}, \mathbf{x}_{i, n})$ , i.e.,

$$\mu_{j \rightarrow i, n}^{(t)}(\mathbf{x}_{i, n}) \propto \int p(\mathbf{z}_{j \rightarrow i, n} | \mathbf{x}_{j, n}, \mathbf{x}_{i, n}) b_j^{(t-1)}(\mathbf{x}_{j, n}) d\mathbf{x}_{j, n}. \quad (4)$$

Finally, the belief at message passing iteration  $t$  is obtained as

$$b_i^{(t)}(\mathbf{x}_{i, n}) \propto \mu_{i, \rightarrow n}(\mathbf{x}_{i, n}) \prod_{j \in \mathcal{N}_i} \mu_{j \rightarrow i, n}^{(t)}(\mathbf{x}_{i, n}). \quad (5)$$

At message passing iteration  $t = 1$ , beliefs are initialized as  $b_j^{(0)}(\mathbf{x}_{j, n}) = \mu_{i, \rightarrow n}(\mathbf{x}_{j, n})$  (cf. (4)).

## 2.3. Particle-Based Processing

Often, the mobility and measurement models in (1) and (2) are non-linear and non-Gaussian and it is impossible to obtain a closed-form solution for the message passing and belief calculation equations in (3)–(5). This problem can be addressed by representing messages and beliefs by  $K$  weighted random samples or “particles” and approximating (3)–(5) by means of Monte Carlo techniques [17, 18]. The solution of the resulting particle-based processing can be arbitrary close to the corresponding true solution of (3)–(5) by choosing  $K$  sufficiently large.

- **Prediction Step:** Let  $\{\mathbf{x}_{i, n-1}^{(k)}, w_{i, n-1}^{(T, k)}\}_{k=1}^K$  be a particle representation of  $b_i^{(T)}(\mathbf{x}_{i, n-1})$ . Then, a particle representation  $\{\mathbf{x}_{i, n}^{(k)}, w_{i, n}^{(k)}\}_{k=1}^K$  of  $\mu_{i, \rightarrow n}(\mathbf{x}_{i, n})$  in (3), can be obtained by drawing, for each  $k \in \{1, \dots, K\}$ , one particle  $\mathbf{x}_{i, n}^{(k)}$  from  $p(\mathbf{x}_{i, n} | \mathbf{x}_{i, n-1}^{(k)})$ , i.e.,

$$\mathbf{x}_{i, n}^{(k)} \sim p(\mathbf{x}_{i, n} | \mathbf{x}_{i, n-1}^{(k)}), \quad w_{i, n}^{(k)} = w_{i, n-1}^{(T, k)}.$$

At  $n = 0$ , we draw particles from the prior, i.e.  $\mathbf{x}_{i, 0}^{(k)} \sim p(\mathbf{x}_{i, 0})$ , and set corresponding weights according to  $w_{i, 0}^{(T, k)} = \frac{1}{K}$ .

- **Update Step:** Following the message multiplication scheme in [17] a particle-based representation of  $b_i^{(t)}(\mathbf{x}_{i, n})$  in (5) is calculated for each message passing iteration  $t \in \{1, \dots, T\}$ . In particular, particle-based messages

$$\phi_{j \rightarrow i}^{(t, k)} = p(\mathbf{z}_{j \rightarrow i, n} | \mathbf{x}_{j, n}^{(k)}, \mathbf{x}_{i, n}^{(k)}) \tilde{w}_{j, n}^{(t-1, k)} \quad (6)$$

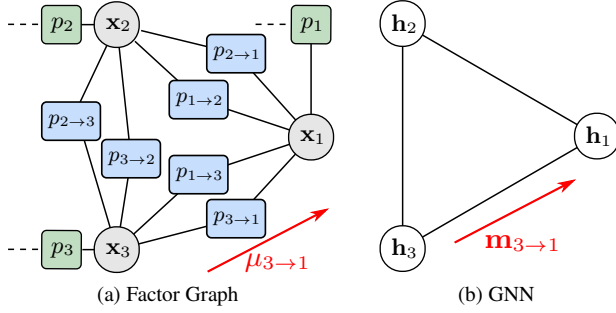
are computed for each neighbor  $j \in \mathcal{N}_i$  and weight update is performed according to

$$\tilde{w}_{i, n}^{(t, k)} = w_{i, n}^{(k)} \prod_{j \in \mathcal{N}_i} \phi_{j \rightarrow i}^{(t, k)}. \quad (7)$$

