Closed-Form Multi-Anchor Wasserstein Barycenters for High-Quality Assumed Gaussian Filtering

Jiachen Zhou and **Uwe D. Hanebeck**
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
jiachen.zhou@kit.edu, uwe.hanebeck@kit.edu

***Abstract*—In this paper, we propose a novel Gaussian Assumed Density Filter (GADF) for high-quality state estimation of nonlinear dynamic systems, with a focus on the measurement update. Recent approaches parameterize a Gaussian Process (GP) in the joint measurement/state space by assigning a set of Gaussian distributions to measurement-space positions and evaluating the state posterior via a Wasserstein barycentric interpolation between two neighboring Gaussians. This yields an efficient, non-Gaussian local approximation of the underlying joint measurement/state density. However, while effective for one-dimensional measurements, identifying the appropriate boundary pair becomes non-trivial beyond 1D. Building on this line of work, we continue to leverage deterministic samples from the joint measurement/state density, but instead parameterize a GP via a 2-Wasserstein barycentric fusion of multiple conditional Gaussian state densities, so that all anchor distributions contribute for every measurement update. By introducing a carefully designed parameterization of the anchor Gaussians, we derive a barycenter in closed form, substantially simplifying parameter estimation while maintaining accuracy and improving computational tractability for nonlinear filtering.**

***Index Terms*—Bayesian inference, nonlinear filtering, Gaussian Assumed Density Filter, Wasserstein distance, maximum likelihood estimation, Gaussian Processes.**

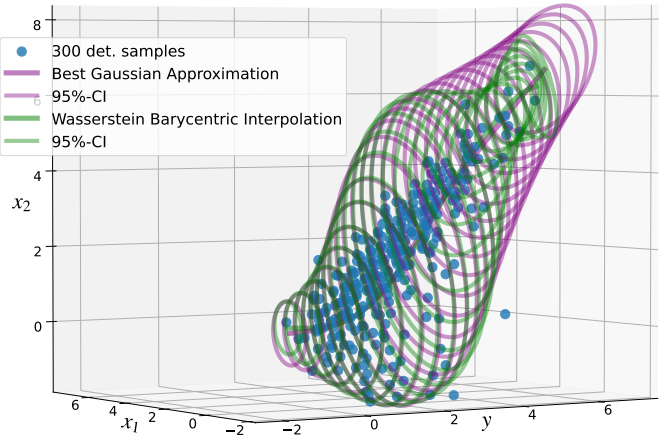


Fig. 1: Illustration of the GP-based local non-Gaussian approximation (green) of the true joint measurement/state density obtained via Wasserstein barycentric interpolation. For clarity, only a subset of the 2D conditional Gaussian state PDFs is shown. The GP is fitted to deterministic samples (blue dots) representing the underlying joint density. For comparison, the best Gaussian approximation of the true state posterior is shown (purple), with conditional means and covariances matching those of the true posterior. Ellipses denote the 95% confidence regions of the 2D conditional Gaussians.

I. Introduction

A. Context

We consider the filter step in the general state estimation problem for a discrete-time stochastic nonlinear dynamic system observed through noisy measurements. Closed-form recursive Bayesian filtering is available only for limited system classes, especially linear continuous-state systems with Gaussian noise. In that setting, the Kalman filter (KF) provides the optimal linear unbiased estimator, and the state estimate can be propagated analytically at each processing step. By contrast, in case of nonlinearities and/or non-Gaussian noise, exact recursive Bayesian updates are typically intractable. Consequently, practical inference must rely on approximations. In this work, we focus on Gaussian filters that approximate the true, in general complex, state Probability Density Function (PDF) by explicitly optimizing the shape of a Gaussian distribution after each processing step. This class of filters is known as Gaussian Assumed Density Filters (GADFs) [1], [2].

The key advantage of GADFs is the compact and constant amount of information to represent their state estimates. By propagating only the mean and covariance within each update step, they avoid the growing complexity compared to filters that operate directly on the more complex state densities [3]. Despite these simplifications, deriving a closed-form measurement update still remains difficult and often entails numerical integration and/or the evaluation of expectations. The Bayesian measurement update requires an explicit likelihood function. In many practical settings, however, deriving it is challenging, for example, when the measurement noise is neither purely additive nor scalar-multiplicative. Furthermore, even with a valid likelihood, the filter step is often analytically intractable.

B. State of the Art

The Gaussian Particle Filter (GPF) [4] is a Monte-Carlo-based GADF inspired by particle filtering. It uses importance sampling to estimate posterior moments, but re-approximates the posterior by a single Gaussian instead of propagating weighted particles. Hence, in contrast to conventional Particle Filters (PFs), its particles act only as temporary numerical integration points for moment matching. Although the GPF converges to the Minimum Mean Square Error (MMSE) estimator as the number of particles approaches infinity, its

computational cost scales poorly with state dimension and requires an explicit likelihood function for particle re-weighting.

Likelihood-free approaches avoid explicit likelihood construction by linearly approximating the nonlinear measurement equation and applying the Kalman update formulas, leading to Nonlinear Kalman Filters (NKFs). However, their accuracy depends strongly on the "strength" of nonlinearity, and severe nonlinearities can make this approximation overly restrictive compared with more general GADFs without such (explicit or implicit) linearization.

One approach to linearization is explicit linearization via Taylor series expansion, as employed by Extended Kalman Filters (EKFs) and its variants [5], [6]. In contrast, NKFs based on statistical linearization implicitly linearize the measurement model by approximating the joint density of measurement and state with a Gaussian, a technique commonly referred to as the second Gaussian assumption. When the first- and second-order moments of nonlinearly transformed densities are unavailable in closed form, they must be approximated from sample-based density representations. These NKFs fall under the category of Linear Regression Kalman Filters (LRKFs) [7], [8]. Notable examples include the Unscented Kalman Filter (UKF) [9] and the Smart Sampling Kalman Filter (S^2KF) [10].

LRKFs are computationally efficient and easy to implement. Unlike PFs and GPFs, they circumvent the issue of sample degeneration due to their likelihood-free nature. However, their main drawback is a further degradation in state-estimation quality in highly nonlinear systems, since continuous state and noise densities are represented by only a finite set of samples.

A more advanced GADF, the Progressive Gaussian Filter (PGF), eliminates the second Gaussian assumption, reducing linearization errors and enhancing estimation quality [11], [12]. It achieves this by decomposing the measurement update into multiple sub-updates, gradually integrating measurement information into state estimates. In [12], an explicit likelihood function is needed for the progression mechanism. Both PGF variants may accumulate errors due to multiple intermediate Gaussian approximations within each recursion step.

A closely related approach is the GADF with Inverse Gaussian Process (IGP) interpolation [13], [14]. Unlike conventional methods, it avoids both the second-Gaussian assumption and an explicit likelihood construction in the measurement update. Instead, it first generates deterministic samples from the true joint measurement/state density. Based on these samples, a Gaussian Process (GP) is then fitted in the joint measurement/state space to obtain a local approximation of the non-Gaussian joint density. Conditioning at any query point in the measurement space directly yields a Gaussian conditional state PDF, realizing efficient backward inference.

Despite strong filtering performance, often comparable to that of the optimal MMSE estimator [4], these methods face practical limitations in computational efficiency and scalability. Specifically, [13] requires two sequential matrix inversions within each measurement update, each with cubic complexity, which can lead to runtimes incompatible with real-time deployment. Moreover, both [13], [14] scale poorly with increasing state and measurement dimensionality, limiting their applicability to high-dimensional problems such as human motion tracking [15] and extended object tracking [16].

