Breaking the Gaussian Barrier: Leveraging ReGVAMP to Extend EKF and IEKF

Fangqing Xiao, Dirk Slock
Communication Systems Department, EURECOM, France
fangqing.xiao@eurecom.fr, dirk.slock@eurecom.fr

***Abstract*—In the non-linear hidden Markov chain (HMC) model, commonly employed in robotics, navigation, signal processing, and control systems, various variants of Kalman filter (KF) like the Extended Kalman Filter (EKF) and Iterated Extended Kalman Filter (IEKF) have been developed for different scenarios, with their limitations and advantages. However, their effectiveness can diminish in the presence of non-Gaussian noise, which is prevalent in many real-world situations. This paper presents a novel approach to tackle non-Gaussian noise leveraging the revisited Generalized Vector Approximate Message Passing (ReGVAMP) within the context of EKF and IEKF. ReGVAMP extends these KF variants to break the Gaussian barrier, thereby enhancing the accuracy of state estimation. A tracking simulation validated the feasibility of our proposed algorithms.**

I. Introduction

The Kalman Filter (KF) [1] is a foundational tool in state estimation, particularly effective for linear systems with Gaussian noise. However, its utility is restricted by the assumption of linearity and Gaussianity, leading to the development of extensions to handle nonlinearities and non-Gaussian noise sources [2]. Key among these extensions are the Extended Kalman Filter (EKF) [3] and Iterated Extended Kalman Filter (IEKF) [4], which play crucial roles in various fields such as robotics, navigation, signal processing, and control systems [5], [6]. These algorithms expand upon the capabilities of the conventional Kalman Filter, enabling it to effectively deal with nonlinear systems. The EKF, an extension of the KF, addresses nonlinear systems by linearizing system dynamics and measurement functions around the current state estimate using first-order Taylor expansion. Despite its wide application, the EKF has limitations. Its linearization step introduces errors, particularly evident in highly nonlinear systems.

Further enhancing filtering accuracy, the IEKF iteratively refines the state estimate. Unlike the EKF, which relies on a single linearization, the IEKF iterates between prediction and update steps, linearizing around the current best estimate at each iteration. This iterative approach reduces errors introduced by linearization, thereby improving filtering performance, especially for highly nonlinear systems. However, achieving IEKF convergence may depend on initial conditions and could require parameter tuning for optimal performance.

Despite their effectiveness, the assumption of Gaussian noise distributions in these filters restricts their applicability in scenarios with non-Gaussian noise processes. While they can be adapted for non-Gaussian noises, achieving Linear Minimum Mean Square Error (LMMSE) performance when precise linearization achieve.

A straightforward approach might involve calculating the true posterior at each time index then approximate it as Gaussian prior for next time index. Setting aside considerations of their linearization accuracy and stability, they consistently yield superior performance compared to LMMSE. However, the solvability of MMSE integration is often challenging due to the high-dimensional integration involved, necessitating the use of approximate approaches for Bayesian inference. Previous research has introduced several numerical methods based on message passing to obtain an analytic solution to the integration in the time update step, including generalized approximate message passing (GAMP) [7], vector approximate message passing (VAMP) [8], variational inference (VI) [9], etc.. Nevertheless, these methods typically assume a high system dimension, aiming to avoid complex matrix inversions, and are sensitive to the measurement matrix, which must be right rotationally invariant for VAMP and consist of i.i.d. sub-Gaussian elements for GAMP. Additionally, these algorithms only provide averaged variances. These limitations motivated the introduction of the revisited generalized vector approximate message passing (ReGVAMP) algorithm [10]. Leveraging the ReGVAMP algorithm [10], we approximate the posterior $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ as Gaussian $q(\boldsymbol{x}_k)$ prior for next time index. The key idea behind ReGVAMP is to approximate the extrinsic information into Gaussian form due to central limit theory. Consequently, the minimization of the Kullback-Leibler divergence (KLD) between $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ and $q(\boldsymbol{x}_k)$ is separated into several moment matching problems. Optimizing the approximated posterior $q(\boldsymbol{x}_k)$ via KLD$(p||q)$ may yield better performance than KLD$(q||p)$ as VI does, as the former provides the exact posterior covariance matrix while the latter only offers correct diagonal precisions even when p is Gaussian. Moreover, while this approximated multivariate Gaussian posterior may not be tight, it captures full second-order moments. Hence, from one perspective, ReGVAMP algorithm minimizes the desirable KLD$(p||q)$ approximately, made feasible by the asymptotic Gaussianity of the extrinsic. In this paper, we propose combining ReGVAMP with EKF and IEKF to handle different nonlinear and non-Gaussian models, coining them as ReGVAMP-EKF and ReGVAMP-IEKF. Additionally, comparison experiments is designed to compare them to the unscented Kalman filter (UKF) [11]. Through our simulation, it is evident that when faced with non-Gaussian noise, introducing the ReGVAMP algorithm to handle such noise significantly enhances the performance of both EKF and IEKF.

A. Notations

The operator $(\cdot)^T$ denotes the matrix transpose. We denote the Kullback–Leibler divergence between distributions p and q as $D_{KL}(p\|q)$. The notation $\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ represents the Gaussian distribution function evaluated at $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\boldsymbol{e}_i$ indicates a unit vector with only the i th entry set to 1 and all others set to 0. $\delta(\boldsymbol{x})$ denotes the Dirac delta distribution. $\mathbf{I}_n$ denotes the $n\times n$ identity matrix. The notation $|x|$ is used to denote the absolute value of x.

II. Generalities on Linear State Space Model

Consider a non-linear state space model at each time step k:

$$\boldsymbol{x}_k = h(\boldsymbol{x}_{k-1},\mathbf{u}_k,\boldsymbol{w}_k) \tag{1a}$$
$$\boldsymbol{y}_k = g(\boldsymbol{x}_k,\boldsymbol{v}_k), \tag{1b}$$

where $\boldsymbol{x}_k\in\mathbb{R}^N$, $\mathbf{u}_k\in\mathbb{R}^O$, and $\boldsymbol{y}_k\in\mathbb{R}^M$ represent the true unobserved state vector of the system, control vector, and only observation (measurement) respectively. The process noise $\boldsymbol{w}_k\in\mathbb{R}^L$ drives the dynamic system and the observation noise is denoted as $\boldsymbol{v}_k\in\mathbb{R}^C$. This entire process can be interpreted as a hidden Markov chain (HMC) [12]. While the standard Kalman filter is a powerful estimation tool, its algorithm begin to break down when the system being estimated is nonlinear. Fortunately, a version of the standard Kalman filter, known as the extended Kalman filter (EKF), has been extended to nonlinear systems and relies on linearization in estimating these nonlinear systems. Linearization operates on the principle that at a small section around a selected operating point a nonlinear function can be approximated as a linear function. This linearized function can be derived from the nonlinear function using the first-order terms in a Taylor series expansion shown in below:

