# High-Quality Assumed Gaussian Filtering Based on Wasserstein Barycentric Interpolation

**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 introduce a novel Gaussian Assumed Density Filter (GADF) for high-quality state estimation in discrete-time stochastic nonlinear dynamic systems, with a primary focus on the measurement update. Rooted in optimal transport theory, the Wasserstein distance is employed as a powerful metric for comparing probability distributions. Building on this foundation, we utilize the unique, explicit Wasserstein barycentric interpolation between Gaussian distributions to parameterize an initial Gaussian Process (GP) in the joint measurement/prior state space. Deterministic samples drawn from the true joint measurement/state density are then used with likelihood-based parameter estimation techniques to optimize the parameters of this Gaussian Process. As a result, the derived Gaussian Process provides a local non-Gaussian approximation to the true joint density. This approach eliminates the need for a second Gaussian assumption on the joint density and avoids an explicit likelihood function, making it a higher-quality plug-in replacement for the commonly used Linear Regression Kalman Filter (LRKF).**

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

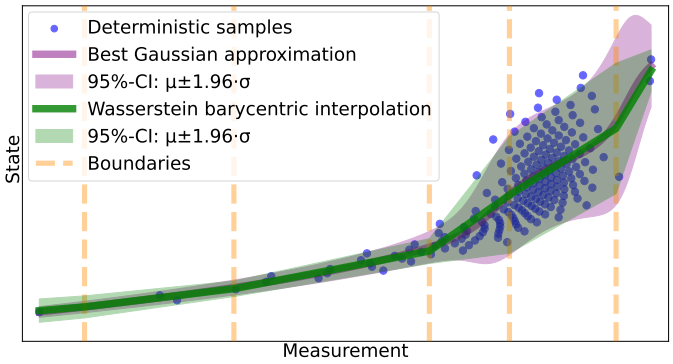


Fig. 1: Exemplary illustration of the proposed non-Gaussian local approximation (green) of the true joint measurement/state density using Wasserstein barycentric interpolation between each pair of adjacent Gaussians positioned at the boundaries (orange). The approach is based on samples (blue) drawn from the true joint. For comparison, the family of conditional Gaussian densities (purple) for each measurement (horizontal axis), with mean and variance identical to the true posterior density.

## I. Introduction

### A. Context

We consider the general state estimation problem for a discrete-time stochastic nonlinear dynamic system with noisy measurements. Specifically, 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).

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 [1]. Gaussian mixture estimators can then be constructed based on these filters [2], [3], [4]. However, even with these simplifications, it is still challenging to derive closed-form solutions for the time update, and especially the measurement update. A Bayesian measurement update requires an explicit likelihood function, but deriving one for non-additive measurement noise is challenging. Even when available, an analytical execution of the filter step is often infeasible.

### B. State of the Art

The Gaussian Particle Filter (GPF) [5] is a special sequential importance sampling Particle Filter (PF) [6], [7] that approximates posterior distributions by single Gaussians. It employs Monte Carlo integration for moment matching and converges to the Minimum Mean Square Error (MMSE) estimator as the sample size approaches infinity. However, its computational cost scales poorly with state dimension due to the curse of dimensionality, and it still requires an explicit likelihood function for sample re-weighting.

To further ease the measurement update, likelihood-free approaches have been introduced. These methods approximate the nonlinear mapping between the prior state and noisy measurements linearly and then apply the Kalman filter formulas, leading to Nonlinear Kalman Filters (NKFs). While they eliminate the need for an explicit likelihood function, their effectiveness depends on the "strength" of nonlinearity. In cases of strong nonlinearity, this approximation can be overly simplistic, resulting in reduced estimation accuracy compared to more general GADFs without such linearization.

One approach to linearization is explicit linearization via Taylor series expansion, as employed by Extended Kalman Filters (EKFs) and its variants [8], [9]. 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 required first- and second-order moments are computed using sample-based density representations, these NKFs fall under the category of Linear Regression Kalman Filters (LRKFs) [10], [11]. Notable examples include the Unscented Kalman Filter (UKF) [12] and the Smart Sampling Kalman Filter ($S^2$KF) [13].

LRKFs offer many advantages in terms of computational efficiency and ease of implementation. Unlike PFs and GPFs, they circumvent the issue of sample degeneration due to their likelihood-free nature. Furthermore, their flexible design enables effective application to highly nonlinear systems. However, their primary limitation lies in the reduced state estimation accuracy, as they approximate continuous state and noise densities using 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 [14], [15]. It achieves this by decomposing the measurement update into multiple sub-updates, gradually integrating measurement information into state estimates. In [15], 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.

In particular, a closely related approach is the GADF based on the so-called Inverse Gaussian Process (IGP) Interpolation [16]. Unlike conventional methods, it avoids both the second Gaussian assumption and the need for an explicit likelihood function during measurement updates. Instead, it first generates deterministic samples from the true joint measurement/state density and subsequently approximates this joint density through a sequential use of two IGPs. As a result, this GADF achieves performance comparable to the optimal MMSE estimator. However, the necessity of two sequential matrix inversions, each with cubic time complexity, leads to lengthy runtimes that impede real-time applications.

### C. Contributions

In this work, we propose a GADF with a novel likelihood-free measurement update. Our method does not rely on the second Gaussian assumption for the true joint measurement/state density, leading to better performance than the LRKFs, which heavily depend on this simplification. Instead, we perform a non-Gaussian local approximation of the true joint density through a two-stage approach. First, within the joint space of measurement and prior state, we construct a Gaussian Process (GP) model using the unique, explicit Wasserstein barycentric interpolation technique. In this way, at each measurement value along the measurement axis, a corresponding Gaussian conditional state PDF is readily obtained, effectively fulfilling the goal of backward inference. In the next step, deterministic and equally weighted samples of high quality are drawn from the true joint measurement/state density. Using these samples, likelihood-based parameter estimation techniques are applied in a data-driven manner to optimize the parameters of the GP. Together, the resulting GP fully characterizes the conditional Gaussian densities of the hidden state on concrete measurements as the approximate posterior state estimates. To enhance visualization, we employ a rotated Cartesian coordinate system throughout this paper, with measurements on the horizontal axis and prior/posterior states on the vertical axis. For instance, this configuration is illustrated in Figure 1.