C. Contributions

In this work, we propose a GADF with a novel likelihood-free measurement update that avoids explicit likelihood evaluation. Unlike LRKFs, the new GADF does not rely on the second Gaussian assumption for the true joint measurement/state density. Instead, building upon the IGP-interpolation-based filtering introduced in [13], [14], we perform a non-Gaussian local approximation of the true joint density.

Furthermore, the proposed formulation addresses the main practical limitations of prior IGP-based methods, namely high computational cost and limited scalability in higher-dimensional state and measurement spaces. Compared with [14], we replace pairwise interpolation by a multi-anchor construction, where each anchor is a multivariate Gaussian conditional state PDF located in the measurement space. For any measurement value, the corresponding conditional state density is obtained as the unique 2-Wasserstein barycenter of all anchors, thereby defining a smooth continuum of conditional Gaussians over the measurement space and enabling efficient backward inference.

Unlike [14], where interpolation is based on a two-anchor subset, our barycenter is computed jointly from all anchors. This yields a globally consistent interpolation rule and naturally supports extrapolation. Moreover, in multi-dimensional measurement spaces, anchor locations are treated as learnable and unconstrained (i.e., non-grid, non-ordered) variables.

Subsequently, we generate high-quality deterministic samples from the true joint measurement/state density. Based on these samples, we fit the parameters of the proposed conditional Gaussian model by minimizing the negative conditional log-density objective. The learned model then provides a continuum of Gaussian conditional state densities over the measurement space, from which the approximate posterior state estimate is obtained by evaluating the model at the concrete measurement. An example is illustrated in Figure 1.

Similar to [13], [14], our proposed filter can be integrated as a higher-quality plug-in replacement for the commonly used LRKF during an online filter step. In addition, the sample-based nature of our method greatly enhances versatility. Essentially, it can relax the additive Gaussian noise assumption, allowing for non-Gaussian and non-additive noise models, provided that high-quality (preferably deterministic) samples can be drawn from the true joint measurement/state density.

We validate the proposed filtering method in two representative settings. As a sanity check, we first consider a linear constant-velocity model with process and measurement noise, both assumed Gaussian, for which the KF yields the exact MMSE reference solution. We then study a high-dimensional, nonlinear estimation problem in the context of extended object tracking, where the time-varying position and length of a stick-shaped target are estimated under both additive and multiplicative measurement noise [17], [12].

II. Problem Formulation

A. Models

We consider estimating the hidden state $\underline{\boldsymbol{x}}_k$ of a discrete-time stochastic nonlinear dynamic system based on noisy measurements, consisting of a time update (or prediction step) and a measurement update (or filter step). This work focuses specifically on the challenging measurement update step. The relationship between the measurement random vector $\underline{\boldsymbol{y}}_k$, the system state $\underline{\boldsymbol{x}}_k$, and the measurement noise $\underline{\boldsymbol{v}}_k$ is described by the following generative measurement model

$$\underline{\boldsymbol{y}}_k = \underline{h}_k(\underline{\boldsymbol{x}}_k,\ \underline{\boldsymbol{v}}_k)\ , \tag{1}$$

where $\underline{h}_k(\cdot,\cdot)$ denotes the deterministic vector-valued measurement nonlinearity and k denotes the discrete time step.

A predicted state density is received by performing a time update and then approximating it as a Gaussian through moment matching, i.e., the Gaussian PDF of the state at time step k conditioned on the measurements $\tilde{\underline{y}}_1, \ldots, \tilde{\underline{y}}_{k-2}, \tilde{\underline{y}}_{k-1}$

$$f^{\mathrm{p}}_{\boldsymbol{x}_k}(\underline{x}_k) = f_{\boldsymbol{x}_k}(\underline{x}_k \mid \tilde{\underline{y}}_{1:k-1}) \approx \mathcal{N}(\underline{x}_k;\ \hat{\underline{x}}^{\mathrm{p}}_k,\ \boldsymbol{\Sigma}^{\mathrm{p}}_k)\ . \tag{2}$$

The goal is to correct this Gaussian prior state estimate by incorporating a newly received concrete measurement $\tilde{\underline{y}}_k$ at time step k. In general, it is done by using Bayes' rule. The generative model (1) is first converted into a probabilistic model as the conditional density $f_{\boldsymbol{y}_k}(\underline{y}_k \,|\, \underline{x}_k)$ of $\boldsymbol{y}_k$ given $\underline{x}_k$

$$f_{\boldsymbol{y}_k}(\underline{y}_k \,|\, \underline{x}_k) = \int_{\mathbb{R}^{D_{\boldsymbol{v}}}} \delta\Big(\underline{y}_k - \underline{h}_k(\underline{x}_k,\ \underline{v}_k)\Big)\, f_{\boldsymbol{v}_k}(\underline{v}_k)\, d\underline{v}_k\ . \tag{3}$$

It turns into a likelihood function $f^{\mathrm{L}}_k(\underline{x}_k) \overset{\mathrm{def}}{=} f_{\boldsymbol{y}_k}(\tilde{\underline{y}}_k \mid \underline{x}_k)$ for a given specific measurement $\tilde{\underline{y}}_k$. However, as discussed previously, an explicit likelihood formulation is generally difficult to obtain, especially when $\underline{\boldsymbol{v}}$ cannot be represented as purely additive or scalar-multiplicative noise.

B. Likelihood-free Measurement Update

Hence, we adopt a likelihood-free formulation that avoids explicit construction or evaluation of the likelihood and instead operates directly on the joint density of the prior state and the measurement $f^{\boldsymbol{x},\boldsymbol{y}}_k(\underline{x}_k,\ \underline{y}_k \mid \tilde{\underline{y}}_{1:k-1})$, so the corrected state estimate can be written as

$$f^{\mathrm{e}}_{\boldsymbol{x}_k}(\underline{x}_k) = f_{\boldsymbol{x}_k}(\underline{x}_k \mid \tilde{\underline{y}}_{1:k}) = \frac{f^{\boldsymbol{x},\boldsymbol{y}}_k(\underline{x}_k,\ \tilde{\underline{y}}_k \mid \tilde{\underline{y}}_{1:k-1})}{f_{\boldsymbol{y}_k}(\tilde{\underline{y}}_k \mid \tilde{\underline{y}}_{1:k-1})}\ , \tag{4}$$

in which $\tilde{\underline{y}}_k$ determines where to condition the joint distribution in order to get the posterior state density.

C. Considered Problem

Direct access to the function $f^{\boldsymbol{x},\boldsymbol{y}}_k(\underline{x}_k,\ \tilde{\underline{y}}_k \mid \tilde{\underline{y}}_{1:k-1})$, as well as direct sampling from it, are generally infeasible. Otherwise, the filtering step would be essentially trivial. Furthermore, an analytic expression of the underlying true joint density is typically unavailable. Nevertheless, samples drawn from the joint density are easily available, using either deterministic or stochastic sampling schemes. As a result, the problem at hand is to locally approximate this true joint density using high-quality limited-quantity samples drawn from it. The key idea to solve this problem is given in the next section. By subsequently conditioning this approximate joint density representation on the actual received measurement, we aim to derive an approximate posterior state density. Moreover, the posterior state density may assume an arbitrary shape and is generally non-Gaussian, even if the prior is Gaussian. To maintain a recursive Gaussian filtering framework, we re-approximate it as a single Gaussian distribution through moment matching. This approach preserves computational consistency across processing steps.