At message passing iteration  $t = 1$ , (6) is initialized as  $\phi_{j \rightarrow i}^{(1, k)} = p(\mathbf{z}_{j \rightarrow i, n} | \mathbf{x}_{j, n}^{(k)}, \mathbf{x}_{i, n}^{(k)}) w_{j, n}^{(k)}$ . Finally, particles and unnormalized weights  $\{\mathbf{x}_{i, n}^{(k)}, \tilde{w}_{i, n}^{(t, k)}\}_{k=1}^K$  are exchanged among neighboring agents.

After the last iteration ( $t = T$ ), the weights are normalized, i.e.,  $w_{i, n}^{(T, k)} = \tilde{w}_{i, n}^{(T, k)} / \sum_{k'=1}^K \tilde{w}_{i, n}^{(T, k')}$ ,  $k \in \{1, \dots, K\}$ , and an approximation of the MMSE estimate can be obtained as  $\hat{\mathbf{x}}_{i, n} = \sum_{k=1}^K w_{i, n}^{(T, k)} \mathbf{x}_{i, n}^{(k)}$ . The particle representation  $\{\mathbf{x}_{i, n}^{(k)}, w_{i, n}^{(T, k)}\}_{k=1}^K$  is also needed for the prediction step at the next time step  $n + 1$ .

For future reference, we introduce the weight vector  $\mathbf{w}_i = [w_i^{(1)} \cdots w_i^{(K)}]^T$  of agent  $i$  after the prediction step and the joint weight vector  $\mathbf{w} = [\mathbf{w}_i]_{i \in \mathcal{I}}$ . Similarly, we introduce the vectors of particle-based BP messages  $\phi_{j \rightarrow i}^{(t)} = [\phi_{j \rightarrow i}^{(t, 1)} \cdots \phi_{j \rightarrow i}^{(t, K)}]^T$  and  $\phi^{(t)} = [\phi_{j \rightarrow i}^{(t)}]_{i \in \mathcal{I}, j \in \mathcal{N}_i}$ .



**Fig. 1:** Factor graph for CL (a) and corresponding graph neural network (GNN) (b) for a single time slot  $n$ . The BP and GNN messages related to the messages  $3 \rightarrow 1$  are also shown. The time index  $n$  is omitted and the short notations  $p_{j \rightarrow i} \triangleq p(\mathbf{z}_{j \rightarrow i, n} | \mathbf{x}_{j, n}, \mathbf{x}_{i, n})$ ,  $p_i \triangleq p(\mathbf{x}_{i, n} | \mathbf{x}_{i, n-1})$ ,  $\mu_{j \rightarrow i} = \mu_{j \rightarrow i, n}(\mathbf{x}_{i, n})$ , and  $\mathbf{m}_{j \rightarrow i} = \mathbf{m}_{j \rightarrow i}^{(t)}$  are used.

### 3. NEBP FOR CL

In this section, we will review GNNs and present the proposed particle-based NEBP framework for CL. In particular, at each time step  $n$  we complement iterative particle-based BP (6)–(7) by a GNN. Since we limit our discussion to a single time step, we will omit the time index  $n$  in what follows.

#### 3.1. Graph Neural Networks (GNNs)

GNNs [12] extend neural networks to graph-structured data. We consider the message passing neural network (MPNN) [19] which is a variant of GNNs that generalizes graph convolutional networks [20] and implements a message passing mechanism similar to BP. A MPNN is defined on a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  where  $\mathcal{E}$  induces the sets of neighbors  $\mathcal{N}_i = \{j \in \mathcal{V} | (i, j) \in \mathcal{E}\}$ . There is one neural network for each node and each edge in the graph. In many applications all node networks and all edge networks share sets of parameters, respectively.