Like [16], our proposed filter can also be integrated as a higher-quality plug-in replacement for the commonly used LRKF during an online filter step. Evaluations on a canonical benchmark demonstrate that the proposed filter not only achieves state estimation performance comparable to state-of-the-art GADFs, but also significantly improves computational efficiency relative to [16]. Moreover, the data-driven, sample-based nature of this approach 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.

## II. Problem Formulation

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 hidden system state $\underline{\boldsymbol{x}}_k$, and the measurement noise $\underline{\boldsymbol{v}}_k$ is described by the nonlinear 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 known vector-valued measurement nonlinearity and the subscript $k$ 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 $\tilde{\underline{y}}_1, \ldots, \tilde{\underline{y}}_{k-2}, \tilde{\underline{y}}_{k-1}$

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

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 measurement model (1) is first converted into a probabilistic model as the conditional density $f_{\boldsymbol{y}_k}(\underline{y}_k \,|\, \underline{x}_k)$ of $\underline{y}_k$ given $\underline{x}_k$, which 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 mentioned before, an explicit description of the likelihood function is hard to derive in the case of non-additive measurement noise. Hence, we adopt a likelihood-free point of view, which instead operates directly on the joint density of prior state and 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

$$f^{\mathrm{e}}_{\boldsymbol{x}_k}(\underline{x}_k) = f_{\boldsymbol{x}_k}(\underline{x}_k \mid \tilde{\underline{y}}_{1:k}) = \frac{f^{\boldsymbol{x},\boldsymbol{y}}_k(\underline{x}_k,\ \tilde{\underline{y}}_k \mid \tilde{\underline{y}}_{1:k-1})}{f_{\boldsymbol{y}_k}(\tilde{\underline{y}}_k \mid \tilde{\underline{y}}_{1:k-1})}\ , \tag{3}$$

in which a newly received concrete measurement $\underline{\tilde{y}}_k$ determines where to condition the joint distribution in order to get the posterior state density as the corrected state estimate.

*Problem at hand:* Directly obtaining the true joint density is still a burden. Nevertheless, samples drawn from it are easily available, regardless of whether deterministic or random sampling methods are used. 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 concrete measurement, we aim to derive an approximate posterior state density. In addition, this posterior state density can take an arbitrary form and is not necessarily Gaussian, even when starting with a Gaussian prior. Therefore, it is re-approximated as a single Gaussian distribution through moment matching. This approach ensures computational consistency in recursive processing.

## III. Key Idea And Groundwork

### A. Optimal Mass Transport

We present a brief overview of the Optimal Mass Transport (OMT) theory, highlighting only the aspects relevant to this work. For a more comprehensive treatment, see [17].

OMT seeks to transport mass from one distribution to another, preserving total mass while minimizing the transportation cost. Concretely, let $\nu_0$ and $\nu_1$ be two probability measures over a common space $\mathbb{R}^n$. In the classical OMT framework, the goal is to determine a transportation map $T : \mathbb{R}^n \to \mathbb{R}^n$ that pushes the initial distribution $\nu_0$ forward to the target distribution $\nu_1$, denoted as $\nu_1 = T_\sharp \nu_0$. This map is required to minimize the total transportation cost

$$\int_{\mathbb{R}^n} c\left(\underline{x}, T(\underline{x})\right)\, \nu_0(d\underline{x}) \,, \tag{4}$$

where $c(\underline{x}, \underline{y})$ is the cost of moving a unit probability mass from $\underline{x}$ to $\underline{y}$. To ensure this integral (4) is finite, both $\nu_0$ and $\nu_1$ are assumed to lie in the space of probability measures with finite second moments.

However, due to the nonlinear dependence of the transportation cost (4) on the transport map, a minimizer may not always be guaranteed [17]. To overcome this issue, the OMT problem is reformulated in terms of a joint distribution $\Pi(\nu_0, \nu_1)$ over $\mathbb{R}^n \times \mathbb{R}^n$, where the marginals are constrained to match $\nu_0$ and $\nu_1$ along their respective coordinate axes

$$\inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^n \times \mathbb{R}^n} c(\underline{x}, \underline{y})\, d\pi(\underline{x}, \underline{y}) \,. \tag{5}$$

In the Kantorovich formulation, an optimal joint distribution $\pi^*$ always exists [18]. If $\nu_0$ and $\nu_1$ are absolutely continuous with respect to the Lebesgue measure, having PDFs $\rho_0$ and $\rho_1$, respectively, and the cost is defined as the squared Euclidean distance, i.e., $c(\underline{x}, \underline{y}) = \|\underline{x} - \underline{y}\|_2^2$, then this joint distribution $\pi^*$ is also unique [19]. The square root of the minimal transportation cost in (5) thus defines the Riemannian Wasserstein metric $W_2$ on the space of probability densities, providing a natural framework for comparison, interpolation, and averaging.

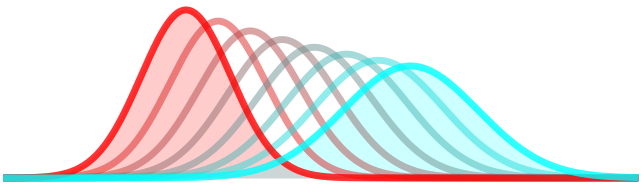

Fig. 2: Illustration of barycentric interpolation between two one-dimensional Gaussian distributions, depicted by the leftmost red and rightmost blue curves.

### B. Gaussian Marginal Distributions

When both marginals $\nu_0$ and $\nu_1$ are continuous Gaussian distributions, the problem simplifies significantly, allowing for a unique closed-form solution. Let $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ be Gaussian random vectors on a common space $\mathbb{R}^n$, with PDFs $f_{\boldsymbol{x}}(\underline{x})$ and $f_{\boldsymbol{y}}(\underline{y})$, and corresponding means $\underline{\mu}^{\boldsymbol{x}}$, $\underline{\mu}^{\boldsymbol{y}}$, and covariance matrices $\Sigma^{\boldsymbol{x}}$, $\Sigma^{\boldsymbol{y}}$. Under the squared Euclidean distance cost, the optimal transportation cost becomes