III. Key Idea And Groundwork

A. Wasserstein Barycenters

We provide a brief overview of barycenters in Wasserstein space and focus on aspects relevant to this work. For a comprehensive treatment, see [18], [19]. In the Euclidean space, consider a set of points $\underline{x}_1, \ldots, \underline{x}_L \in \mathbb{R}^n$ with weights $w_1, \ldots, w_L$ satisfying $w_i > 0$ and $\sum_{i=1}^L w_i = 1$. The Fréchet mean or barycenter with the Euclidean metric is defined as the minimizer of the weighted sum of squared distances

$$\underline{x}^* = \arg\min_{\underline{x}\in\mathbb{R}^n}\ \sum\nolimits_{i=1}^{L} w_i\ \|\underline{x}_i - \underline{x}\|_2^2\ , \tag{5}$$

In this case, the barycenter is unique and given in closed form by the weighted arithmetic average $\underline{x}^* = \sum_{i=1}^L w_i\, \underline{x}_i$.

In many problems, however, the quantities to be averaged are probability distributions rather than vectors. For instance, a PDF is constrained to be nonnegative and normalized, so pointwise averaging may be poorly aligned with the geometry of the space of distributions. A natural remedy is to keep the Fréchet-mean principle but replace the Euclidean distance by a metric on probability measures. For that reason, we focus on the Wasserstein distance [20], [21].

Let $\mathcal{P}_2(\mathbb{R}^n)$ denote the set of probability measures on $\mathbb{R}^n$ that are absolutely continuous with respect to the Lebesgue measure and have finite second moment, and let $f,\, g : \mathbb{R}^n \to \mathbb{R}_+$ denote the corresponding PDFs, the 2-Wasserstein distance between f and g is defined as

$$W_2(f,\, g) = \left[\inf_{h\in H(f,\,g)} \int \|\underline{x} - \underline{y}\|_2^2\, h(\underline{x},\, \underline{y})\, \mathrm{d}\underline{x}\, \mathrm{d}\underline{y}\right]^{1/2}\ , \tag{6}$$

where $H(f,\, g)$ represents the set of all joint densities $h\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ with marginals g and f, i.e.,

$$\int_{\mathbb{R}^n} h(\underline{x},\, \underline{y})\, \mathrm{d}\underline{y} = f(\underline{x}) \ \text{ and } \ \int_{\mathbb{R}^n} h(\underline{x},\, \underline{y})\, \mathrm{d}\underline{x} = g(\underline{y})\,. \tag{7}$$

Unlike Euclidean averaging, W_2 compares distributions through optimal transport, i.e., via the minimal quadratic cost of redistributing probability mass, leading to barycenters that better respect the geometry of probability measures.

Given a set of PDFs $p_1, \ldots, p_L \in \mathcal{P}_2(\mathbb{R}^n)$ and weights $\{w_i\}_i^L$, the barycenter induced by the 2-Wasserstein distance W_2 (6) is then defined as

$$p^* \in \arg\min_{p\in\mathcal{P}_2(\mathbb{R}^n)}\ \sum\nolimits_{i=1}^{L} w_i\, W_2^2(p_i,\, p)\,. \tag{8}$$

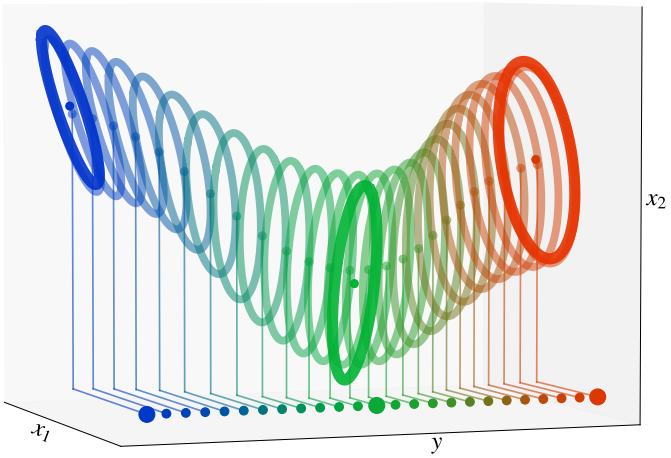


Fig. 2: Wasserstein barycenters (thin, color-blended ellipses) between three given 2D conditional Gaussian state distributions (thick blue, green, and red ellipses) conditioned on three distinct measurement locations along the y-axis.

B. Barycenters in Gaussian Distributions

Although theoretically appealing, the efficient computation of Wasserstein barycenters remains challenging. Many works focus on numerical methods [22], [23]. In particular, while W_2 is tractable on the real line, no easy closed-form solution is available in higher dimensions.

A notable exception arises for multivariate Gaussian distributions. Consider L Gaussians $\mathcal{N}(\underline{\mu}_i, \mathbf{\Sigma}_i)$ on $\mathbb{R}^n$ with nonsingular covariance matrices $\mathbf{\Sigma}_i > \mathbf{0}$ and weights $w_i \geq 0$, $\sum_{i=1}^{L} w_i = 1$. The Wasserstein barycenter with respect to (6) is again Gaussian, i.e., $\mathcal{N}(\underline{\mu}^*, \mathbf{\Sigma}^*)$ [18]. Moreover, the barycenter mean admits a closed form

$$\underline{\mu}^* = \sum_{i=1}^{L} w_i \, \underline{\mu}_i \,, \tag{9}$$

whereas the corresponding covariance $\mathbf{\Sigma}^*$ generally has no analytical expression, except in one dimension or under additional assumptions. Instead, $\mathbf{\Sigma}^*$ is computed as a positive definite fixed point of a nonlinear equation involving $\{\mathbf{\Sigma}_i\}_{i=1}^{L}$ [18]. In particular, [24] proposes an alternative fixed-point method with convergence guarantees towards the true Gaussian barycenter and has improved practical efficiency. The fixed-point iteration used in this work is given by

$$\begin{aligned} &\Psi(\mathbf{\Sigma}) := \mathbf{\Sigma}^{-\frac{1}{2}} \left[\sum_{i=1}^{L} w_i \left(\mathbf{\Sigma}^{\frac{1}{2}} \mathbf{\Sigma}_i \mathbf{\Sigma}^{\frac{1}{2}} \right)^{\frac{1}{2}} \right]^2 \mathbf{\Sigma}^{-\frac{1}{2}} \,, \\ &\mathbf{\Sigma}^{(k+1)} = \Psi(\mathbf{\Sigma}^{(k)}), \text{ iteration step } k \,. \end{aligned} \tag{10}$$

Figure 2 provides a visual illustration of barycentric interpolation in the Gaussian case.

IV. Method Derivation

Leveraging the Gaussian Wasserstein barycenter formulation (10) introduced above, we now use it to design a novel measurement update. Specifically, we construct a continuum of conditional Gaussian state PDFs over the measurement space, which induces a Gaussian-process model in the joint measurement/state space and enables efficient backward inference by conditioning on the received measurement.

A. Gaussian Process Construction

A classical GP is a stochastic process whose random variables are indexed by an input domain, e.g., time or space, and whose finite-dimensional marginals are jointly Gaussian [25]. This viewpoint motivates representing state uncertainty by a continuum of measurement-indexed conditional Gaussian state PDFs. However, a strict GP specification would also require the cross-covariance structure between different measurements. The proposed likelihood-free measurement update does not require such a full stochastic-process structure, since backward inference only depends on the conditional state density at the currently received measurement. Furthermore, our approach to constructing the GP model differs from the conventional data-driven training pipeline, where one specifies a parametric kernel function and then maximizes the marginal likelihood to optimize its hyperparameters.