$$\begin{aligned} h(\boldsymbol{x}_{k-1},\mathbf{u}_k,\boldsymbol{w}_k) &\approx h(\hat{\boldsymbol{x}}_{k-1},\hat{\boldsymbol{w}}_k,\mathbf{u}_k)+h_{\boldsymbol{x}_{k-1}}(\hat{\boldsymbol{x}}_{k-1},\hat{\boldsymbol{w}}_k,\mathbf{u}_k) \\ &\quad (\boldsymbol{x}_{k-1}-\hat{\boldsymbol{x}}_{k-1})+h_{\boldsymbol{w}_k}(\hat{\boldsymbol{x}}_{k-1},\hat{\boldsymbol{w}}_k,\mathbf{u}_k)(\boldsymbol{w}_k-\hat{\boldsymbol{w}}_k), \end{aligned} \tag{2a}$$
$$\begin{aligned} g(\boldsymbol{x}_k,\boldsymbol{v}_k) &\approx g(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{v}}_k)+g_{\boldsymbol{x}_k}(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{v}}_k)(\boldsymbol{x}_k-\hat{\boldsymbol{x}}_k) \\ &\quad + g_{\boldsymbol{v}_k}(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{v}}_k)(\boldsymbol{v}_k-\hat{\boldsymbol{v}}_k), \end{aligned} \tag{2b}$$

where $h_{\boldsymbol{x}_{k-1}}(\hat{\boldsymbol{x}}_{k-1},\hat{\boldsymbol{w}}_k,\mathbf{u}_k)$ and $h_{\boldsymbol{w}_k}(\hat{\boldsymbol{x}}_{k-1},\hat{\boldsymbol{w}}_k,\mathbf{u}_k)$ donate the partial derivation with respect to (w.r.t.) $\boldsymbol{x}_{k-1}$ and $\boldsymbol{w}_k$ respectively at point $(\hat{\boldsymbol{x}}_{k-1},\hat{\boldsymbol{w}}_k,\mathbf{u}_k)$; $g_{\boldsymbol{x}_k}(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{v}}_k)$ and $g_{\boldsymbol{v}_k}(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{v}}_k)$ donate the partial derivation w.r.t. $\boldsymbol{x}_k$ and $\boldsymbol{v}_k$ respectively at point $(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{v}}_k)$. Neglecting Taylor expansion terms of second and higher orders in (2), we can havesimplifies to:

$$\boldsymbol{x}_k = \boldsymbol{H}_k\boldsymbol{x}_{k-1}+\boldsymbol{F}_k\boldsymbol{w}_k+\boldsymbol{h}_k, \tag{3a}$$
$$\boldsymbol{y}_k = \boldsymbol{G}_k\boldsymbol{x}_k+\boldsymbol{B}_k\boldsymbol{v}_k+\mathbf{g}_k, \tag{3b}$$

where the matrices $\boldsymbol{H}_k\in\mathbb{R}^{N\times N}$, $\boldsymbol{F}_k\in\mathbb{R}^{N\times O}$, $\boldsymbol{G}_k\in\mathbb{R}^{M\times N}$ and $\boldsymbol{B}_k\in\mathbb{R}^{M\times C}$, and vectors $\boldsymbol{h}_k\in\mathbb{R}^{N\times 1}$ and $\mathbf{g}_k\in\mathbb{R}^{M\times 1}$ can be expressed as

$$\boldsymbol{H}_k = h_{\boldsymbol{x}_{k-1}}(\hat{\boldsymbol{x}}_{k-1},\hat{\boldsymbol{w}}_k,\mathbf{u}_k),\quad \boldsymbol{F}_k = h_{\boldsymbol{w}_k}(\hat{\boldsymbol{x}}_{k-1},\hat{\boldsymbol{w}}_k,\mathbf{u}_k), \tag{4a}$$
$$\boldsymbol{G}_k = g_{\boldsymbol{x}_k}(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{v}}_k),\quad \boldsymbol{B}_k = g_{\boldsymbol{v}_k}(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{v}}_k), \tag{4b}$$
$$\boldsymbol{h}_k = h(\hat{\boldsymbol{x}}_{k-1},\hat{\boldsymbol{w}}_k,\mathbf{u}_k)-\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k-1}-\boldsymbol{F}_k\hat{\boldsymbol{w}}_k, \tag{4c}$$
$$\mathbf{g}_k = g(\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{v}}_k)-\boldsymbol{G}_k\hat{\boldsymbol{x}}_k-\boldsymbol{B}_k\hat{\boldsymbol{v}}_k. \tag{4d}$$

In accordance with the HMC, at each time step k, for brevity, we omit writing $\boldsymbol{y}_{1:k-1}$ in the probability density function (pdf) in the rest of paper, such as $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$ denoted as $p(\boldsymbol{x}_{k-1})$. Here, the estimated pdf of the previous state $\boldsymbol{x}_{k-1}$ is represented as $p(\boldsymbol{x}_{k-1})=\mathcal{N}(\boldsymbol{x}_{k-1}|\hat{\boldsymbol{x}}_{k-1},\boldsymbol{P}_{k-1})$. For the prediction phase, leveraging Bayes's rule, the predictive pdf of $\boldsymbol{x}_k$ given $\boldsymbol{w}_k$ can be deduced as follows:

$$\begin{aligned} &p(\boldsymbol{x}_k|\boldsymbol{w}_k) = \textstyle\int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1},\boldsymbol{w}_k)p(\boldsymbol{x}_{k-1})\mathrm{d}\boldsymbol{x}_{k-1} \\ &= \textstyle\int \delta(\boldsymbol{H}_k\boldsymbol{x}_{k-1}+\boldsymbol{F}_k\boldsymbol{w}_k+\boldsymbol{h}_k-\boldsymbol{x}_k)\mathcal{N}(\boldsymbol{x}_{k-1};\hat{\boldsymbol{x}}_{k-1},\boldsymbol{P}_{k-1})\mathrm{d}\boldsymbol{x}_{k-1} \\ &= \mathcal{N}(\boldsymbol{x}_k;\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k-1}+\boldsymbol{F}_k\boldsymbol{w}_k+\boldsymbol{h}_k,\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^T). \end{aligned} \tag{5}$$