$$\begin{aligned}\mathbb{E}_{f_{\boldsymbol{x},\boldsymbol{y}}}\left\{\left\|\underline{x}-\underline{y}\right\|_2^2\right\} &= \mathbb{E}_{f_{\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}}}\left\{\left\|\underline{\tilde{x}}+\underline{\mu}^{\boldsymbol{x}}-\underline{\tilde{y}}-\underline{\mu}^{\boldsymbol{y}}\right\|_2^2\right\} \\ &= \underbrace{\mathbb{E}_{f_{\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}}}\left\{\underline{\tilde{x}}^\top\,\underline{\tilde{x}}\right\}}_{=\operatorname{trace}(\Sigma^{\boldsymbol{x}})} + \underbrace{\mathbb{E}_{f_{\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}}}\left\{\underline{\tilde{y}}^\top\,\underline{\tilde{y}}\right\}}_{=\operatorname{trace}(\Sigma^{\boldsymbol{y}})} \\ &\quad - 2\,\underbrace{\mathbb{E}_{f_{\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}}}\left\{\underline{\tilde{x}}^\top\,\underline{\tilde{y}}\right\}}_{=\operatorname{trace}(S)} + \left\|\underline{\mu}^{\boldsymbol{x}}-\underline{\mu}^{\boldsymbol{y}}\right\|_2^2 \,,\end{aligned} \tag{6}$$

where $\underline{\tilde{\boldsymbol{x}}}$, $\underline{\tilde{\boldsymbol{y}}}$ are the zero mean versions of $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$, and $S$ is the cross-covariance matrix $\mathbb{E}_{f_{\boldsymbol{x},\boldsymbol{y}}}\left\{\underline{x}\,\underline{y}^\top\right\}$. The transport cost is minimized over all feasible Gaussian joint distributions, yielding the optimal cross-covariance matrix $S^*$. This leads to

$$\max_S \left\{ \operatorname{trace}(S) \;\middle|\; \begin{bmatrix} \Sigma^{\boldsymbol{x}} & S \\ S^\top & \Sigma^{\boldsymbol{y}} \end{bmatrix} \geq 0 \right\}. \tag{7}$$

The positive semi-definiteness of the block matrix is equivalent, via the Schur Complement Condition, to the requirement

$$\begin{aligned} &0 \leq \Sigma^{\boldsymbol{x}} - S\left(\Sigma^{\boldsymbol{y}}\right)^{-1} S^\top, \\ &S \leq \left(\Sigma^{\boldsymbol{x}}\right)^{1/2} \left[\left(\Sigma^{\boldsymbol{x}}\right)^{1/2} \Sigma^{\boldsymbol{y}} \left(\Sigma^{\boldsymbol{x}}\right)^{1/2}\right]^{1/2} \left(\Sigma^{\boldsymbol{x}}\right)^{-1/2}. \end{aligned} \tag{8}$$

Consequently, the maximization in (7) is attained by the unique closed-form cross-covariance

$$S^* = \left(\Sigma^{\boldsymbol{x}}\right)^{1/2} \left[\left(\Sigma^{\boldsymbol{x}}\right)^{1/2} \Sigma^{\boldsymbol{y}} \left(\Sigma^{\boldsymbol{x}}\right)^{1/2}\right]^{1/2} \left(\Sigma^{\boldsymbol{x}}\right)^{-1/2}. \tag{9}$$

In 1D cases, the cross-covariance term $s^*$ equals $\sigma^{\boldsymbol{x}}\,\sigma^{\boldsymbol{y}}$, indicating a linear relationship between the two random variables.

Building on these results, barycentric interpolation traces the 2-Wasserstein geodesic between any two Gaussian distributions sharing the same support [20], [21]. Along this path, the interpolant minimizes the weighted sum of the Wasserstein distances to the given Gaussians and is itself Gaussian. Its mean and covariance matrix are given in closed form by

$$\begin{aligned} \underline{\mu}_{\mathrm{t}} &= (1-\mathrm{t})\,\underline{\mu}^{\boldsymbol{x}} + \mathrm{t}\,\underline{\mu}^{\boldsymbol{y}}, \;\text{weight } \mathrm{t} \in [0,1] \,, \\ \Sigma_{\mathrm{t}} &= \left(\Sigma^{\boldsymbol{x}}\right)^{-\frac{1}{2}} \left[(1-\mathrm{t})\,\Sigma^{\boldsymbol{x}} + \mathrm{t}\,\Sigma\right]^2 \left(\Sigma^{\boldsymbol{x}}\right)^{-\frac{1}{2}} \,, \\ \Sigma &= \left[\left(\Sigma^{\boldsymbol{x}}\right)^{\frac{1}{2}} \Sigma^{\boldsymbol{y}} \left(\Sigma^{\boldsymbol{x}}\right)^{\frac{1}{2}}\right]^{\frac{1}{2}} \,. \end{aligned} \tag{10}$$

See Fig. 2 for an illustrative example of this interpolation.

## IV. Method Derivation

Leveraging closed-form Wasserstein barycentric interpolation introduced above, we now turn to its use for novel measurement updates. Several conditional Gaussian state densities on discrete measurements are interpolated. This yields a smooth approximate joint measurement/state density, from which the backward inference is directly conducted.

### A. Gaussian Process Construction

Instead of multiple formal definitions of the GP method as found in the literature for diverse applications, such as regression and classification, we follow an intuitive view in [22]. A GP is a stochastic process composed of random variables indexed by time or space, where any finite subset follows a multivariate Gaussian distribution. For backward inference in measurement updates, this definition is helpful and provides the basis for modeling the hidden state's conditional Gaussian density on each measurement value along the measurement axis. Furthermore, our approach to building the GP model deviates from the conventional GP's data-driven training process, where the marginal likelihood function is maximized to tune the hyperparameters in a predefined kernel function. This optimization problem, due to its inherent non-convexity, does not ensure a globally optimal solution and is also computationally demanding.

In this work, we propose a novel approach that leverages a unique, explicit Wasserstein barycentric interpolation. This method directly derives an interpolated Gaussian PDF between any two given Gaussian distributions. For simplicity, we assume that the measurement random variable is a scalar throughout this paper. Within the joint space of measurement and state, we first initialize a set of boundaries $\{b_i\}_{i=1}^{M}$ along the measurement axis. At each boundary $b_i$, a conditional Gaussian state PDF $f_{\boldsymbol{x}}(\underline{x} \,|\, y = b_i) = \mathcal{N}(\underline{x};\, \underline{\mu}_i,\, \Sigma_i)$ is defined such that its realizations are represented in the state space. Subsequently, for each adjacent pair of conditional Gaussian distributions $\mathcal{N}(\underline{x};\, \underline{\mu}_i,\, \Sigma_i)$ and $\mathcal{N}(\underline{x};\, \underline{\mu}_{i+1},\, \Sigma_{i+1})$, we employ Wasserstein barycentric interpolation to derive a continuum of intermediate conditional Gaussians, each characterized by explicit first and second moments in (10). This approach yields an initial GP spanning the entire joint measurement/state space, assigning a conditional Gaussian state PDF to each concrete measurement. An exemplary three-dimensional illustration is provided in Figure 3, where the joint measurement/state space is augmented by an additional dimension representing the conditional density function values.