Instead, we parameterize a measurement-indexed conditional Gaussian model via the Wasserstein barycentric interpolation (10) of multiple conditional Gaussian state PDFs placed at selected measurement-space locations, referred to as *anchor Gaussians*. For any received measurement, this construction directly yields the corresponding approximate posterior state density, whose mean is available in closed form, while the covariance is obtained via a fixed-point iteration (10).

In the multi-dimensional measurement space, we first select a fixed number M of anchor locations $\{\underline{b}_i \in \mathbb{R}^{D_{\boldsymbol{y}}}\}_{i=1}^{M}$. In contrast to [14], they do not require any ordering, e.g., along a certain axis, and are treated as unordered locations in the measurement space. At each $\underline{b}_i$, we define an anchor conditional Gaussian state PDF with parameters $\{\underline{\mu}_i, \mathbf{\Sigma}_i\}_{i=1}^{M}$

$$f_{\boldsymbol{x}}(\underline{x} \mid \underline{y} = \underline{b}_i) = \mathcal{N}(\underline{x}; \underline{\mu}_i \,, \mathbf{\Sigma}_i) \,, \; \underline{\mu}_i \in \mathbb{R}^{D_{\boldsymbol{x}}}, \, \mathbf{\Sigma}_i > \mathbf{0} \,. \tag{11}$$

These parameters $\{\underline{\mu}_i, \mathbf{\Sigma}_i\}_{i=1}^{M}$ are learnable and will be optimized. Subsequently, we apply Wasserstein barycentric interpolation to the anchor Gaussians to obtain, for any concrete measurement $\tilde{\underline{y}}$, the corresponding conditional Gaussian state PDF. This yields a continuum of conditional Gaussians over the measurement space and thereby induces an initial model in the joint measurement/state space.

More importantly, to ensure that the parameterized model accurately approximates the true underlying joint measurement/state density (4), the anchor positions must also be highly flexible, continuously adapting themselves over the measurement space. Moreover, when samples from the true joint density are readily available (preferably deterministic), we can directly learn both the optimal anchor locations and the anchor Gaussian parameters in a data-driven manner.

B. Deterministic Gaussian Sampling

Samples from the true joint measurement/state density are informative for inferring model parameters. Since this density is generally unavailable in closed form, we generate representative measurement/state pairs within the GADF framework by sampling the Gaussian prior state and independent Gaussian measurement noise, either randomly or deterministically.

The samples are then propagated through the deterministic measurement equation (1), yielding samples from the measurement/state density induced by the measurement model.

Deterministic Dirac-mixture approximations of Gaussian distributions include, for example, the sigma points used in the UKF [9]. Here, we employ generalized Fibonacci grids [26], [27], which yield homogeneous coverage using few equally weighted samples at low computational cost. The construction is inspired by sunflower spiral packing, the 2D Fibonacci grid, and its higher-dimensional generalization [28].

C. Parameter Optimization for the Gaussian Process

In this subsection, we optimize the model parameters using samples $\{(\underline{x}_i, \underline{y}_i)\}_{i=1}^N$ drawn from the true joint measurement/state density. Given a measurement $\underline{y}_i$, we first compute nonnegative, normalized interpolation weights $\{w_m(\underline{y}_i, \underline{b}_{1:M})\}_{m=1}^M$ based on a softmax over squared Euclidean distances to the anchor locations $\{\underline{b}_m\}_{m=1}^M$

$$w_m(\underline{y}_i, \underline{b}_{1:M}) \coloneqq \frac{\exp\big(-\|\underline{y}_i - \underline{b}_m\|_2^2 / \rho^2\big)}{\sum_{\ell=1}^M \exp\big(-\|\underline{y}_i - \underline{b}_\ell\|_2^2 / \rho^2\big)}, \tag{12}$$

where $\rho > 0$ is a bandwidth parameter controlling the locality of the interpolation. Intuitively, anchors closer to $\underline{y}_i$ receive larger weights, while distant anchors contribute less.

Although (12) has a kernel-like form, the proposed method is conceptually different from kernel density estimation (KDE). KDE centers a kernel at each sample to approximate the joint measurement/state density and yields a mixture-type conditional density after conditioning. In contrast, we use the samples for parameter estimation and place a fixed number of learnable anchors in the measurement space, each associated with a conditional Gaussian state density. The weights serve only as barycentric coordinates for fusing these anchors in the 2-Wasserstein geometry. Thus, after parameter estimation, conditioning at a concrete measurement is performed through the learned M anchors rather than through N sample-centered kernels, and directly returns a Gaussian state posterior.

With a fixed number of anchors M, the learnable parameters are $\Theta \coloneqq \{\rho, \underline{b}_{1:M}, \underline{\mu}_{1:M}, \mathbf{\Sigma}_{1:M}\}$. The resulting conditional Gaussian state density is fully determined by Θ as

$$f_{\boldsymbol{x}}(\underline{x} \mid \underline{y}_i, \Theta) = \mathcal{N}\big(\underline{x};\, \underline{\mu}_\Theta^{\mathrm{WBI}}(\underline{y}_i),\, \mathbf{\Sigma}_\Theta^{\mathrm{WBI}}(\underline{y}_i)\big), \tag{13}$$

whose mean is available in closed form, while the covariance is obtained by the fixed-point iteration in (10). In backward inference, each state realization $\underline{x}_i$ is modeled as a sample from the conditional Gaussian distribution given $\underline{y}_i$ in (13). The resulting interpolation can be written as

$$\begin{aligned}
\underline{x}_i &\sim f_{\boldsymbol{x}}(\underline{x} \mid \underline{y}_i, \Theta) = \mathcal{N}\big(\underline{x};\, \underline{\mu}_\Theta^{\mathrm{WBI}}(\underline{y}_i),\, \mathbf{\Sigma}_\Theta^{\mathrm{WBI}}(\underline{y}_i)\big), \\
\underline{\mu}_\Theta^{\mathrm{WBI}}(\underline{y}_i) &= \sum\nolimits_{m=1}^M w_m(\underline{y}_i, \underline{b}_{1:M})\, \underline{\mu}_m, \\
\Psi(\mathbf{\Sigma}) &\coloneqq \mathbf{\Sigma}^{-\frac{1}{2}} \left[\sum\nolimits_{m=1}^M w_m \left(\mathbf{\Sigma}^{\frac{1}{2}} \mathbf{\Sigma}_m \mathbf{\Sigma}^{\frac{1}{2}}\right)^{\frac{1}{2}}\right]^2 \mathbf{\Sigma}^{-\frac{1}{2}}, \\
\mathbf{\Sigma}^{(k+1)} &= \Psi(\mathbf{\Sigma}^{(k)}), \text{ iteration step } k, \\
\mathbf{\Sigma}_\Theta^{\mathrm{WBI}}(\underline{y}_i) &= \lim_{k\to\infty} \mathbf{\Sigma}^{(k)} = \mathbf{\Sigma}^*.
\end{aligned} \tag{14}$$

The parameter set Θ can be estimated from the sample set $\{(\underline{x}_i, \underline{y}_i)\}_{i=1}^N$ via Maximum Likelihood Estimation (MLE). Specifically, MLE determines the optimal parameter set Θ^* that maximizes the (training-)data likelihood of the observed samples under the assumed parametric model.