Each node  $i \in \mathcal{V}$  is associated with a vector  $\mathbf{h}_i$  called node embedding. At message passing iteration  $t \in \{1, \dots, T\}$ , the following operations are performed for each node  $i \in \mathcal{V}$  in parallel. First, messages are exchanged with neighboring nodes  $j \in \mathcal{N}_i$ . In particular, the GNN message sent from node  $i \in \mathcal{V}$  to its neighbor  $j \in \mathcal{N}_i$  is given by

$$\mathbf{m}_{i \rightarrow j}^{(t)} = g_e(\mathbf{h}_i^{(t)}, \mathbf{h}_j^{(t)}, \mathbf{a}_{i \rightarrow j}) \quad (8)$$

where  $g_e(\cdot)$  is a neural network with trainable parameters and  $\mathbf{a}_{i \rightarrow j}$  is the edge attribute. Next, the node embedding  $\mathbf{h}_i^{(t)}$  is updated by incorporating the sum of received messages  $\mathbf{m}_{j \rightarrow i}^{(t)}$ ,  $j \in \mathcal{N}_i$ , i.e.,

$$\mathbf{h}_i^{(t)} = g_n\left(\mathbf{h}_i^{(t-1)}, \sum_{j \in \mathcal{N}_i} \mathbf{m}_{j \rightarrow i}^{(t-1)}\right). \quad (9)$$

Here,  $g_n(\cdot)$  is again a neural network with trainable parameters.

Since  $g_e(\cdot)$  and  $g_n(\cdot)$  share the same parameters across edges and nodes, respectively, they can be trained on small graphs even if then used in large scale inference problems. For future reference, we introduce the joint vector of node embeddings  $\mathbf{h} = [\mathbf{h}_i]_{i \in \mathcal{I}}$  and the joint vector of messages  $\mathbf{m} = [\mathbf{m}_{j \rightarrow i}]_{i \in \mathcal{I}, j \in \mathcal{N}_i}$ .

#### 3.2. Particle-based NEBP

The main idea of particle-based NEBP is to use the particle representation  $\{\mathbf{x}_{i, n}^{(k)}, \mathbf{w}_{i, n}^{(t, k)}\}_{k=1}^K$  of a continuous state  $\mathbf{x}_{i, n}$  as if it would be the probability mass function (PMF) of a discrete random variable, i.e., the particles  $\mathbf{x}_{i, n}^{(k)}$ ,  $k \in \{1, \dots, K\}$  are the possible outcomes and the particle weights  $w_{i, n}^{(k)}$  are the probabilities of these outcomes.

The GNN that is complementary to one time step of the CL factor graph is shown in Fig. 1(b). Since the CL factor graph only consists of pairwise interactions, we can use a simpler GNN compared to the one originally proposed for NEBP [15], i.e. a GNN that only models variables nodes in the factor graph by a corresponding GNN node. In what follows, we denote the NEBP messages and their joint vector by  $\underline{\phi}^{(t)}$  and  $\underline{\phi}_{j \rightarrow i}^{(t)}$ , respectively.

NEBP for CL consists of the following three steps:

1. First, classical BP runs for one iteration, i.e.,

$$\phi^{(t)} = \text{BP}(\underline{\phi}^{(t-1)}, \mathbf{w}). \quad (10)$$

Here  $\text{BP}(\cdot)$  is the function that takes NEBP messages  $\underline{\phi}^{(t-1)}$  and  $\mathbf{w}$  as inputs and returns the BP messages  $\phi^{(t)}$  by first computing (7) (with  $t$  replaced by  $t-1$  and  $\phi_{j \rightarrow i}^{(t, k)}$  replaced by  $\phi_{j \rightarrow i}^{(t-1, k)}$ ) followed by (6) for all edges  $(j, i)$ ,  $i \in \mathcal{I}$ ,  $j \in \mathcal{N}_i$  in the network. At iteration  $t = 1$ ,  $\underline{\phi}^{(t-1)}$  in (10) is replaced by the all-ones vector with dimension  $K \sum_{i=1}^I |\mathcal{N}_i|$ .