More importantly, to ensure that the derived GP accurately approximates the true underlying joint measurement/state density (3), the boundary positions must be highly flexible, continuously adapting along the measurement axis. Likewise, the conditional Gaussian distributions defined at these boundaries must dynamically adjust their parameters. Moreover, since samples from the true joint density are readily available, whether obtained deterministically or randomly, we can directly learn both the optimal boundary locations and the corresponding Gaussian parameters in a data-driven manner.

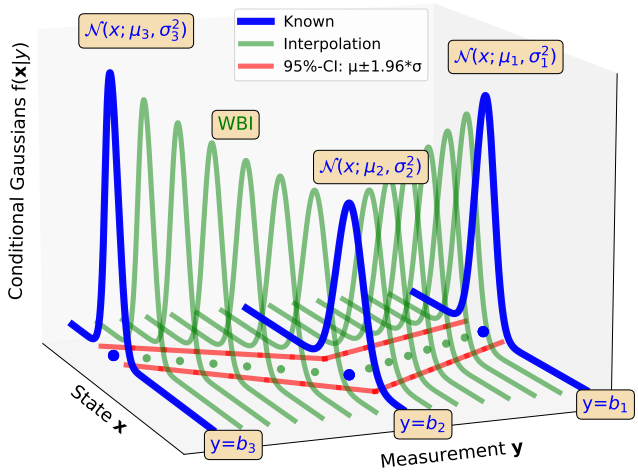


Fig. 3: An exemplary three-dimensional visualization of the constructed GP model. Along the measurement axis, three boundary points $\{b_i\}_{i=1}^{3}$ are defined, each associated with a conditional Gaussian state density $\mathcal{N}(x;\, \mu_i,\, \sigma_i^2)$ (blue). Between adjacent boundaries, the Wasserstein barycentric interpolation (WBI) generates a continuum of interpolated conditional Gaussian distributions (green) for the state, with the means (green dots) and variances (red) linearly interpolated in the one-dimensional case. For clarity, only a subset of the interpolation results is displayed.

### B. Deterministic Gaussian Sampling

Samples drawn from the true joint measurement/state density provide essential and valuable information for estimating all parameters that characterize the GP. However, direct sampling from this joint density is challenging. To overcome this difficulty, we adopt the approach in [16] by assuming that the measurement noise is Gaussian-distributed. In this way, we transform the problem into sampling from the Gaussian joint density of the prior state and measurement noise. This sampling process is simpler and more tractable. The obtained samples are then propagated through the known measurement equation to generate corresponding measurement realizations, thereby enabling indirect sampling of the true joint density.

Several techniques exist for computing deterministic Gaussian Dirac mixture approximations, such as those employed by the UKF and Gaussian Filter (GF). We adopt a sampling method based on generalized Fibonacci grids [23], which efficiently covers the joint measurement/state space with fewer, equally weighted samples at low cost. This approach is inspired by the unique spiral packing observed in sunflower heads, the 2D Fibonacci grid, and its generalization to higher dimensions [24]. Specifically, a uniformly distributed sample set in higher dimensions is transformed to deterministically approximate arbitrary multivariate Gaussian distributions.

### C. Parameter Optimization for the Gaussian Process

In this subsection, we detail the parameterization of the initial GP using samples $\{(\underline{x}_i,\, y_i)\}_{i=1}^{N}$ drawn from the true joint measurement/state density. This procedure ensures that the optimized GP closely approximates the underlying joint density. By leveraging the explicit formulation provided by Wasserstein barycentric interpolation, we introduce a comprehensive set of parameters for the GP model. The param-

eter set $\Theta = \{b_2, \ldots, b_{M+1}\} \cup \{\underline{\theta}_i\}_{i=1}^{M+2}$ comprises $M$ sequentially varying internal boundaries and $M+2$ parameter sets. These $M$ internal boundaries are not fixed and can flexibly shift along the measurement axis within the joint measurement/state space, enabling adaptive partitioning of the measurement space. Additionally, two fixed boundary points, $b_1$ and $b_{M+2}$, are defined a priori, serving as the outermost limits of the entire sample set. At each boundary position $b_i \in \{b_1^{\text{fixed}}, b_2^{\text{var.}}, \ldots, b_{M+1}^{\text{var.}}, b_{M+2}^{\text{fixed}}\}$, we define a conditional Gaussian distribution for the state vector $\underline{\boldsymbol{x}}$. Each of these Gaussian distributions $f_{\underline{\boldsymbol{x}}}(\underline{x} \,|\, b_i, \underline{\theta}_i) = \mathcal{N}(\underline{x};\, \underline{\mu}_i,\, \Sigma_i)$ is parameterized by $\underline{\theta}_i = \{\underline{\mu}_i,\, \Sigma_i\}$ for $i = 1, \ldots, M+2$.