$$\mathcal{L}(\Theta) \coloneqq f\left(\{\underline{x}_i\}_{i=1}^N \;\middle|\; \{\underline{y}_i\}_{i=1}^N, \Theta\right). \tag{15}$$

Building on the above formulation, we now describe how to estimate the parameter set Θ^* via MLE. We first construct the likelihood function $\mathcal{L}(\Theta)$ and, for simplicity, assume that the sample pairs $(\underline{x}_i, \underline{y}_i)$ are mutually independent.

$$\begin{aligned}
\Theta^* &= \arg\max_\Theta \mathcal{L}(\Theta) = \arg\max_\Theta f\left(\{\underline{x}_i\}_{i=1}^N \;\middle|\; \{\underline{y}_i\}_{i=1}^N, \Theta\right) \\
&= \arg\max_\Theta \prod_{i=1}^N \mathcal{N}\left(\underline{x}_i;\, \underline{\mu}_\Theta^{\mathrm{WBI}}, \mathbf{\Sigma}_\Theta^{\mathrm{WBI}}\right) \\
&= \arg\max_\Theta \log \prod_{i=1}^N \mathcal{N}\left(\underline{x}_i;\, \underline{\mu}_\Theta^{\mathrm{WBI}}, \mathbf{\Sigma}_\Theta^{\mathrm{WBI}}\right) \\
&= \arg\max_\Theta \sum\nolimits_{i=1}^N \log \mathcal{N}\left(\underline{x}_i;\, \underline{\mu}_\Theta^{\mathrm{WBI}}, \mathbf{\Sigma}_\Theta^{\mathrm{WBI}}\right) \\
&= \arg\min_\Theta \sum\nolimits_{i=1}^N \log|\mathbf{\Sigma}_\Theta^{\mathrm{WBI}}| + \underline{r}_i^\top (\mathbf{\Sigma}_\Theta^{\mathrm{WBI}})^{-1} \underline{r}_i,
\end{aligned} \tag{16}$$

with $\underline{r}_i \coloneqq \underline{x}_i - \underline{\mu}_\Theta^{\mathrm{WBI}}$. Taking the negative logarithm of the likelihood yields the objective in (16), which, for each sample pair $(\underline{x}_i, \underline{y}_i)$, comprises a log-determinant term and a Mahalanobis term. Minimizing this negative log-likelihood yields the MLE estimate Θ^*. Importantly, solving (16) remains nontrivial, since it requires repeated evaluations of $\log|\mathbf{\Sigma}_\Theta^{\mathrm{WBI}}|$ and $(\mathbf{\Sigma}_\Theta^{\mathrm{WBI}})^{-1}$. In addition, $\mathbf{\Sigma}_\Theta^{\mathrm{WBI}}$ is obtained only implicitly via the fixed-point mapping (10) that repeatedly involves matrix square-root operations. These operations become particularly costly in high-dimensional state spaces.

D. Analytical Simplification via Shared Eigen-Basis

To make likelihood evaluation analytically tractable and computationally efficient, we impose a structural assumption on the covariance matrices. Specifically, since any real symmetric matrix admits an orthogonal eigendecomposition, we assume that the anchor covariances $\{\mathbf{\Sigma}_m\}_{m=1}^M$ and the barycenter covariance $\mathbf{\Sigma}$ share a common orthogonal eigenbasis $\mathbf{U}$. Equivalently, the involved covariances can be written as $\mathbf{\Sigma}_m = \mathbf{U}\mathbf{D}_m\mathbf{U}^\top$ and $\mathbf{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ for a common orthogonal matrix $\mathbf{U}$ and diagonal matrices $\mathbf{D}_m$ and $\mathbf{D}$. Accordingly, the learnable parameter set becomes $\Theta \coloneqq \{\rho, \mathbf{U}, \underline{b}_{1:M}, \underline{\mu}_{1:M}, \underline{\lambda}_{1:M}\}$, where $\underline{\lambda}_m = [\lambda_m^1, \ldots, \lambda_m^{D_{\boldsymbol{x}}}]^\top$ collects the eigenvalues of the m-th anchor covariance $\mathbf{\Sigma}_m$. To preserve the orthogonality constraint $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$ during optimization, we parameterize $\mathbf{U}$ by unconstrained rotation parameters rather than optimizing its entries directly. In the 2D case, $\mathbf{U}$ can be written as

$$\mathbf{U}(\alpha) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}, \quad \alpha \in \mathbb{R}. \tag{17}$$

For general $D_{\boldsymbol{x}}$, we use a Givens-rotation parameterization [29] with $D_{\boldsymbol{x}}(D_{\boldsymbol{x}}-1)/2$ unconstrained parameters.

Under the proposed shared eigen-basis assumption, the matrix square roots of $\mathbf{\Sigma}$ in (10) simplify to

$$\mathbf{\Sigma}^{\frac{1}{2}} = \mathbf{U}\,\mathbf{D}^{\frac{1}{2}}\,\mathbf{U}^\top,\ \mathbf{D}^{\frac{1}{2}} = \operatorname{diag}\Big(\sqrt{\lambda^1},\,\ldots,\,\sqrt{\lambda^{D_{\boldsymbol{x}}}}\Big)\,. \quad (18)$$

Consequently, the inner term becomes

$$\begin{aligned}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Sigma}_m\mathbf{\Sigma}^{\frac{1}{2}} &= \mathbf{U}\,\mathbf{D}^{\frac{1}{2}}\,\mathbf{U}^\top\,\mathbf{U}\,\mathbf{D}_m\,\mathbf{U}^\top\,\mathbf{U}\,\mathbf{D}^{\frac{1}{2}}\,\mathbf{U}^\top \\ &= \mathbf{U}\,\mathbf{D}^{\frac{1}{2}}\,\mathbf{D}_m\,\mathbf{D}^{\frac{1}{2}}\,\mathbf{U}^\top = \mathbf{U}\,\mathbf{D}\,\mathbf{D}_m\,\mathbf{U}^\top, \quad (19)\end{aligned}$$

Thus, its square root can be taken element-wise in the $\mathbf{U}$-basis

$$\Big(\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Sigma}_m\mathbf{\Sigma}^{\frac{1}{2}}\Big)^{\frac{1}{2}} = \mathbf{U}\,\mathbf{D}^{\frac{1}{2}}\,\mathbf{D}_m^{\frac{1}{2}}\,\mathbf{U}^\top. \quad (20)$$

Substituting this result into the fixed-point map in (10) yields

$$\sum\nolimits_{m=1}^{M} w_m \Big(\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Sigma}_m\mathbf{\Sigma}^{\frac{1}{2}}\Big)^{\frac{1}{2}} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}\left(\sum\nolimits_{m=1}^{M} w_m\mathbf{D}_m^{\frac{1}{2}}\right)\mathbf{U}^\top. \quad (21)$$