Combining this with the prior of $\boldsymbol{w}_k$, the predictive prior of $\boldsymbol{x}_k$ can be articulated as:

$$p(\boldsymbol{x}_k) = \textstyle\int p(\boldsymbol{x}_k|\boldsymbol{w}_k)\prod_{i=1}^{L}p(w_{ki})\mathrm{d}\boldsymbol{w}_k = \int p(\boldsymbol{x}_k,\boldsymbol{w}_k)\mathrm{d}\boldsymbol{w}_k. \tag{6}$$

In the measurement phase, the likelihood pdf $p(\boldsymbol{y}_k|\boldsymbol{x}_k)$ can be computed as:

$$\begin{aligned} p(\boldsymbol{y}_k|\boldsymbol{x}_k) &= \textstyle\int p(\boldsymbol{y}_k|\boldsymbol{x}_k,\boldsymbol{v}_k)p(\boldsymbol{v}_k)\mathrm{d}\boldsymbol{v}_k \\ &= \textstyle\int \delta(\boldsymbol{y}_k-\boldsymbol{G}_k\boldsymbol{x}_k-\boldsymbol{B}_k\boldsymbol{v}_k-\mathbf{g}_k)\prod_{j=1}^{C}p(v_{kj})\mathrm{d}\boldsymbol{v}_k. \end{aligned} \tag{7}$$

With (6) and (7), in the update step, the posterior of $\boldsymbol{x}_k$ given $\boldsymbol{y}_k$ can be calculated as:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_k) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k)}{\int p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k} \tag{8a}$$
$$= \frac{\iint p(\boldsymbol{y}_k|\boldsymbol{v}_k,\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{w}_k)p(\boldsymbol{w}_k)p(\boldsymbol{v}_k)\mathrm{d}\boldsymbol{v}_k\mathrm{d}\boldsymbol{w}_k}{\iiint p(\boldsymbol{y}_k|\boldsymbol{v}_k,\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{w}_k)p(\boldsymbol{w}_k)p(\boldsymbol{v}_k)\mathrm{d}\boldsymbol{v}_k\mathrm{d}\boldsymbol{w}_k\mathrm{d}\boldsymbol{x}_k} \tag{8b}$$
$$= \frac{\iint p(\boldsymbol{y}_k,\boldsymbol{x}_k,\boldsymbol{w}_k,\boldsymbol{v}_k)\mathrm{d}\boldsymbol{w}_k\mathrm{d}\boldsymbol{v}_k}{\iiint p(\boldsymbol{y}_k,\boldsymbol{x}_k,\boldsymbol{w}_k,\boldsymbol{v}_k)\mathrm{d}\boldsymbol{w}_k\mathrm{d}\boldsymbol{v}_k\mathrm{d}\boldsymbol{x}_k}. \tag{8c}$$

When both measurement and process noises follow Gaussian distributions, (8a) remains Gaussian, facilitating straightforward calculation and leading to the Kalman filter. However, if either the measurement noise $\boldsymbol{v}_k$ or process noise $\boldsymbol{w}_k$ deviates from Gaussianity, exhibiting non-Gaussian distributions like Gaussian mixture model (GMM) distributions, the integration in (6) and (7) might become intractable or involve costly computational complexity. Therefore, we introduce the ReGVAMP-EKF algorithm in the next section.

III. Derivation of the ReGVAMP-EKF

This iterative approach reduces errors introduced by linearization, thereby improving filtering performance, especially for highly nonlinear systems. In the ReGVAMP-EKF algorithm framework, the primary objective is to approximate $p(\boldsymbol{x}_k,\boldsymbol{y}_k,\boldsymbol{w}_k,\boldsymbol{v}_k)$ in (8c) using a Gaussian pdf $q(\boldsymbol{x}_k,\boldsymbol{y}'_k,\boldsymbol{w}_k,\boldsymbol{v}_k)$ with $\boldsymbol{y}'_k=\boldsymbol{y}_k-\mathbf{g}_k$ such that:

$$\begin{aligned} &p(\boldsymbol{x}_k,\boldsymbol{y}_k,\boldsymbol{w}_k,\boldsymbol{v}_k)\approx q(\boldsymbol{x}_k,\boldsymbol{y}'_k,\boldsymbol{w}_k,\boldsymbol{v}_k) \\ &= p(\boldsymbol{x}_k|\boldsymbol{w}_k)\delta(\boldsymbol{y}'_k-\boldsymbol{G}_k\boldsymbol{x}_k-\boldsymbol{B}_k\boldsymbol{v}_k)\textstyle\prod_{i=1}^{L}f(w_{ki})\prod_{j=1}^{C}f(v_{kj}). \end{aligned} \tag{9}$$

All $f(w_{ki})$ and $f(v_{kj})$ in (9) are assumed to be Gaussian factors, expressed as:

$$f(w_{ki})=\mathcal{N}(w_{ki};\hat{w}_{ki},\tau_{w_{ki}}),\, f(\boldsymbol{w}_k)=\mathcal{N}(\boldsymbol{w}_k;\hat{\boldsymbol{w}}_k,\boldsymbol{T}_{\boldsymbol{w}_k}); \tag{10a}$$
$$f(v_{kj})=\mathcal{N}(v_{kj};\hat{v}_{kj},\tau_{v_{kj}}),\, f(\boldsymbol{v}_k)=\mathcal{N}(\boldsymbol{v}_k;\hat{\boldsymbol{v}}_k,\boldsymbol{T}_{\boldsymbol{v}_k}), \tag{10b}$$

where the ith entry of $\hat{\boldsymbol{w}}_k\in\mathbb{R}^L$ and the jth entry of $\hat{\boldsymbol{v}}_k\in\mathbb{R}^C$ are denoted as $\hat{w}_{ki}$ and $\hat{v}_{kj}$ respectively. Moreover, $\boldsymbol{T}_{\boldsymbol{w}_k}\in\mathbb{R}^{L\times L}$ and $\boldsymbol{T}_{\boldsymbol{v}_k}\in\mathbb{R}^{C\times C}$ are diagonal matrices with the ith and jth diagonal entries represented as $\tau_{w_{ki}}$ and $\tau_{v_{kj}}$ respectively.

The ReGVAMP-EKF algorithm for each time index, detailed in Algorithm 1, will be explained as below.