The parameter set $\Theta$ can be estimated based on the sample set $\{(\underline{x}_i,\, y_i)\}_{i=1}^N$ using Maximum Likelihood Estimation (MLE). MLE determines the optimal parameter set $\Theta^*$ that maximizes the likelihood function $f\left(\{\underline{x}_i\}_{i=1}^N \,\middle|\, \{\underline{y}_i\}_{i=1}^N,\, \Theta\right) \overset{\text{def}}{=} \ell(\Theta)$ characterizing the probability of the observed data under the given parametric model. During backward inference, each realization of the state random vector $\{\underline{x}_i\}_{i=1}^N$ is assumed to be drawn from a conditional Gaussian distribution. More specifically, the key idea is that, for each concrete measurement $y_i \in \{\underline{y}_i\}_{i=1}^N$ that falls within the interval $[b_j^i,\, b_{j+1}^i]$, $j = 1, \ldots, M+1$, the conditional Gaussian distribution, from which $\underline{x}_i$ of $(\underline{x}_i,\, y_i)$ is drawn, is fully characterized by interpolation results. These results are obtained via the Wasserstein barycentric interpolation (10) between the two distinct Gaussian distributions defined at $b_j^i$ and $b_{j+1}^i$. The parameters of these two Gaussians are given by $\underline{\theta}_{j:j+1}^i = \{\underline{\mu}_j^i,\, \Sigma_j^i,\, \underline{\mu}_{j+1}^i,\, \Sigma_{j+1}^i\}$. In particular, we derive closed-form expressions for the interpolated mean vector $\underline{\mu}_i^{\text{WBI}}(y_i,\, \underline{b}_{j:j+1}^i,\, \underline{\theta}_{j:j+1}^i)$ and covariance matrix $\Sigma_i^{\text{WBI}}(y_i,\, \underline{b}_{j:j+1}^i,\, \underline{\theta}_{j:j+1}^i)$, respectively. In the following, the interpolation results are mathematically expressed as

$$
\begin{aligned}
y_i &\in [b_j^i,\, b_{j+1}^i]\,, \\
\underline{x}_i &\sim \mathcal{N}\big(\underline{x};\, \underline{\mu}_i^{\text{WBI}},\, \Sigma_i^{\text{WBI}}\big)\,, \\
\text{weight } \alpha &= \frac{y_i - b_j^i}{b_{j+1}^i - b_j^i} \in [0,\, 1]\,, \\
\underline{\mu}_i^{\text{WBI}}(y_i,\, \underline{b}_{j:j+1}^i,\, \underline{\theta}_{j:j+1}^i) &= (1-\alpha)\,\underline{\mu}_j^i + \alpha\,\underline{\mu}_{j+1}^i\,, \\
\Sigma_i^{\text{WBI}}(y_i,\, \underline{b}_{j:j+1}^i,\, \underline{\theta}_{j:j+1}^i) &= (\Sigma_j^i)^{\frac{-1}{2}} \left[(1-\alpha)\,\Sigma_j^i + \alpha\,\Sigma\right]^2 \\
&\qquad (\Sigma_j^i)^{\frac{-1}{2}}\,, \\
\Sigma &= \left[(\Sigma_j^i)^{\frac{1}{2}}\,\Sigma_{j+1}^i\,(\Sigma_j^i)^{\frac{1}{2}}\right]^{\frac{1}{2}}\,. \qquad (11)
\end{aligned}
$$

Building upon the setting described above, we now highlight how the optimal parameter set $\Theta^*$ can be determined using MLE in the interval $[b_1,\, b_{M+2}]$. In what follows, we first focus on the construction of the likelihood function $\ell(\Theta)$. For simplicity, we assume that the individual sample pairs $(\underline{x}_i,\, y_i)$ are independent from each other.

$$
\begin{aligned}
\Theta^* &= \arg\max_{\Theta} \ell(\Theta) = \arg\max_{\Theta} f\left(\{\underline{x}_i\}_{i=1}^N \,\middle|\, \{\underline{y}_i\}_{i=1}^N,\, \Theta\right) \\
&= \arg\max_{\Theta} \prod_{i=1}^N f\left(\underline{x}_i \,\middle|\, \underline{y}_i,\, \Theta\right) \\
&= \arg\max_{\Theta} \prod_{i=1}^N \mathcal{N}\left(\underline{x}_i;\, \underline{\mu}_i^{\text{WBI}},\, \Sigma_i^{\text{WBI}}\right) \\
&= \arg\max_{\Theta}\, \log \prod_{i=1}^N \mathcal{N}\left(\underline{x}_i;\, \underline{\mu}_i^{\text{WBI}},\, \Sigma_i^{\text{WBI}}\right) \\
&= \arg\max_{\Theta} \sum_{i=1}^N \log\, \mathcal{N}\left(\underline{x}_i;\, \underline{\mu}_i^{\text{WBI}},\, \Sigma_i^{\text{WBI}}\right) \\
&= \arg\max_{\Theta} \sum_{i=1}^N -\tfrac{1}{2}\, \log\, |\Sigma_i^{\text{WBI}}| \\
&\quad - \tfrac{1}{2}\left(\underline{x}_i - \underline{\mu}_i^{\text{WBI}}\right)^\top (\Sigma_i^{\text{WBI}})^{-1} \left(\underline{x}_i - \underline{\mu}_i^{\text{WBI}}\right) \\
&= \arg\min_{\Theta} \sum_{i=1}^N \log\, |\Sigma_i^{\text{WBI}}| \\
&\quad + \left(\underline{x}_i - \underline{\mu}_i^{\text{WBI}}\right)^\top (\Sigma_i^{\text{WBI}})^{-1} \left(\underline{x}_i - \underline{\mu}_i^{\text{WBI}}\right) \qquad (12)
\end{aligned}
$$

By transitioning from maximizing the (log) likelihood function to minimizing the negative log-likelihood, we derive an objective function that consists of two primary terms for each sample pair $(\underline{x}_i,\, y_i)$. The first term, the log-determinant term, acts as a regularization term that penalizes excessively large or small covariance matrices, thus ensuring numerical stability. The second term, the Mahalanobis distance term, quantifies the deviation of each given sample $\underline{x}_i$ from the interpolated mean vector, normalized by the corresponding interpolated covariance matrix. This term helps the estimated parameters closely match the actual data. The optimal parameter set $\Theta^*$ is obtained by minimizing the negative log-likelihood.

However, this optimization problem is generally non-convex due to the fact that small moves of boundary positions can reassign which data pairs $(\underline{x}_i,\, y_i)$ fall into which segment, unexpectedly changing the gradient with respect to $\underline{b}_{j:j+1}^i$ and $\underline{\theta}_{j:j+1}^i$. Furthermore, even though the Wasserstein interpolation formulas have closed-form expressions, matrix square roots are present, which complicates the problem and tends to break convexity with respect to $\Sigma_j^i$ and $\Sigma_{j+1}^i$. Thus, there is no guarantee that standard optimization procedures will reach the global optimum. Multiple local optima typically exist, and the algorithm may converge to any one of them.