Finally, left- and right-multiplication of the squared bracketed term (21) by $\mathbf{\Sigma}^{-\frac{1}{2}} = \mathbf{U}\,\mathbf{D}^{-\frac{1}{2}}\,\mathbf{U}^\top$ cancels the $\mathbf{D}$ terms

$$\Psi(\mathbf{\Sigma}) = \mathbf{U}\left(\sum\nolimits_{m=1}^{M} w_m\,\mathbf{D}_m^{\frac{1}{2}}\right)^2\mathbf{U}^\top, \quad (22)$$

with $\mathbf{U}\mathbf{U}^\top = \mathbf{U}^\top\mathbf{U} = \mathbf{I}$. Thus, under the proposed shared eigen-basis assumption, the fixed-point map no longer depends on the current iterate $\mathbf{\Sigma}$. Therefore, for any initialization $\mathbf{\Sigma}_0$ compatible with the same $\mathbf{U}$-basis, e.g., $\mathbf{\Sigma}_0 = \mathbf{U}\mathbf{U}^\top = \mathbf{I}$, the iteration $\mathbf{\Sigma}^{(k+1)} = \Psi(\mathbf{\Sigma}^{(k)})$ converges in a single step (22). We then obtain the component-wise closed-form solution for the barycenter covariance

$$\mathbf{\Sigma}^* = \mathbf{U}\,\mathbf{D}^*\,\mathbf{U}^\top, \quad \mathbf{D}^* = \operatorname{diag}(s_1^2,\,\ldots,\,s_{D_{\boldsymbol{x}}}^2)\,, \quad (23)$$

$$s_j \coloneqq \sum\nolimits_{m=1}^{M} w_m\,\sqrt{\lambda_m^j},\ j = 1,\,\ldots,\,D_{\boldsymbol{x}}\,. \quad (24)$$

The closed-form expression in (22) renders the likelihood function analytically tractable under the shared eigen-basis assumption and enables an efficient evaluation of the log-likelihood. This is because it turns the otherwise costly evaluation of $\log|\mathbf{\Sigma}_\Theta^{\text{WBI}}|$ and $(\mathbf{\Sigma}_\Theta^{\text{WBI}})^{-1}$ in (16) into simple component-wise operations. Since $\mathbf{U}$ is orthogonal with $|\det(\mathbf{U})| = 1$, the log-determinant term simplifies to

$$\begin{aligned}\log|\mathbf{\Sigma}_\Theta^{\text{WBI}}| = \log|\mathbf{D}_\Theta^{\text{WBI}}| &= \log\left(\prod\nolimits_{j=1}^{D_{\boldsymbol{x}}} s_{j,\,i}^2\right) \\ &= \sum\nolimits_{j=1}^{D_{\boldsymbol{x}}}\log(s_{j,\,i}^2) = 2\sum\nolimits_{j=1}^{D_{\boldsymbol{x}}}\log(s_{j,\,i})\,, \quad (25)\end{aligned}$$

$$s_{j,\,i} = \sum\nolimits_{m=1}^{M} w_m(\underline{y}_i)\,\sqrt{\lambda_m^j} > 0\,. \quad (26)$$

Moreover, the inverse is given by

$$(\mathbf{\Sigma}_\Theta^{\text{WBI}})^{-1} = \mathbf{U}\,(\mathbf{D}_\Theta^{\text{WBI}})^{-1}\,\mathbf{U}^\top, \quad (27)$$

$$(\mathbf{D}_\Theta^{\text{WBI}})^{-1} = \operatorname{diag}(s_{1,\,i}^{-2},\,\ldots,\,s_{D_{\boldsymbol{x}},\,i}^{-2})\,. \quad (28)$$

Let $\tilde{\underline{r}}_i \coloneqq \mathbf{U}^\top\underline{r}_i$, the quadratic form in (16) then becomes

$$\begin{aligned}\underline{r}_i^\top\,(\mathbf{\Sigma}_\Theta^{\text{WBI}})^{-1}\,\underline{r}_i &= \underline{r}_i^\top\,\mathbf{U}\,(\mathbf{D}_\Theta^{\text{WBI}})^{-1}\,\mathbf{U}^\top\underline{r}_i \\ &= \tilde{\underline{r}}_i^\top\,(\mathbf{D}_\Theta^{\text{WBI}})^{-1}\,\tilde{\underline{r}}_i = \sum\nolimits_{j=1}^{D_{\boldsymbol{x}}}\frac{\tilde{r}_{j,\,i}^{\,2}}{s_{j,\,i}^2}. \quad (29)\end{aligned}$$

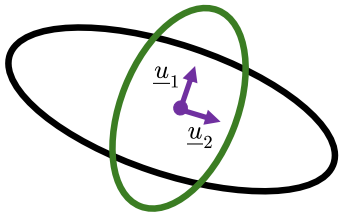


Fig. 3: A 2D illustration of the shared eigen-basis assumption. Both Gaussians share the orthonormal eigenvectors $\mathbf{U} = [\underline{u}_1,\,\underline{u}_2]$ (purple), but differ in their eigenvalues, yielding different ellipse shapes along the same directions.

Substituting (25) and (29) into the negative log-likelihood yields the equivalent objective

$$\begin{aligned}\Theta^* &= \arg\min_\Theta\ \sum\nolimits_{i=1}^{N}\big[\log|\mathbf{\Sigma}_\Theta^{\text{WBI}}| + \underline{r}_i^\top(\mathbf{\Sigma}_\Theta^{\text{WBI}})^{-1}\underline{r}_i\big] \\ &= \arg\min_\Theta\ \sum\nolimits_{i=1}^{N}\Big[\sum\nolimits_{j=1}^{D_{\boldsymbol{x}}}\big(2\,\log(s_{j,\,i}) + \frac{\tilde{r}_{j,\,i}^{\,2}}{s_{j,\,i}^2}\big)\Big]. \quad (30)\end{aligned}$$

The shared eigen-basis assumption replaces the implicit fixed-point evaluation of the Gaussian Wasserstein barycenter covariance by a closed-form expression. This makes the covariance-dependent MLE terms dimension-wise separable, allowing the log-determinant and Mahalanobis term to be evaluated by inexpensive component-wise operations. This substantially reduces the computational burden and improves numerical robustness in high-dimensional state spaces.

V. Evaluation

We evaluate the proposed approach in several scenarios and compare it with multiple state-of-the-art methods.

A. Sanity Check: Linear System

We consider first a linear constant-velocity model subject to Gaussian process noise $\underline{\boldsymbol{w}}$ and Gaussian measurement noise $\boldsymbol{\nu}$. The state $\underline{\boldsymbol{x}}_k = [\boldsymbol{p}_k,\,\boldsymbol{v}_k]^\top \in \mathbb{R}^2$ is two-dimensional, where $\boldsymbol{p}_k$ denotes position and $\boldsymbol{v}_k$ velocity. The system equation is

$$\underline{\boldsymbol{x}}_{k+1} = \mathbf{A}\,\underline{\boldsymbol{x}}_k + \underline{\boldsymbol{w}}\,, \quad \underline{\boldsymbol{w}} \sim \mathcal{N}(\underline{w};\,\underline{0},\,\mathbf{Q})\,, \quad (31)$$

with

$$\mathbf{A} = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}, \quad \mathbf{Q} = \operatorname{diag}(0.1^2,\,0.05^2)\,. \quad (32)$$

The scalar measurement equation follows

$$\boldsymbol{y}_k = \boldsymbol{p}_k + \boldsymbol{\nu}\,, \quad \boldsymbol{\nu} \sim \mathcal{N}(\nu;\,0,\,1.0^2)\,. \quad (33)$$

We initialize the filter with the Gaussian prior state estimate

$$f_{\boldsymbol{x}_0}^{\text{p}}(\underline{x}_0) = \mathcal{N}(\underline{x}_0;\,[0,\,0]^\top,\,\mathbf{P}_0)\,,\ \mathbf{P}_0 = \begin{bmatrix}1.0 & 1.15\\ 1.15 & 1.5\end{bmatrix}. \quad (34)$$