Algorithm 1 ReGVAMP-EKF Algorithm at time index k

Input: $p(\boldsymbol{x}_{k-1})$ and $\boldsymbol{y}_k$

Initialization: All $f(w_{ki}), f(v_{kj})$, $\boldsymbol{H}_k$, $\boldsymbol{F}_k$, $\boldsymbol{G}_k$, $\boldsymbol{B}_k$, $\boldsymbol{h}_k$, g_k and $\boldsymbol{y}'_k$

repeat

  Initilization of index $i \leftarrow 0$ and $j \leftarrow 0$

  Update $q(\boldsymbol{y}'_k|\boldsymbol{w}_k)$ in (11)

  repeat

    $i \leftarrow i+1$

    Update extrinsic $m(w_{ki})$ of w_{ki} in (15)

    Update approximated posterior $\hat{b}_i(w_{ki})$ of w_{ki} via KL divergence in (18)

    Update $f(w_{ki})$ in (20)

  until $i = L$

  Update $q(\boldsymbol{y}'_k|\boldsymbol{v}_k)$ in (22)

  repeat

    $j \leftarrow j+1$

    Update extrinsic $m(v_{kj})$ of v_{kj} in (26)

    Update approximated posterior $\hat{b}_j(v_{kj})$ of v_{kj} via KL divergence in (29)

    Update $f(v_{kj})$ in (31)

  until $j = C$

until All $f(w_{ki})$ and $f(z_{kj})$ converge or maximum iteration times

Calculate approximated posterior $q(\boldsymbol{x}_k|\boldsymbol{y}'_k)$ in (34)

Output: Treat $q(\boldsymbol{x}_k|\boldsymbol{y}'_k)$ as a prior $p(\boldsymbol{x}_k)$ of $\boldsymbol{x}_k$ for $k+1$ index

A. Initialization

If we know the prior distributions of the noises, we can initialize the parameters in (10a) and (10b) with the actual mean and variance of each process noise $p(w_{ki})$ and measurement noise $p(v_{kj})$ respectively. However, when all parameters are unknown, combining ReGVAMP-EKF with Expectation Maximization (EM) [13] allows for a feasible solution. In such cases, initializing all means to 0 and all variances to 1 serves as a pragmatic approach. Nonetheless, addressing this scenario is considered beyond the current scope of our contribution. For the Taylor expansion, we choose the expanding point around the posterior mean of the process noise, measurement noise, and previous state mean. For initialization, we use the prior means.

B. Update $q(\boldsymbol{y}'_k|\boldsymbol{w}_k)$

With initialized $f(z_{kj})$, according to Bayes' rule, the approximated likelihood $q(\boldsymbol{y}'_k|\boldsymbol{w}_k)$ can be expressed as:

$$\begin{aligned} q(\boldsymbol{y}'_k|\boldsymbol{w}_k) &= \iint p(\boldsymbol{x}_k|\boldsymbol{w}_k)\delta(\boldsymbol{y}'_k - \boldsymbol{G}_k\boldsymbol{x}_k - \boldsymbol{B}_k\boldsymbol{v}_k) f(\boldsymbol{v}_k)\mathrm{d}\boldsymbol{x}_k \\ &= \mathcal{N}(\boldsymbol{y}'_k; \boldsymbol{G}_k\boldsymbol{F}_k\boldsymbol{w}_k + \boldsymbol{G}_k\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{G}_k\boldsymbol{h}_k + \boldsymbol{B}_k\hat{\boldsymbol{v}}_k, \\ &\qquad \boldsymbol{G}_k\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^T\boldsymbol{G}_k^T + \boldsymbol{B}_k\boldsymbol{T}_{\boldsymbol{v}_k}\boldsymbol{B}_k^T). \end{aligned} \tag{11}$$

C. Update all $f(w_{ki})$

1) Update extrinsic $m(w_{ki})$ of w_{ki}: Since both $q(\boldsymbol{w}_k)$ and $f(w_{kn})$ are Gaussian pdfs, we can define an approximate posterior $\tilde{q}(\boldsymbol{w}_k)$ of $\boldsymbol{w}_k$ as follows:

$$\tilde{q}(\boldsymbol{w}_k) = \mathcal{N}(\boldsymbol{w}_k; \mathbf{m}_{\boldsymbol{w}_k}, \mathbf{C}_{\boldsymbol{w}_k}) = \frac{q(\boldsymbol{y}'_k|\boldsymbol{w}_k)\prod_{i=1}^{L} f(w_{ki})}{\int q(\boldsymbol{y}'_k|\boldsymbol{w}_k)\prod_{i=1}^{L} f(w_{ki})\mathrm{d}\boldsymbol{w}_k}, \tag{12}$$

where

$$\mathbf{C}_{\boldsymbol{w}_k} = (\boldsymbol{T}_{\boldsymbol{w}_k}^{-1} + \boldsymbol{F}_k^T\boldsymbol{G}_k^T(\boldsymbol{G}_k\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^T\boldsymbol{G}_k^T + \boldsymbol{B}_k\boldsymbol{T}_{\boldsymbol{v}_k}\boldsymbol{B}_k^T)^{-1}\boldsymbol{G}_k\boldsymbol{F}_k)^{-1}; \tag{13}$$

$$\begin{aligned} \mathbf{m}_{\boldsymbol{w}_k} = \mathbf{C}_{\boldsymbol{w}_k}[&\boldsymbol{G}_k^T\boldsymbol{H}_k^T(\boldsymbol{T}_{\mathbf{z}_k} + \boldsymbol{F}_k^T\boldsymbol{G}_k^T(\boldsymbol{G}_k\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^T\boldsymbol{G}_k^T + \boldsymbol{B}_k\boldsymbol{T}_{\boldsymbol{v}_k}\boldsymbol{B}_k^T)^{-1} \\ &(\boldsymbol{y}'_k - (\boldsymbol{G}_k\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{G}_k\boldsymbol{h}_k + \boldsymbol{B}_k\hat{\boldsymbol{v}}_k)) + \boldsymbol{T}_{\boldsymbol{w}_k}^{-1}\hat{\boldsymbol{w}}_k]. \end{aligned} \tag{14}$$

Removing $f(w_{ki})$, the extrinsic $m(w_{ki})$ can be calculated as:

$$m(w_{ki}) = \frac{\int \tilde{q}(\boldsymbol{w}_k)\mathrm{d}\boldsymbol{w}_{k\setminus i}/f(w_{ki})}{\iint \tilde{q}(\boldsymbol{w}_k)\mathrm{d}\boldsymbol{w}_{k\setminus i}/f(w_{ki})\mathrm{d}w_{ki}} = \mathcal{N}(w_{ki}; \mu_{w_{ki}}, \xi_{w_{ki}}), \tag{15}$$

where

$$\begin{aligned} \xi_{w_{ki}} &= \frac{\tau_{w_{ki}}\boldsymbol{e}_i^T\mathbf{C}_{\boldsymbol{w}_k}\boldsymbol{e}_i}{\tau_{w_{ki}} - \boldsymbol{e}_i^T\mathbf{C}_{\boldsymbol{w}_k}\boldsymbol{e}_i}; \\ \mu_{w_{ki}} &= \xi_{w_{ki}}\left(\frac{\boldsymbol{e}_i^T\mathbf{m}_{\boldsymbol{w}_k}}{\boldsymbol{e}_i^T\mathbf{C}_{\boldsymbol{w}_k}\boldsymbol{e}_i} - \frac{\hat{w}_{ki}}{\tau_{w_{ki}}}\right) \end{aligned} \tag{16}$$