### *D. Prior Constraints Consideration*

While the pure MLE in the previous subsection attempts to fit the sample set $\{(\underline{x}_i,\, y_i)\}_{i=1}^N$ by maximizing the likelihood $\ell(\Theta)$, it lacks a mechanism to address two critical scenarios. First, it does not provide any built-in means to extrapolate beyond the observed range. When measurements fall outside the interval $[b_1,\, b_{M+2}]$ covered by the data, the model parameters remain undefined. Second, in regions where data

are sparse or even completely absent, relying on MLE alone can yield unreliable or unstable estimates. To overcome these limitations, we impose a prior constraint on the parameter set $\Theta$. By design, this prior encodes domain knowledge and steers the model toward more plausible solutions in underrepresented or unobserved regions. Generally, we impose:

- constraints on the external segments, so that for measurements lying outside the interval populated by data, the model maintains sensible behavior, and
- smoothness constraints on the internal segments and on the transitions between these segments and the outer segments, which prevents abrupt and large jumps in the mean or variance, thereby leading to smoother interpolations.

Let $\pi(\Theta)$ denote the prior distribution over the parameter set $\Theta$, encapsulating our domain knowledge regarding the constraints above. By combining this prior with a likelihood term that evaluates the data fit (12), we obtain a penalized maximum likelihood formulation, also known as Maximum A Posteriori Estimation (MAP) estimation. This concept is inspired by the application of GP model in regression. In GP regression, a function-space prior is established by specifying a mean function and a covariance kernel. The mean function reflects our expectations regarding the function's output, while the kernel encodes its smoothness properties and correlation across the input domain. Consequently, even in regions with sparse or no data, the GP's predictions revert to the prior mean in an elegant way, with the associated predictions' uncertainty determined by the kernel.

For simplicity, let us consider a scalar-valued state. The prior constraints $\pi(\Theta)$ can then be defined as follows.

- External segments: The two external segments have $\{\mu^{\mathrm{L}},\, \sigma^{\mathrm{L}},\, \mu^{\mathrm{R}},\, \sigma^{\mathrm{R}}\}$ to parameterize their respective Gaussian distributions. Each of these parameters follows a univariate Gaussian prior,

$$\mu^{\mathrm{L}},\, \mu^{\mathrm{R}} \sim \mathcal{N}\big(\mu^*,\, \sigma_{\mathrm{tol}}^2\big),\; \sigma^{\mathrm{L}},\, \sigma^{\mathrm{R}} \sim \mathcal{N}\big(\sigma^*,\, \sigma_{\mathrm{tol}}^2\big)\;, \quad (13)$$

  where $\{\mu^*,\, \sigma_{\mathrm{tol}}\}$ are user-specified hyperparameters.
- Internal segments: For each internal segment, the mean $\mu_i$ and standard deviation $\sigma_i$ likewise follow univariate Gaussian priors. For $i = 1, \ldots, M+2$,

$$\mu_i \sim \mathcal{N}\big(\mu^*,\, \sigma_{\mathrm{tol}}^2\big),\; \sigma_i \sim \mathcal{N}\big(\sigma^*,\, \sigma_{\mathrm{tol}}^2\big)\;. \quad (14)$$

- Smoothness between adjacent internal segments: For each neighboring pair $(i,\, i+1)$, $i = 1, \ldots, M+1$, we impose

$$(\mu_{i+1} - \mu_i),\; (\sigma_{i+1} - \sigma_i) \sim \mathcal{N}\big(0,\, \sigma_{\mathrm{tol}}^2\big), \quad (15)$$

  preventing large or abrupt changes in the mean and variance across adjacent internal segments.
- Smoothness between external and internal segments:

$$\begin{aligned}(\mu_1 - \mu^{\mathrm{L}}),\; (\mu_{M+2} - \mu^{\mathrm{R}}) &\sim \mathcal{N}\big(0,\, \sigma_{\mathrm{tol}}^2\big),\\ (\sigma_1 - \sigma^{\mathrm{L}}),\; (\sigma_{M+2} - \sigma^{\mathrm{R}}) &\sim \mathcal{N}\big(0,\, \sigma_{\mathrm{tol}}^2\big)\;. \end{aligned} \quad (16)$$

- More spline-like smoothness: The second-order difference of the mean and standard deviation across consecutive segments can also be penalized. This second-order difference approximates the "local curvature" in the functions. By penalizing large curvatures, the solution becomes smoother and more spline-like. For $i = 2, \ldots, M$,

$$(\mu_{i+1} - 2\mu_i + \mu_{i-1}) \sim \mathcal{N}(0, \sigma_{\mathrm{tol}_2}^2), \quad (17)$$

$$(\sigma_{i+1} - 2\sigma_i + \sigma_{i-1}) \sim \mathcal{N}(0, \sigma_{\mathrm{tol}_2}^2)\;. \quad (18)$$

  where $\sigma_{\mathrm{tol}_2}$ is another user-defined hyperparameter.

Collecting all these components yields a product of univariate Gaussian priors over each parameter. As a result, the prior distribution $\pi(\Theta)$ provides the necessary "pull" in data-sparse or out-of-range regimes, ensuring that the solution remains well-defined and smooth. The optimal parameter set $\Theta^*$ is then obtained by minimizing the negative log-posterior

$$\Theta^* = \arg\min_{\Theta} \Big[ -\log \ell(\Theta) - \log \pi(\Theta) \Big]. \quad (19)$$

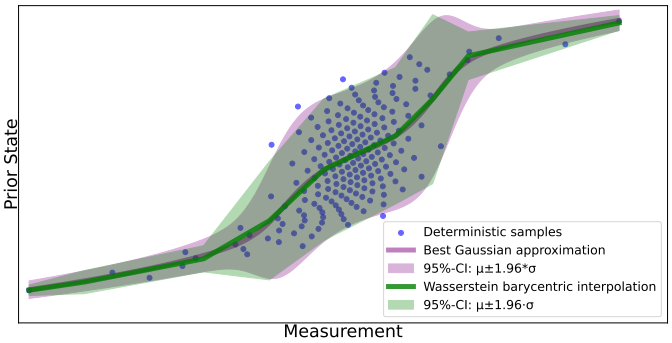


Fig. 4: Illustration of the derived Gaussian Process model (green), whose optimal parameters are determined via Maximum A Posteriori estimation from samples (blue) drawn from the true joint measurement/state density. This model achieves a non-Gaussian local approximation of the true joint density by assigning a conditional Gaussian state PDF to each concrete measurement along the horizontal axis. For comparison, we also depict the family of conditional Gaussian densities (purple) for each measurement value, with means and variances matching those of the true posterior density.