2. Next, the output of the GNN is computed, i.e.,

$$[\mathbf{h}^{(t+1)\top} \mathbf{m}^{(t)\top}]^\top = \text{GNN}(\mathbf{h}^{(t)}, \phi^{(t)})$$

where  $\text{GNN}(\cdot)$  is the function that calculates GNN messages  $\mathbf{m}^{(t)}$  and updated node embeddings  $\mathbf{h}^{(t+1)}$  from the current node embeddings  $\mathbf{h}^{(t)}$  and the classical BP messages  $\phi^{(t)}$  using (8)–(9). In particular, the classical BP messages  $\phi_{j \rightarrow i}^{(t)}$ ,  $i \in \mathcal{I}$ ,  $j \in \mathcal{N}_i$  are used as the edge attributes  $\mathbf{a}_{j \rightarrow i}$  in (8).

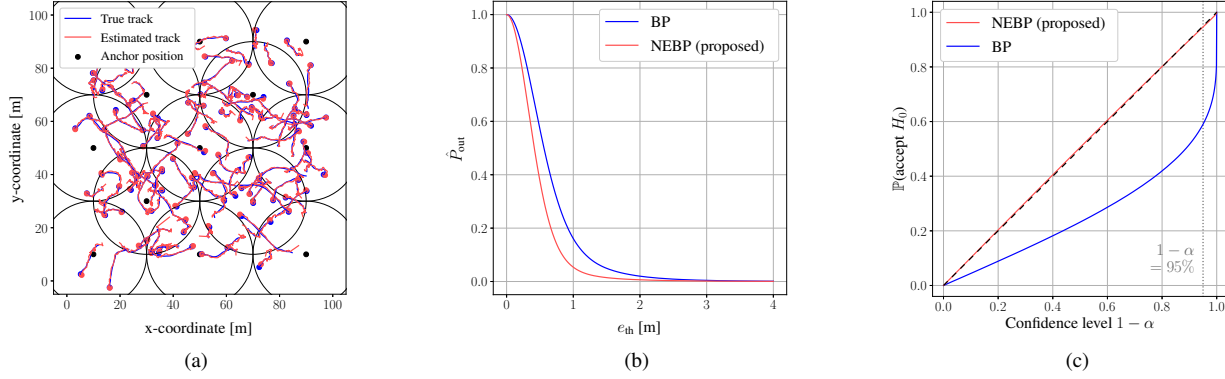
At  $t = 1$ ,  $\mathbf{h}^{(1)}$  is initialized by setting  $\mathbf{h}_i^{(1)} = [\hat{\mathbf{x}}_i^\top \text{vec}(\hat{\mathbf{C}}_i)^\top]^\top$  for all  $i \in \mathcal{I}$ , where  $\hat{\mathbf{x}}_i = \sum_{k=1}^K w_i^{(k)} \mathbf{x}_i^{(k)}$  is the sample mean,  $\hat{\mathbf{C}}_i = \sum_{k=1}^K w_i^{(k)} (\mathbf{x}_i^{(k)} - \hat{\mathbf{x}}_i)(\mathbf{x}_i^{(k)} - \hat{\mathbf{x}}_i)^\top$  is the sample covariance, and  $\text{vec}(\cdot)$  creates a vector from the input matrix by taken elements columnwise.

3. Finally, for all edges  $(j, i)$ ,  $i \in \mathcal{I}$ ,  $j \in \mathcal{N}_i$ , GNN messages  $\mathbf{m}_{j \rightarrow i}^{(t)}$  and BP messages  $\phi_{j \rightarrow i}^{(t)}$  are combined according to

$$\underline{\phi}_{j \rightarrow i}^{(t)} = g_s(\mathbf{m}_{j \rightarrow i}^{(t)}) \phi_{j \rightarrow i}^{(t)} + g_v(\mathbf{m}_{j \rightarrow i}^{(t)}).$$

The functions  $g_s(\cdot)$  and  $g_v(\cdot)$  are neural networks with learnable parameters and output a positive scalar and a positive vector, respectively.