2) Update approximated posterior (belief) $\hat{b}_i(w_{ki})$ of w_{ki} via KL divergence: Firstly, defining $\hat{b}_i(w_{ki})$ as a Gaussian pdf:

$$\hat{b}_i(w_{ki}) = \mathcal{N}(w_{ki}; \alpha_{w_{ki}}, \beta_{w_{ki}}), \tag{17}$$

then combining the extrinsic $m(w_{ki})$ from (15) with the actual prior $p(w_{ki})$ and minimizing the Kullback–Leibler (KL) divergence to obtain an approximate marginal posterior (belief) $\hat{b}_i(w_{ki})$:

$$\hat{b}_i(w_{ki}) = \arg\min_{b_i(w_{ki})} D_{KL}\left[\frac{m(w_{ki})p(w_{ki})}{\int m(w_{ki})p(w_{ki})\mathrm{d}w_{ki}}\middle\| b_i(w_{ki})\right]. \tag{18}$$

With some mathematical algebra, the solution to (18) can be provided as:

$$\alpha_{w_{ki}} = \frac{\int w_{ki}m(w_{ki})p(w_{ki})\mathrm{d}w_{ki}}{\int m(w_{ki})p(w_{ki})\mathrm{d}w_{ki}} = g_{w_{ki}}(\mu_{w_{ki}}, \xi_{w_{ki}}); \tag{19a}$$

$$\beta_{w_{ki}} = \frac{\int(\alpha_{w_{ki}} - w_{ki})^2 m(w_{ki})p(w_{ki})\mathrm{d}w_{ki}}{\int m(w_{ki})p(w_{ki})\mathrm{d}w_{ki}} = \xi_{w_{ki}}\frac{\partial g_{w_{ki}}(\mu_{w_{ki}}, \xi_{w_{ki}})}{\partial \mu_{w_{ki}}}. \tag{19b}$$

3) Update $f(w_{ki})$: Once we obtain $\hat{b}_i(w_{ki})$, removing the extrinsic $m(w_{ki})$, we can update $f(w_{ki})$ as:

$$f(w_{ki}) = \frac{\hat{b}_i(w_{ki})/m(w_{ki})}{\int \hat{b}_i(w_{ki})/m(w_{ki})dw_{ki}} = \mathcal{N}(w_{ki}; \hat{w}_{ki}, \tau_{w_{ki}}), \tag{20}$$

where the updated variance $\tau_{w_{ki}}$ and updated mean $\hat{w}_{ki}$ can be expressed as:

$$\begin{aligned} \tau_{w_{ki}} &= \frac{\xi_{w_{ki}}\beta_{w_{ki}}}{\xi_{w_{ki}} - \beta_{w_{ki}}}, \\ \hat{w}_{ki} &= \tau_{w_{ki}}\left(\frac{\beta_{w_{ki}}}{\alpha_{w_{ki}}} - \frac{\mu_{w_{ki}}}{\xi_{w_{ki}}}\right). \end{aligned} \tag{21}$$

D. Update $q(\boldsymbol{y}'_k|\boldsymbol{v}_k)$

Similar to III-B, with $f(\boldsymbol{w}_k)$, the approximated likelihood $q(\boldsymbol{y}'_k|\boldsymbol{v}_k)$ can be expressed as:

$$\begin{aligned} q(\boldsymbol{y}'_k|\boldsymbol{v}_k) &= \iint p(\boldsymbol{x}_k|\boldsymbol{w}_k)\delta(\boldsymbol{y}'_k - \boldsymbol{G}_k\boldsymbol{x}_k - \boldsymbol{B}_k\boldsymbol{v}_k)f(\boldsymbol{w}_k)\mathrm{d}\boldsymbol{x}_k\mathrm{d}\boldsymbol{w}_k \\ &= \mathcal{N}(\boldsymbol{y}'_k; \boldsymbol{B}_k\boldsymbol{v}_k + \boldsymbol{G}_k\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{G}_k\boldsymbol{F}_k\boldsymbol{w}_k + \boldsymbol{G}_k\boldsymbol{h}_k, \\ &\qquad \boldsymbol{G}_k\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^T\boldsymbol{G}_k^T + \boldsymbol{G}_k\boldsymbol{F}_k\boldsymbol{T}_{\boldsymbol{w}_k}\boldsymbol{F}_k^T\boldsymbol{G}_k^T). \end{aligned} \tag{22}$$

E. Update all $f(v_{kj})$

1) Update extrinsic $m(v_{kj})$ of v_{kj}: Since both $q(\boldsymbol{y}'_k|\boldsymbol{v}_k)$ and $f(v_{kj})$ are Gaussian pdfs, we can define an approximated posterior $\tilde{q}(\boldsymbol{v}_k)$ of $\boldsymbol{v}_k$ as follows:

$$\tilde{q}(\boldsymbol{v}_k) = \mathcal{N}(\boldsymbol{v}_k; \mathbf{m}_{\boldsymbol{v}_k}, \mathbf{C}_{\boldsymbol{v}_k}) = \frac{q(\boldsymbol{y}'_k|\boldsymbol{v}_k)\prod_{j=1}^{C} f(v_{kj})}{\int q(\boldsymbol{y}'_k|\boldsymbol{v}_k)\prod_{j=1}^{C} f(v_{kj})\mathrm{d}\boldsymbol{v}_k}, \tag{23}$$

where

$$\begin{aligned} \mathbf{C}_{\boldsymbol{v}_k} = (\boldsymbol{T}_{\boldsymbol{v}_k}^{-1} + \boldsymbol{B}_k^T(&\boldsymbol{G}_k\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^T\boldsymbol{G}_k^T \\ &+ \boldsymbol{G}_k\boldsymbol{F}_k\boldsymbol{T}_{\boldsymbol{w}_k}\boldsymbol{F}_k^T\boldsymbol{G}_k^T)^{-1}\boldsymbol{B}_k)^{-1}; \end{aligned} \tag{24}$$

$$\mathbf{m}_{\boldsymbol{v}_k} = \mathbf{C}_{\boldsymbol{v}_k}[\boldsymbol{B}_k^T(\boldsymbol{G}_k\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^T\boldsymbol{G}_k^T\boldsymbol{G}_k\boldsymbol{F}_k\boldsymbol{T}_{\boldsymbol{w}_k}\boldsymbol{F}_k^T\boldsymbol{G}_k^T)^{-1} (\boldsymbol{y}_k' - (\boldsymbol{G}_k\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{G}_k\boldsymbol{F}_k\boldsymbol{w}_k + \boldsymbol{G}_k\boldsymbol{h}_k)) + \boldsymbol{T}_{\boldsymbol{v}_k}^{-1}\hat{\boldsymbol{v}}_k]. \tag{25}$$