Figure 4 presents an exemplary illustration of the derived GP model with optimal parameters. This GP model provides a robust non-Gaussian local approximation of the true joint density. More specifically, it assigns a conditional Gaussian state PDF to each concrete measurement along the horizontal axis, thereby representing the approximate posterior state density. In data-sparse regions, incorporating prior information into the parameter set prevents large or abrupt variations in both the mean and variance. Moreover, our framework supports the simultaneous acquisition of multiple measurements within a single filter step. By conditioning this approximate joint density on several measurements, we can select the best result as the posterior state estimate.

## V. Evaluation

In this study, we evaluate the performance of the proposed GADF using the well-known discrete-time cubic sensor problem. This problem is defined by the measurement equation, corrupted by additive, zero-mean, state-independent Gaussian measurement noise.

$$\boldsymbol{y}_k = 0.1\, \boldsymbol{x}_k^3 + \boldsymbol{v}_k\;. \quad (20)$$

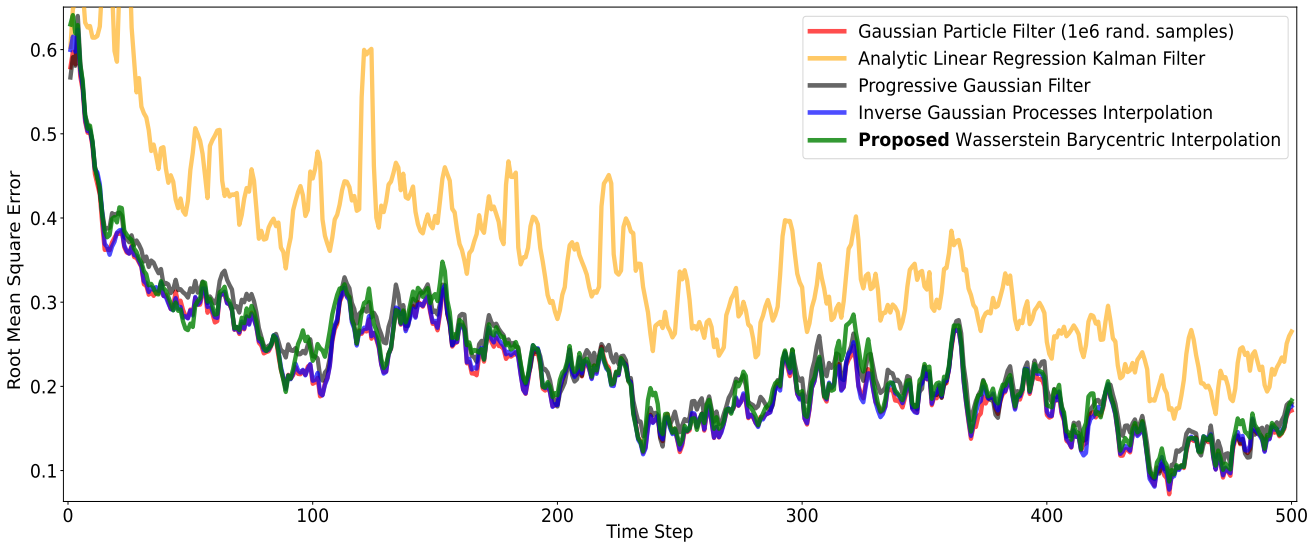


Fig. 5: Comparison of the average estimation quality across 100 independent measurement sequences over 500 time steps for the proposed algorithm and state-of-the-art methods. To improve visualization, all results are smoothed using a moving average with a window size of five time steps.

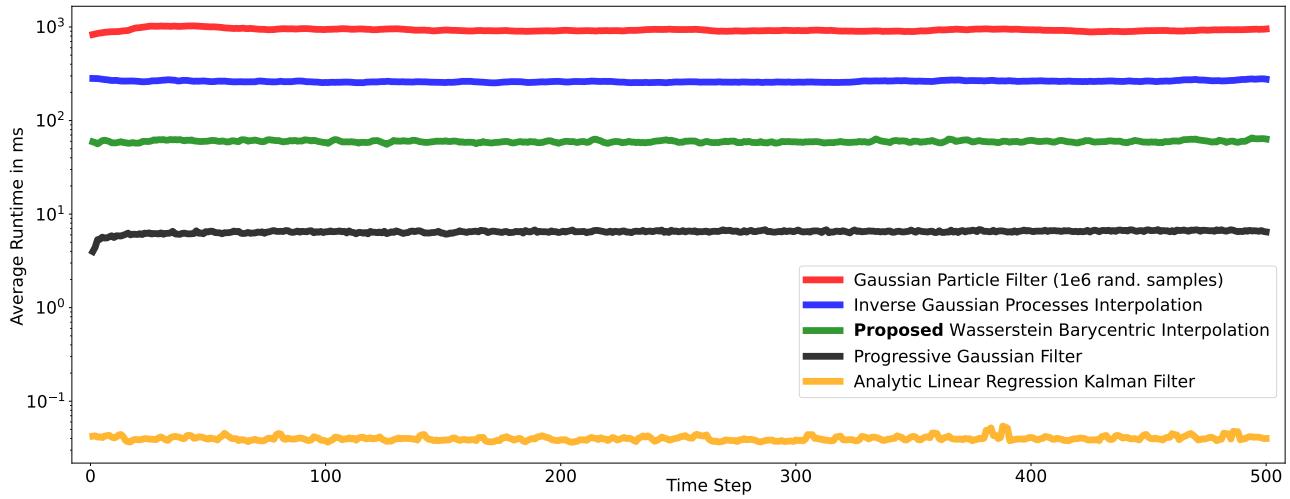


Fig. 6: Average measurement update runtimes in milliseconds of different filters at each time step.