After  $T$  iterations, the particle representation  $\{\mathbf{x}_i^{(k)}, \mathbf{w}_i^{(T, k)}\}_{k=1}^K$  of the NEBP belief  $b_i^{(T)}(\mathbf{x}_i)$  is obtained by calculating weights  $\mathbf{w}_i^{(T, k)}$  based on (7) with BP messages  $\phi_{j \rightarrow i}^{(T, k)}$  replaced by NEBP messages  $\underline{\phi}_{j \rightarrow i}^{(T, k)}$ . The particle representation  $\{\mathbf{x}_i^{(k)}, \mathbf{w}_i^{(T, k)}\}_{k=1}^K$  can then be used to calculate an approximate MMSE estimate  $\hat{\mathbf{x}}_i$  and the corresponding approximate covariance matrix. Since we aim to reduce the MSE of the position estimate  $\hat{\mathbf{p}}_i$ , the neural networks  $g_e(\cdot)$ ,  $g_n(\cdot)$ ,  $g_s(\cdot)$ , and  $g_v(\cdot)$  are trained based on the loss function  $\sum_{i=1}^I \|\hat{\mathbf{p}}_i - \mathbf{p}_i\|^2$ .



**Fig. 2:** Numerical evaluation of the proposed NEBP method in a CL scenario with 13 static anchors and 100 mobile agents. (a) One realization of true and estimated mobile agent tracks with dots indicating the final positions. (b) Outage probability versus threshold  $e_{\text{th}}$ . (c) Probability that an estimate is consistent versus confidence levels. (The gray dotted-line indicates a confidence level of 95%.)

#### 4. EXPERIMENTS

In this section, we compare the performance the proposed NEBP method with BP. The neural networks  $g_e(\cdot)$ ,  $g_s(\cdot)$ ,  $g_v(\cdot)$  are multi-layer perceptrons (MLPs) with a single hidden layer and leaky rectified linear units (ReLU) [21], with exception that the output layers of  $g_s(\cdot)$  and  $g_v(\cdot)$  use sigmoid and ReLU activations, respectively. We set the number of message passing iterations to  $T = 1$ , in which case the node function  $g_n(\cdot)$  is not needed. During training, the parameters of  $g_e(\cdot)$ ,  $g_s(\cdot)$ , and  $g_v(\cdot)$  are updated through back-propagation. The dimension of node embeddings  $\mathbf{h}_i$  is 20 and the dimension of GNN messages  $\mathbf{m}_{j \rightarrow i}$  is 32.

##### 4.1. Dataset and Training Procedure

We consider agent networks in two dimensional (2-D) space. The state of each agent at time  $n$  is defined as  $\mathbf{x}_{i,n} = [\mathbf{p}_{i,n}^T \mathbf{v}_{i,n}^T]^T \in \mathbb{R}^4$  where  $\mathbf{p}_{i,n} \in \mathbb{R}^2$  and  $\mathbf{v}_{i,n} \in \mathbb{R}^2$  are the 2-D position and velocity, respectively. We use a constant-velocity motion model with drag force and Gaussian driving noise with standard deviation  $\sigma_a = 0.05$  (see [22] for details). Furthermore, we consider measurements of the distance  $z_{j \rightarrow i,n} = \|\mathbf{p}_{j,n} - \mathbf{p}_{i,n}\| + r_{j \rightarrow i,n}$ , where  $r_{j \rightarrow i,n}$  is zero-mean Gaussian noise with standard deviation  $\sigma_r = 1$ .