Removing $f(v_{kj})$, the extrinsic $m(v_{kj})$ can be calculated as follows:

$$m(v_{kj}) = \frac{\int \tilde{q}(\boldsymbol{v}_k)\mathrm{d}\boldsymbol{v}_{k\setminus j}/f(v_{kj})}{\iint \tilde{q}(\boldsymbol{v}_k)\mathrm{d}\boldsymbol{v}_{k\setminus j}/f(v_{kj})\mathrm{d}v_{kj}} = \mathcal{N}(v_{kj}; \mu_{v_{kj}}, \xi_{v_{kj}}), \tag{26}$$

where

$$\xi_{v_{kj}} = \frac{\tau_{v_{kj}}\boldsymbol{e}_j^T\mathbf{C}_{\boldsymbol{v}_k}\boldsymbol{e}_j}{\tau_{v_{kj}} - \boldsymbol{e}_j^T\mathbf{C}_{\boldsymbol{v}_k}\boldsymbol{e}_j}; \tag{27a}$$

$$\mu_{v_{kj}} = \xi_{v_{kj}}\left(\frac{\boldsymbol{e}_j^T\mathbf{m}_{\boldsymbol{v}_k}}{\boldsymbol{e}_j^T\mathbf{C}_{\boldsymbol{v}_k}\boldsymbol{e}_j} - \frac{\hat{v}_{kj}}{\tau_{v_{kj}}}\right). \tag{27b}$$

2) Update approximated posterior (belief) $\hat{b}_j(v_{kj})$ of v_{kj} via KL divergence: Defining $\hat{b}_j(v_{kj})$ as a Gaussian pdf:

$$\hat{b}_j(v_{kj}) = \mathcal{N}(v_{kj}; \alpha_{v_{kj}}, \beta_{v_{kj}}). \tag{28}$$

We combine the extrinsic $m(v_{kj})$ with the real prior $p(v_{kj})$ and minimize the KL divergence to obtain an approximate marginal posterior (belief) $\hat{b}_j(v_{kj})$.

$$\hat{b}_j(v_{kj}) = \arg\min_{b_j(v_{kj})} D_{KL}\left[\frac{m(v_{kj})p(v_{kj})}{\int m(v_{kj})p(v_{kj})\mathrm{d}v_{kj}}\middle\| b_j(v_{kj})\right]. \tag{29}$$

Following the approach used for (18), we can derive the solution for (29) as:

$$\alpha_{v_{kj}} = \frac{\int v_{kj}m(v_{kj})p(v_{kj})\mathrm{d}v_{kj}}{\int m(v_{kj})p(v_{kj})\mathrm{d}v_{kj}} = g_{v_{kj}}(\mu_{v_{kj}}, \xi_{v_{kj}}); \tag{30a}$$

$$\beta_{v_{kj}} = \frac{\int(\alpha_{v_{kj}} - v_{kj})^2 m(v_{kj})p(v_{kj})\mathrm{d}z_{kj}}{\int m(v_{kj})p(v_{kj})\mathrm{d}v_{kj}} = \xi_{v_{kj}}\frac{\partial g_{v_{kj}}(\mu_{v_{kj}}, \xi_{v_{kj}})}{\partial \mu_{v_{kj}}}. \tag{30b}$$

3) Update $f(v_{kj})$: Removing the extrinsic $m(v_{kj})$ from belief $\hat{b}_j(v_{kj})$, the approximated factor $f(v_{kj})$ can be updated as

$$f(v_{kj}) = \frac{\hat{b}_j(v_{kj})/m(v_{kj})}{\int \hat{b}_j(v_{kj})/m(v_{kj})\mathrm{d}v_{kj}} = \mathcal{N}(v_{kj}; \hat{v}_{kj}, \tau_{v_{kj}}), \tag{31}$$

where the updated variance $\tau_{v_{kj}}$ and mean $\hat{v}_{kj}$ can be expressed as

$$\tau_{v_{kj}} = \frac{\xi_{v_{kj}}\beta_{v_{kj}}}{\xi_{v_{kj}} - \beta_{v_{kj}}}, \tag{32a}$$

$$\hat{v}_{kj} = \tau_{v_{kj}}\left(\frac{\beta_{v_{kj}}}{\alpha_{v_{kj}}} - \frac{\mu_{v_{kj}}}{\xi_{v_{kj}}}\right). \tag{32b}$$

F. Convergence judgment criteria

Convergence is not assured with the ReGVAMP. In the majority of cases, however, convergence is assured with the ReGVAMP algorithm within simulations. As for the convergence judgment criteria, we choose:

$$\textstyle\sum_{i=1}^{L}\|\hat{w}_{ki}^{new} - \hat{w}_{ki}\|_2^2 + \sum_{j=1}^{C}\|\hat{v}_{kj}^{new} - \hat{v}_{kj}\|_2^2 < \epsilon, \tag{33}$$

where $\hat{w}_{ki}^{new}$ and $\hat{v}_{kj}^{new}$ indicate the updated mean of w_{ki} and v_{kj}, respectively. ϵ is set to be 10^{-3} in our numerical simulation, and the maximum iteration time is set to be 10.