For the Monte Carlo simulations, we generate 50 trajectories of length $T = 50$ starting from $\underline{x}_0^{\text{true}} = [0,\,0]^\top$. At each time step k, we draw $N = 300$ equally weighted deterministic samples from the Gaussian joint distribution of the state and measurement noise, propagate them through the measurement equation to obtain pairs $\{(\underline{x}_i,\,y_i)\}_{i=1}^N$, and fit the parameterized WBI model with 6 anchors via MLE-optimization. Conditioning the resulting continuum of conditional Gaussians

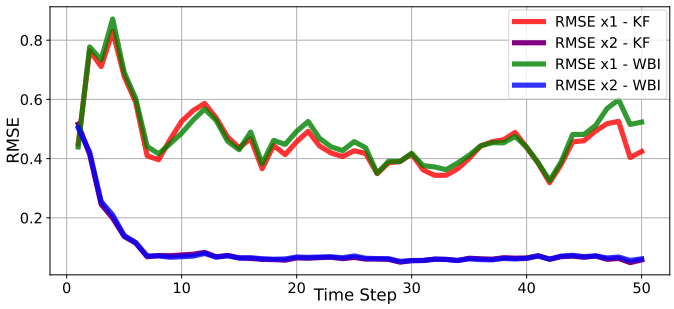


Fig. 4: Sanity check of the proposed approach on a linear constant-velocity model with Gaussian process/measurement noise. RMSE trajectories for both state components over 50 time steps, averaged over 50 Monte Carlo runs.

on the received measurement $\tilde{y}_k$ yields the posterior state estimate. Performance is reported as the Root Mean Square Error (RMSE) at each time step for each state component, i.e., for position $\boldsymbol{x}_k^1 = \boldsymbol{p}_k$ and velocity $\boldsymbol{x}_k^2 = \boldsymbol{v}_k$, averaged over Monte Carlo runs. From Fig. 4, we observe that the proposed filter achieves performance comparable to the KF, which provides the exact MMSE reference solution.

B. Extended Object Tracking Scenario

In the following, we consider an extended object tracking application, where we estimate the length $\boldsymbol{l}$ of a stick together with its one-dimensional position $\boldsymbol{c}$. The state vector is $\underline{\boldsymbol{x}}_k = [\boldsymbol{c}_k, \boldsymbol{l}_k]^\top \in \mathbb{R}^2$. At each time step k, a measurement can originate from an arbitrary point on the stick and is corrupted by additive noise. Following the multiplicative-noise extent model in [17], we employ the measurement equation

$$\boldsymbol{y}_k = \boldsymbol{c}_k + \boldsymbol{l}_k \boldsymbol{u} + \boldsymbol{\nu}\,, \tag{35}$$

where $\boldsymbol{\nu} \sim \mathcal{N}(\nu;\, 0,\, 0.15^2)$ denotes additive state-independent Gaussian noise and $\boldsymbol{u} \sim \mathcal{U}(-1,\, 1)$ accounts for the uniformly distributed multiplicative noise. To capture the temporal evolution of the state, we adopt an identity transition random walk model $\underline{\boldsymbol{x}}_{k+1} = \underline{\boldsymbol{x}}_k + \underline{\boldsymbol{w}}$. Here, $\underline{\boldsymbol{w}}$ denotes state-independent system noise modeled as $\underline{\boldsymbol{w}} \sim \mathcal{N}\big(\underline{w};\, \underline{0},\, \mathrm{diag}(0.1^2,\, 0.5^2)\big)$.

A similar setup was studied in [12], where the PGF was shown to outperform the other methods considered. Here, we compare the proposed approach with the state-of-the-art LRKF S^{2}KF [10], the PGF [12], and a GPF [4]. For the GPF, we draw 10^6 random samples per time step for moment matching. Although this requires a high computational cost and introduces stochastic variability, the GPF converges asymptotically to the optimal MMSE estimator under the Gaussian posterior approximation and therefore serves as a reference for the estimation quality attainable by a GADF. To ensure a comparison that is as fair as possible, we match the sampling budget per measurement update. Specifically, the proposed method and the S^{2}KF each use 300 equally weighted deterministic Gaussian samples. For the PGF, we set the maximal progression steps to 10 and use 30 samples per progression step. In addition, we use 6 anchor Gaussians.

All filters are initialized with the same Gaussian prior, i.e., $\underline{\mu}_0^{\mathrm{e}} = [-1,\, 4]^\top$ and $\boldsymbol{\Sigma}_0^{\mathrm{e}} = \mathbf{I}_2$. We generate a ground-truth trajectory over 200 time steps in which both the target position and its extent vary sinusoidally. At each time step, we obtain a batch of 50 noisy measurements. The results in Fig. 5 show the posterior mean of the Gaussian state approximations produced by different GADFs at each time step, averaged over 20 independent Monte Carlo runs. The S^{2}KF fails to reliably estimate the stick length in this scenario. In contrast, the proposed filter produces estimates that closely match the GPF-based reference, while using only 300 equally weighted deterministic samples per measurement update. Overall, both the proposed method and the PGF achieve strong performance, with the proposed method exhibiting a marginal advantage over the PGF.

VI. Conclusion

In this paper, we proposed a novel GADF for nonlinear state estimation, with a particular focus on the measurement update, which is often analytically intractable. To address this challenge, we construct a well-defined continuum of conditional Gaussian state densities over the measurement space via a multi-anchor 2-Wasserstein barycentric interpolation, capturing the local non-Gaussian structure of the joint measurement/state density without explicit likelihood evaluation.

To reduce the computational burden of barycenter covariance computation, we introduced a shared eigen-basis parameterization that yields a closed-form expression for the barycenter covariance and enables inexpensive MLE-based parameter optimization from equally weighted deterministic samples. Our proposed filter can thus be integrated as a higher-quality plug-in replacement for the commonly used LRKF during an online filter step. As evidenced by the evaluations, the sample-based nature of our method supports non-Gaussian and non-additive noise models while delivering consistently high filtering performance.

Future work will address potential MLE instabilities under data sparsity in higher-dimensional spaces. In addition, we will investigate scalable update schemes, such as mini-batch optimization and parallel GPU implementations.

References

[1] Sebastian Thrun. “Probabilistic robotics”. In: *Communications of the ACM* 45.3 (2002), pp. 52–57.

[2] Dan Simon. *Optimal State Estimation*. 1st. Hoboken, NJ: Wiley & Sons, 2006.

[3] Andrew H Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.

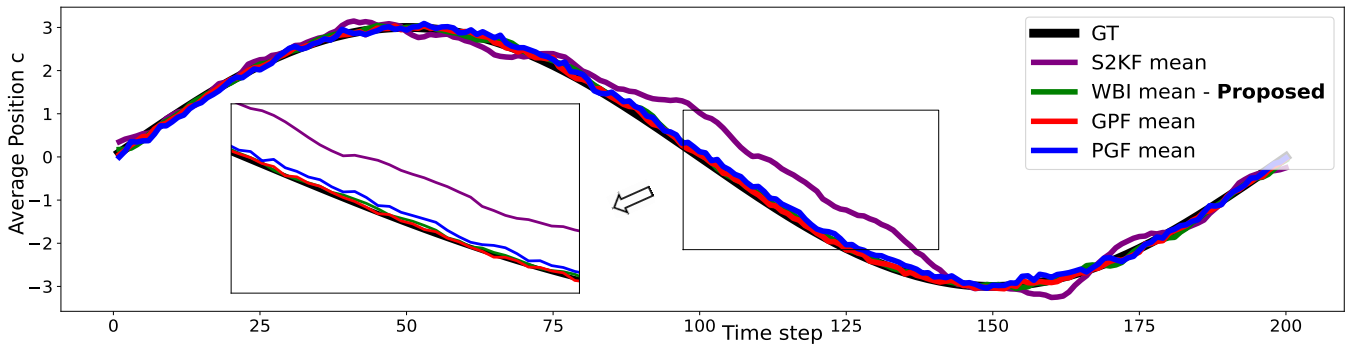
[4] Jayesh H. Kotecha and Petar M. Djuric. “Gaussian Particle Filtering”. In: *IEEE Transactions on Signal Processing* 51.10 (2003), pp. 2592–2601. DOI: 10.1109/TSP.2003.817204.