We compare our filter against state-of-the-art GADFs in recursive hidden-state estimation. The GPF draws $10^6$ random samples in each step for moment matching Gaussian distributions. Despite its high computational cost and non-reproducible results, it converges asymptotically to the optimal MMSE estimator and thus establishes a benchmark for the estimation quality achievable by a GADF. Since the measurement equation (20) is a polynomial in the state, all moments required to parameterize the joint Gaussian density of the state and measurement under the second Gaussian assumption admit closed-form expressions [13, eqs. (14)–(16)]. For example, the expected measurement $\hat{y}_k = \mathbb{E}[\boldsymbol{y}_k]$ can be computed as

$$\hat{y}_k = 0.1 \int_{-\infty}^{\infty} x_k^3 \, f_{\boldsymbol{x}_k}^{\mathrm{P}}(x_k) \, \mathrm{d}x_k = 0.1 \, \hat{x}_k^{\mathrm{P}} + 0.3 \, \hat{x}_k^{\mathrm{P}} \, \sigma_k^{\mathrm{P}} \, .$$

Building on this analytic statistical linearization, we derive an Analytic Linear Regression Kalman Filter (ALRKF), which delivers the highest estimation quality among all LRKFs. Additionally, we analyze a PGF [15] that, like our approach, avoids linearizing the measurement model and outperforms conventional filters. We also consider the GADF based on the IGP-Interpolation, using the same deterministic sampling strategy [23] and sample size as our approach. This controlled setup allows us to compare estimation accuracy, computational efficiency, and to demonstrate whether our filter can match or exceed performance while reducing computational cost.

A Monte Carlo study, following [16], [25], was run with 100 independent state trajectories $x_k^i$, $i = 1, \ldots, 100$, over 500 time steps $k = 1, \ldots, 500$. Each trajectory evolves under the identity dynamics $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{w}_k$, subject to Gaussian noise $\boldsymbol{w}_k$ with standard deviation of 1.0. The initial state is fixed at $x_0 = 0.0$. Noisy measurements are generated via the prescribed measurement model (20) with additive Gaussian noise of standard deviation 0.3. For each trajectory,

filter estimates $\hat{x}_k^{\mathrm{e},i}$, $k = 1,\ldots,500$, $i = 1,\ldots,100$ are computed. Performance is assessed by the Root Mean Square Error (RMSE) between the estimates and the true states. All experiments were implemented in `Julia` and executed on an Intel Core 1355U CPU with 1.70 GHz.

Figure 5 demonstrates that our filter's estimates nearly coincide with the benchmark provided by the GPF, despite using only 200 deterministic samples per measurement update. While minor RMSE increases appear in certain areas, overall performance remains competitive with other state-of-the-art GADFs, offering a robust, high-quality solution for nonlinear state estimation. Furthermore, our filter easily outperforms the ALRKF by a substantial margin by avoiding any linearization of the nonlinear measurement model. Table I compares the computational complexities of various filters during measurement updates. For sample-based methods, $N$ denotes the total number of samples drawn, whether via random or deterministic sampling. $M$ represents the number of internal boundaries in our approach, which is treated as a user-defined parameter. As Fig. 6 demonstrates, although our approach incurs only a slight accuracy loss, it significantly outperforms the IGP-based method in average measurement-update runtime.

TABLE I: Computational and memory complexities of different filters.

| Filter | Sample-Based | Time Complex. | Space Complex. |
|---|---|---|---|
| **Proposed** | Yes | $O(N \cdot M)$ | $O(N + M)$ |
| IGP [16] | Yes | $O(N^3)$ | $O(N^2)$ |
| GPF [5] | Yes | $O(N)$ | $O(N)$ |
| PGF [15] | Yes | $O(N)$ | $ON)$ |
| ALRKF | No | $O(1)$ | $O(1)$ |

## VI. Conclusion

In this paper, we introduce a novel GADF for nonlinear state estimation. Our likelihood-free measurement update builds a non-Gaussian local approximation of the true joint measurement/state density based on samples from it. Specifically, a GP is defined over the joint measurement/state space via an innovative, closed-form Wasserstein barycentric interpolation, parameterizing a conditional Gaussian state PDF for each measurement value. The optimal parameters of the GP model are inferred by MAP estimation. By conditioning the resulting approximate joint density on concrete measurements, our method yields high-quality Gaussian posterior state estimates.

In 1D cases, MAP estimation easily yields analytic updates for the scalar means and variances of conditional Gaussians, either by fixing boundaries as hyperparameters or by means of an alternating optimization scheme. However, in higher dimensions this exact MAP may become computationally intractable and lose its closed-form solution. Future work will adopt a variational lower-bound formulation that smooths the objective function and enables efficient mini-batch updates.

## References


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

[2] D. Alspach and H. Sorenson. "Nonlinear Bayesian estimation using Gaussian sum approximations". In: *IEEE Transactions on Automatic Control* 17.4 (1972), pp. 439–448. DOI: 10.1109/TAC.1972.1100034.

[3] Daniel Frisch and Uwe D. Hanebeck. "Gaussian Mixture Particle Filter Step based on Method of Moments". In: *Proceedings of the 27th International Conference on Information Fusion (Fusion 2024)*. Venice, Italy, July 2024.

[4] Daniel Frisch and Uwe D. Hanebeck. "Progressive Bayesian Filtering with Coupled Gaussian and Dirac Mixtures". In: *Proceedings of the 23rd International Conference on Information Fusion (Fusion 2020)*. Virtual, July 2020.

[5] 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.

[6] Branko Ristic, Sanjeev Arulampalam, and Neil Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Norwood, MA: Artech House Publishers, 2004.

[7] Arnaud Doucet and Adam M. Johansen. "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later". In: *The Oxford Handbook of Nonlinear Filtering*. Ed. by Dan Crisan and Boris Rozovskii. Oxford: Oxford University Press, 2009, pp. 656–704.

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

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

[10] 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.

[11] 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. DOI: 10.1080/00207170410001713462.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] Cédric Villani. "Topics in optimal transportation." In: *OR/MS Today* 30.3 (2003), pp. 66–67.

[18] Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2008.

[19] Yann Brenier. "Polar factorization and monotone rearrangement of vector-valued functions". In: *Communications on pure and applied mathematics* 44.4 (1991), pp. 375–417.

[20] Robert J McCann. "A convexity principle for interacting gases". In: *Advances in mathematics* 128.1 (1997), pp. 153–179.

[21] Clark R Givens and Rae Michael Shortt. "A class of Wasserstein metrics for probability distributions." In: *Michigan Mathematical Journal* 31.2 (1984), pp. 231–240.

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

[23] 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.

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

[25] Uwe D. Hanebeck and Jannik Steinbring. "Progressive Gaussian Filtering Based on Dirac Mixture Approximations". In: *Proceedings of the 15th International Conference on Information Fusion (Fusion 2012)*. Singapore, July 2012.