We consider  $I = 25$  agents on the area of interest  $[0, 60]\text{m} \times [0, 60]\text{m}$ . There are five static anchors at perfectly known locations, i.e., their state transition model and prior distribution are given by  $p(\mathbf{x}_{i,n}|\mathbf{x}_{i,n-1}) = \delta(\mathbf{x}_{i,n} - \mathbf{x}_{i,n-1})$  and  $p(\mathbf{x}_{i,0}) = \delta(\mathbf{x}_{i,0} - [\bar{\mathbf{p}}_i^T \ 0 \ 0]^T)$  where  $\bar{\mathbf{p}}_i$  is the true anchor position. In each realization, the mobile agents are uniformly placed over the area  $[15, 45]\text{m} \times [15, 45]\text{m}$  and their velocity is randomly drawn from  $\mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I}_2)$  with  $\sigma_p = 0.1$ . For each agent, a track that consists of 50 time steps is generated and range measurements are obtained by assuming a connectivity of 20m, i.e.  $j \in \mathcal{N}_i$  if and only if  $\|\mathbf{p}_{j,n} - \mathbf{p}_{i,n}\| \leq 20\text{m}$ . For inference, the initial prior distribution is  $p(\mathbf{x}_{i,0}) = \mathcal{N}(\boldsymbol{\mu}_{i,0}, \boldsymbol{\Sigma}_{i,0})$ . Here  $\boldsymbol{\Sigma}_{i,0} = \text{diag}\{10, 10, 0.01, 0.01\}$  and  $\boldsymbol{\mu}_{i,0}$  is randomly drawn from  $\mathcal{N}(\mathbf{x}_{i,0}, \boldsymbol{\Sigma}_{i,0})$ , where  $\mathbf{x}_{i,0}$  is the true initial state of agent  $i$ .

For training, an Adam optimizer [23] with learning rate  $10^{-4}$  and batch size of 2 is used. Furthermore, 100 realizations of agent tracks and 10 passes of the entire training dataset are considered. A larger agent network is employed for performance evaluation to show the generalization ability of our NEBP method. In particular, we consider a network that consists of 13 static anchors and 100 mobile agents, i.e.,  $I = 100$  in the area  $[0, 100]\text{m} \times [0, 100]\text{m}$  (see [6])

and generate another 400 realizations of agents tracks. The agents are uniformly placed over  $[10, 90]\text{m} \times [10, 90]\text{m}$  at time  $n = 0$ . All the other parameters are set as during training. Anchor positions and a realization of mobile agent tracks are shown in Fig 2a.

##### 4.2. Performance Evaluation

To evaluate the performance of different localization algorithms, we use the outage probability  $P_{\text{out}} = \mathbb{P}(\|\hat{\mathbf{p}}_{i,n} - \mathbf{p}_{i,n}\| > e_{\text{th}})$ , where  $\mathbf{p}_{i,n}$  is the true position,  $\hat{\mathbf{p}}_{i,n}$  is the estimate, and  $e_{\text{th}} > 0$  is the error threshold. Fig. 2b shows the outage probability versus threshold  $e_{\text{th}}$ . It can be seen that our proposed NEBP algorithm yields significantly reduced outage probability compared to BP. Fig. 2a shows one realization of true and estimated agent tracks.

To assess the consistency of BP and NEBP estimates, we conduct two-sided chi-square tests. We first calculate the normalized estimation error squared (NEES) [1],  $e_{i,n} = (\hat{\mathbf{p}}_{i,n} - \mathbf{p}_{i,n})^T \hat{\boldsymbol{\Sigma}}_{\mathbf{p},i,n}^{-1} (\hat{\mathbf{p}}_{i,n} - \mathbf{p}_{i,n})$ , where  $\hat{\boldsymbol{\Sigma}}_{\mathbf{p},i,n} \in \mathbb{R}^{2 \times 2}$  is the estimated position covariance matrix. Assuming that the true posterior distribution is Gaussian, the NEES follows a chi-square distribution with degree of freedom 2. Let  $H_0$  be the hypothesis that the belief of agent  $i$  is consistent, i.e.,  $\hat{\mathbf{p}}_{i,n}$  and  $\hat{\boldsymbol{\Sigma}}_{\mathbf{p},i,n}$  are the true mean and covariance matrix.  $H_0$  is accepted if  $e_{i,n} \in [r_1, r_2]$ , where  $r_1, r_2$  is determined such that  $\mathbb{P}(e_{i,n} \leq r_1|H_0) = \mathbb{P}(e_{i,n} \geq r_2|H_0) = \frac{\alpha}{2}$  and  $1 - \alpha$  is the confidence level. Ideally, as indicated by the black dashed line in Fig. 2c, at confidence level  $1 - \alpha$ , the probability that  $H_0$  is accepted should be  $1 - \alpha$ . However, the BP solution has  $\mathbb{P}(\text{accept } H_0)$  significantly smaller than  $1 - \alpha$ . At 95% confidence level, 40% of the NEES values  $e_{i,n}$  fall outside the confidence interval, indicating that BP provides inconsistent estimates. On the other hand, the proposed NEBP method is close to the ideal line, where only 5% of NEES values  $e_{i,n}$  fall outside the 95% confidence interval [1]. It can thus be concluded that NEBP significantly improves the consistency of estimates. Notably, more consistent estimates are obtained by not explicitly addressing overconfidence in the loss function used for training of the GNN.