[5] Mohinder S. Grewal and Angus P. Andrews. *Kalman Filtering*. 2nd. New York: Wiley, 2001.

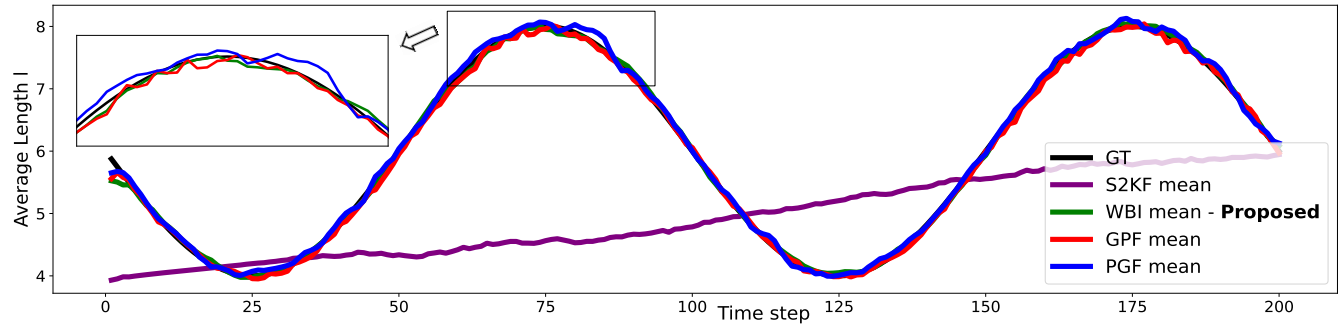
[6] Harold W. Sorenson. *Recursive Estimation for Nonlinear Dynamic Systems*. New York: Wiley, 1988.

[7] Tom Lefebvre, Herman Bruyninckx, and Joris De Schutter. “The Linear Regression Kalman Filter”. In: *Nonlinear Kalman Filtering for Force-Controlled Robot Tasks*. Ed. by Michel Verhaegen and Paul Van Dooren. Vol. 19. Springer, 2005, pp. 205–210.

[8] Tom Lefebvre, Herman Bruyninckx, and Joris De Schutter. “Kalman Filters for Non-Linear Systems: A Comparison of Performance”. In: *International Journal of Control* 77.7 (2004), pp. 639–653.



(a) Averaged posterior mean estimate of the target position.



(b) Averaged posterior mean estimate of the target length.

Fig. 5: Comparison of averaged posterior mean estimates for the target position and length over 20 independent Monte Carlo runs with 200 time steps each.

[9] Simon J. Julier and Jeffrey K. Uhlmann. "Unscented Filtering and Nonlinear Estimation". In: *Proceedings of the IEEE* 92.3 (2004), pp. 401–422. DOI: 10.1109/JPROC.2003.823141.

[10] Jannik Steinbring and Uwe D. Hanebeck. "LRKF Revisited: The Smart Sampling Kalman Filter (S2KF)". In: *Journal of Advances in Information Fusion* 9.2 (Dec. 2014), pp. 106–123.

[11] Uwe D. Hanebeck. "PGF 42: Progressive Gaussian Filtering with a Twist". In: *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*. Istanbul, Turkey, 2013, pp. 1103–1110.

[12] Jannik Steinbring and Uwe D. Hanebeck. "Progressive Gaussian Filtering Using Explicit Likelihoods". In: *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*. Salamanca, Spain, July 2014.

[13] Jiachen Zhou, Daniel Frisch, and Uwe D. Hanebeck. "Inverse Gaussian Process Interpolation for High-Quality Assumed Gaussian Filtering". In: *2024 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2024)*. 2024, pp. 1–8. DOI: 10.1109/MFI62651.2024.10705784.

[14] Jiachen Zhou and Uwe D. Hanebeck. "High-Quality Assumed Gaussian Filtering Based on Wasserstein Barycentric Interpolation". In: *Proceedings of the 28th International Conference on Information Fusion (FUSION 2025)*. Rio de Janeiro, Brazil, July 2025.

[15] Licong Zhang, Jürgen Sturm, Daniel Cremers, and Dongheui Lee. "Real-time human motion tracking using multiple depth cameras". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2012, pp. 2389–2395.

[16] Marcus Baum and Uwe D. Hanebeck. "Shape Tracking of Extended Objects and Group Targets with Star-Convex RHMs". In: *Proceedings of the 14th International Conference on Information Fusion (Fusion 2011)*. Chicago, Illinois, USA, July 2011.

[17] Marcus Baum, Florian Faion, and Uwe D. Hanebeck. "Modeling the Target Extent with Multiplicative Noise". In: *Proceedings of the 15th International Conference on Information Fusion (Fusion 2012)*. Singapore, July 2012.

[18] Martial Agueh and Guillaume Carlier. "Barycenters in the Wasserstein space". In: *SIAM Journal on Mathematical Analysis* 43.2 (2011), pp. 904–924.

[19] Marco Cuturi and Arnaud Doucet. "Fast computation of Wasserstein barycenters". In: *International conference on machine learning*. PMLR. 2014, pp. 685–693.

[20] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2005.

[21] Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2009.

[22] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. "Iterative Bregman projections for regularized transportation problems". In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138.

[23] Guillaume Carlier, Adam Oberman, and Edouard Oudet. "Numerical methods for matching for teams and Wasserstein barycenters". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 49.6 (2015), pp. 1621–1642.

[24] Pedro C Álvarez-Esteban, E Del Barrio, JA Cuesta-Albertos, and C Matrán. "A fixed-point approach to barycenters in Wasserstein space". In: *Journal of Mathematical Analysis and Applications* 441.2 (2016), pp. 744–762.

[25] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006. ISBN: 978-0-262-18253-9.

[26] Daniel Frisch and Uwe D. Hanebeck. "Deterministic Gaussian Sampling With Generalized Fibonacci Grids". In: *Proceedings of the 24th International Conference on Information Fusion (Fusion 2021)*. Sun City, South Africa, Nov. 2021.

[27] Daniel Frisch and Uwe D. Hanebeck. "The Generalized Fibonacci Grid as Low-Discrepancy Point Set for Optimal Deterministic Gaussian Sampling". In: *Journal of Advances in Information Fusion* 18.1 (June 2023), pp. 16–34. ISSN: 1557-6418.

[28] Robert James Purser. *Generalized Fibonacci Grids; A New Class of Structured, Smoothly Adaptive Multi-Dimensional Computational Lattices*. May 2008.

[29] Thomas Frerix and Joan Bruna. "Approximating orthogonal matrices with effective Givens factorization". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 1993–2001.