G. Calculate approximated posterior $q(\boldsymbol{x}_k|\boldsymbol{y}_k')$

With the approximated $f(\boldsymbol{w}_k)$ and $f(\mathbf{z}_k)$, $q(\boldsymbol{x}_k|\boldsymbol{y}_k')$ can be calculated as follows:

$$q(\boldsymbol{x}_k|\boldsymbol{y}_k') = \frac{\iint p(\boldsymbol{x}_k|\boldsymbol{w}_k)\delta(\boldsymbol{y}_k' - \boldsymbol{G}_k\boldsymbol{x}_k - \boldsymbol{B}_k\boldsymbol{v}_k)f(\boldsymbol{w}_k)f(\boldsymbol{v}_k)\mathrm{d}\boldsymbol{v}_k\mathrm{d}\boldsymbol{w}_k}{\iiint p(\boldsymbol{x}_k|\boldsymbol{w}_k)\delta(\boldsymbol{y}_k' - \boldsymbol{G}_k\boldsymbol{x}_k - \boldsymbol{B}_k\boldsymbol{v}_k)f(\boldsymbol{w}_k)f(\boldsymbol{v}_k)\mathrm{d}\boldsymbol{v}_k\mathrm{d}\boldsymbol{w}_k\mathrm{d}\boldsymbol{x}_k} = \mathcal{N}(\boldsymbol{x}_k|\hat{\boldsymbol{x}}_k, \boldsymbol{P}_k), \tag{34}$$

where

$$\boldsymbol{P}_k = [\boldsymbol{G}_k^T(\boldsymbol{B}_k\boldsymbol{T}_{\boldsymbol{v}_k}\boldsymbol{B}_k^T)^{-1}\boldsymbol{G}_k + (\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^T + \boldsymbol{F}_k\boldsymbol{T}_{\boldsymbol{w}_k}\boldsymbol{F}_k^T)^{-1}]^{-1}. \tag{35a}$$

$$\hat{\boldsymbol{x}}_k = \boldsymbol{P}_k[\boldsymbol{G}_k^T(\boldsymbol{B}_k\boldsymbol{T}_{\boldsymbol{v}_k}\boldsymbol{B}_k^T)^{-1}(\boldsymbol{y}_k' - \boldsymbol{B}_k\hat{\boldsymbol{v}}_k) + (\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^T + \boldsymbol{F}_k\boldsymbol{T}_{\boldsymbol{w}_k}\boldsymbol{F}_k^T)^{-1}(\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{F}_k\hat{\boldsymbol{w}}_k + \boldsymbol{h}_k)]. \tag{35b}$$

For the next time step $k+1$, $q(\boldsymbol{x}_k|\boldsymbol{y}_k')$ is set as the approximated pdf of time step $\boldsymbol{x}_k$.

IV. Derivation of the ReGVAMP-IEKF

Further enhancing filtering accuracy, the IEKF iteratively refines the state estimate. Unlike the EKF, which relies on a single linearization, the IEKF iterates between prediction and update steps, linearizing around the current best estimate at each iteration. The ReGVAMP-EKF is extended to the ReGVAMP-IEKF by incorporating iterative updates at each time step, where the state estimate is refined through repeated linearization around the updated estimate until convergence. This approach enhances the accuracy of state estimation in highly nonlinear systems. The ReGVAMP-IEKF can be found in Algorithm 2.

Algorithm 2 ReGVAMP-IEKF Algorithm at time index k

Input: $p(\boldsymbol{x}_{k-1})$ and $\boldsymbol{y}_k$

1: Initialization: All $f(w_{ki}), f(v_{kj})$, $\boldsymbol{H}_k$, $\boldsymbol{F}_k$, $\boldsymbol{G}_k$, $\boldsymbol{B}_k$, $\boldsymbol{h}_k$, $\mathbf{g}_k$ and $\boldsymbol{y}_k'$
2: **repeat**
3: Update $f(w_{ki}), f(v_{kj})$ and $q(\boldsymbol{x}_k|\boldsymbol{y}_k')$ via ReGVAMP-EKF algorithm until converge.
4: Update $\boldsymbol{H}_k$, $\boldsymbol{F}_k$, $\boldsymbol{G}_k$, $\boldsymbol{B}_k$, $\boldsymbol{h}_k$, $\mathbf{g}_k$ and $\boldsymbol{y}_k'$
5: **until** Until $f(w_{ki}), f(v_{kj}), q(\boldsymbol{x}_k|\boldsymbol{y}_k')$ converged
6: Output the approximate posterior $q(\boldsymbol{x}_k|\boldsymbol{y}_k')$

Output: Treat $q(\boldsymbol{x}_k|\boldsymbol{y}_k')$ as a prior $p(\boldsymbol{x}_k)$ of $\boldsymbol{x}_k$ for $k+1$ index

V. Numerical Simulation

We have designed a fully nonlinear state-space model to describe the motion trajectory of an object. The state variables include position p_x, p_y, and velocity v_x, v_y. Both the state transition and observation equations are nonlinear. The specific equations are as follows: The state vector $\boldsymbol{x}_k = [p_{x,k}, p_{y,k}, v_{x,k}, v_{y,k}]^T$ represents the combination of position and velocity. The nonlinear state transition equation is:

$$\boldsymbol{x}_k = \begin{bmatrix} p_{x,k-1} + T\cdot\sin(v_{x,k-1}) \\ p_{y,k-1} + T\cdot\cos(v_{y,k-1}) \\ v_{x,k-1} + T\cdot\sin(p_{x,k-1}) \\ v_{y,k-1} + T\cdot\cos(p_{y,k-1}) \end{bmatrix} + \mathbf{w}_k \tag{36}$$

where T is the sampling time which is set to be 0.05 in our simulation and $\boldsymbol{w}_k$ represents process noise. The initialized state $\boldsymbol{x}_0$ is set to in Gaussian distribution with means $[1\ 1\ 0.1\ 0.1]^T$ and covariance matrix $\mathbf{I}_4$. The observation vector consists of the measured distance and angle, modeled by the following nonlinear observation equation:

$$\boldsymbol{y}_k = \begin{bmatrix} \sqrt{p_{x,k}^2 + p_{y,k}^2} \\ \arctan\left(\frac{p_{y,k}}{p_{x,k}}\right) \end{bmatrix} + \boldsymbol{v}_k \tag{37}$$

where $\boldsymbol{v}_k$ represents observation noise. To simulate a complex noise environment, both process noise $\boldsymbol{w}_k$ and observation noise $\boldsymbol{v}_k$ are modeled as the GMM noise. The GMM is represented as a weighted sum of multiple Gaussian distributions, described as: Process Noise $\boldsymbol{w}_k$:

$$\boldsymbol{w}_k \sim 0.3\mathcal{N}(0, \mathbf{I}_4) + 0.7\mathcal{N}(0, 0.09\mathbf{I}_4); \tag{38a}$$

$$\boldsymbol{v}_k \sim 0.3\mathcal{N}(0, \mathbf{I}_2) + 0.7\mathcal{N}(0, 0.09\mathbf{I}_2). \tag{38b}$$

The *Averaged Cumulative Mean Error* (ACME) is a metric used to evaluate the performance of an estimated vector $\hat{\boldsymbol{x}}$ compared to its ground truth counterpart $\mathbf{x}$ over multiple time steps. The ACME of $\hat{x}_T(i)$is defined as:

$$\text{ACME}_{x_T(i)} = \frac{1}{T}\sum_{t=1}^{T} |x_t(i) - \hat{x}_t(i)|, \tag{39}$$

where T is the total number of time steps. In our simulation, we assume both of them are known during simulations. Fig. 1 illustrates the true positions, velocities, and their estimations and Fig 2 shows the ACME rsults. It can be seen that the performance of EKF and IEKF is improved by combining the ReGVAMP algorithm. Although the UKF demonstrates similar performance, it frequently crashes in our simulations due to arithmetic precision issues. The simulation codes can be found in https://github.com/FqXIAO/ReGVAMP-EKF-IEKF.git.