#### 5. CONCLUSION

In this paper, we propose a particle-based NEBP method for CL that combines the benefits of model-based and data-driven inference. The proposed approach complements BP with learned information provided by a GNN. Simulation results show that NEBP outperforms traditional BP in terms of localization error and consistency of estimates as well as generalizes to larger agent networks.

## 6. REFERENCES

- [1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY: Wiley, 2001.
- [2] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.
- [3] A. Bahr, M. R. Walter, and J. J. Leonard, "Consistent cooperative localization," in *Proc. IEEE ICRA-09*, Kobe, Japan, 2009, pp. 3415–3422.
- [4] R. D. Taranto, S. Muppirisetty, R. Raulefs, D. T. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 102–112, Nov. 2014.
- [5] A. T. Ihler, J. W. Fisher III, R. L. Moses, and A. S. Willsky, "Nonparametric belief propagation for self-localization of sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 809–819, Apr. 2005.
- [6] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, 2009.
- [7] M. Z. Win, F. Meyer, Z. Liu, W. Dai, S. Bartoletti, and A. Conti, "Efficient multi-sensor localization for the Internet-of-Things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 153–167, Sep. 2018.
- [8] F. Meyer, T. Kropfreiter, J. L. Williams, R. A. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, "Message passing algorithms for scalable multitarget tracking," *Proc. IEEE*, vol. 106, no. 2, pp. 221–259, Feb. 2018.
- [9] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [10] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Comput.*, vol. 13, no. 10, pp. 2173–2200, 2001.
- [11] B. Li and Y.-C. Wu, "Convergence of Gaussian belief propagation under general pairwise factorization: Connecting Gaussian MRF with pairwise linear Gaussian model," *JMLR*, vol. 20, no. 144, pp. 1–30, 2019.
- [12] M. Gori, G. Monfardini, and F. Scarselli, "A new model for learning in graph domains," in *Proc. INNS/IEEE IJCNN-05*, vol. 2, Montreal, Canada, 2005, pp. 729–734.
- [13] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, 2009.
- [14] K. Yoon, R. Liao, Y. Xiong, L. Zhang, E. Fetaya, R. Urtasun, R. Zemel, and X. Pitkow, "Inference in probabilistic graphical models by graph neural networks," in *Proc. IEEE Asilomar-19*, Pacific Grove, CA, 2019, pp. 868–875.
- [15] V. G. Satorras and M. Welling, "Neural enhanced belief propagation on factor graphs," in *Proc. AISTATS-21*, Apr. 2021, pp. 685–693.
- [16] F. Meyer, O. Hlinka, and F. Hlawatsch, "Sigma point belief propagation," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 145–149, Feb. 2014.
- [17] F. Meyer, O. Hlinka, H. Wymeersch, E. Riegler, and F. Hlawatsch, "Distributed localization and tracking of mobile networks including noncooperative objects," *IEEE Trans. Signal and Inf. Process. over Networks*, vol. 2, no. 1, pp. 57–71, 2016.
- [18] F. Meyer, H. Wymeersch, M. Fröhle, and F. Hlawatsch, "Distributed estimation with information-seeking control in agent networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, Nov. 2015.
- [19] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proc. ICML-17*, Sydney, Australia, Aug. 2017, pp. 1263–1272.
- [20] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR-17*, Toulon, France, Apr. 2017.
- [21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.
- [22] S. Van de Velde, G. T. de Abreu, and H. Steendam, "Improved censoring and NLOS avoidance for wireless localization in dense networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2302–2312, 2015.
- [23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.