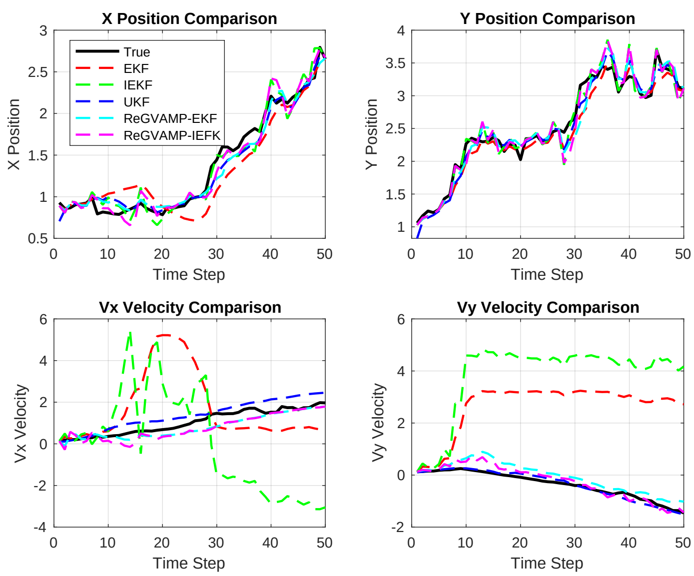

Fig. 1. Estimation Results of Each State

VI. Conclusion

This paper introduces the ReGVAMP-EKF filter to address the filtering problem in stochastic systems characterized by nonlinear system models and non-Gaussian noise. Unlike traditional methods that rely on MMSE, the ReGVAMP algorithm is utilized in this filter to approximate the posterior of each state vector $\boldsymbol{x}_k$ as Gaussian. This capability allows the filter to effectively handle non-Gaussian noise while maintaining computational complexity similar to that of the Extended Kalman Filter (EKF). Numerical examples demonstrate that the ReGVAMP-EKF and ReGVAMP-IEKF achieves high performance, especially compared to the EKF and IEKF.

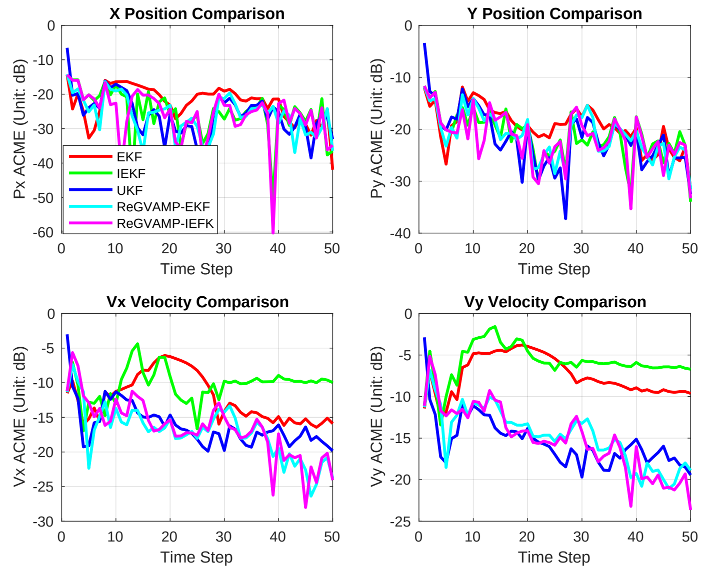

Fig. 2. Averaged Cumulative Mean Error Comparison

However, future research should include more comprehensive comparisons with other state-of-the-art methods. Additionally, further investigation is needed to analyze the performance and convergence of the ReGVAMP-EKF and ReGVAMP-IEFK, particularly in larger systems and with unknown hyperparameters, through extensive simulations.

Acknowledgements EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, Principauté de Monaco, and by the Franco-German projects 5G-OPERA and CellFree6G.

References

[1] R. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *J. Basic. Eng*, vol. 82, no. 1, pp. 35–45, 1960.

[2] K. Jahan and S. K. Rao, "Implementation of Underwater Target Tracking Techniques for Gaussian and non-Gaussian Environments," *Comput. Electr. Eng.*, vol. 87, p. 106783, 2020.

[3] R. Bass, V. Norum, and L. Schwartz, "Optimal Multichannel Nonlinear Filtering," *J. Math. Anal. Appl.*, vol. 16, no. 1, pp. 152–164, 1966.

[4] J. Havlík and O. Straka, "Performance evaluation of iterated extended Kalman filter with variable step-length," *J. Phys. Conf.*, vol. 659, no. 1, p. 012022, nov 2015.

[5] P. Rohal and J. Ochodnicky, "Radar Target Tracking by Kalman and Particle Filter," in *IEEE KIT*, 2017, pp. 1–4.

[6] M. Jiang, S. Guo, H. Luo, Y. Yao, and G. Cui, "A Robust Target Tracking Method for Crowded Indoor Environments Using mmWave Radar," *Remote Sens.*, vol. 15, no. 9, p. 2425, 2023.

[7] S. Rangan, "Generalized Approximate Message Passing for Estimation with Random Linear Mixing," in *IEEE ISIT*, 2011, pp. 2168–2172.

[8] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, 2019.

[9] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Am. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, 2017.

[10] Z. Zhao, F. Xiao, and D. Slock, "Vector Approximate Message Passing for Not So Large N.I.I.D. Generalized I/O Linear Models," in *IEEE ICASSP*, 2024, pp. 13 281–13 285.

[11] E. Wan and R. Van Der Merwe, "The Unscented Kalman Filter for Nonlinear Estimation," in *IEEE AS-SPCC*, 2000, pp. 153–158.

[12] S. R. Eddy, "Hidden Markov models," *Curr. Opin. Struct. Biol.*, vol. 6, no. 3, pp. 361–365, 1996.

[13] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, 1996.