Gaussian Mixture Particle Filter Step based on Method of Moments

Daniel Frisch and **Uwe D. Hanebeck**
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
daniel.frisch@kit.edu, uwe.hanebeck@kit.edu

***Abstract*—We propose a novel update step of a Gaussian mixture particle filter for nonlinear state estimation. The update procedure works as follows: First, unweighted samples are drawn in an optimal deterministic sense from a prior Gaussian mixture. These samples are then assigned weights from the likelihood function, and we compute higher-order moments from this sample-based posterior. These moment approximations converge with L^{-1} instead of $L^{-1/2}$ as our samples are optimal deterministic. Finally, the continuous posterior approximation is determined as the Gaussian mixture that has minimal Fisher information under the constraint of having the aforementioned moments. To achieve this, we employ a closed-form solution of the Fisher information that involves Gaussian root mixture densities.**

***Index Terms*—Bayesian inference, nonlinear filtering, Fisher information, Gaussian sum filter, Gaussian mixture filter, deterministic sampling, Monte Carlo, quasi-Monte Carlo, density approximation,**

I. Introduction

A. Context

We consider the classical general state estimation problem, in particular, the filter step or measurement update, as Bayesian inference: Given prior information is fused with new measurement information through a nonlinear, uncertain measurement model.

B. Considered Problem

The Bayesian inference can be solved in closed form only for limited types of measurement models and prior densities, called conjugate priors [1]. For example, we have the Kalman filter [2] for linear measurement models in conjunction with Gaussian densities or mean and covariance as state representation, and the Wonham filter [3] for discrete state spaces.

In all other cases besides the special cases mentioned, the number of parameters accurately representing the resulting density increases over time with every new measurement, or there is no closed-form representation at all. On the upside, the actual shape of the true state density does not get more and more complicated since uncertainty is added in the alternating prediction step, making it more "Gaussian-like", according to the central limit theorem. Therefore it is a good idea to employ approximate state estimation.

C. State of Art

Linear Regression Kalman Filters: One popular class of filters are the Linear Regression Kalman filters (LRKFs). Thereby, the system models are always linearized in some way to apply the Kalman filter equations, the linear filter for linear systems. For example, the Extended Kalman Filter (EKF) linearizes the models with tangents, making it susceptible to the choice of the linearization point. Alternatively, one may choose, in an s-dimensional problem, $L = s + 1$ samples forming a simplex that supports a secant as linearization [4]. The supporting points' spacing can be chosen according to the extent of the prior uncertainty.

Higher numbers of samples bring us to the actual regression. Very popular is the Unscented Kalman Filter (UKF) with $L = 2s + 1$ samples [5]. More advanced sampling methods can give arbitrary numbers of deterministic samples, enabling an adaption to the induced nonlinearity of the problem at hand. The Smart Sampling Kalman Filter (S^2KF) [6], [7], [8] uses a library of samples obtained by matching their Localized Cumulative Distribution (LCD) [9] to the LCD of the Gaussian density [10], [11]. Alternatively, uniform low-discrepancy point sets, like the Fibonacci lattice, can be transformed to certain densities, yielding deterministic Gaussian samples [12], [13].

All these methods have in common the 2nd Gaussian assumption. That is, the probabilistic measurement equation is represented as multivariate Gaussian density in the joint state and measurement space. This facilitates very much the inversion of the linearized measurement model but also limits the expressiveness of the state-measurement relationship.

Particle Filters: Particle filters use weighted samples or Dirac mixture representations of the state density which can be processed by nonlinear model equations much easier than continuous densities. A major advantage is that no linearization of the relation between state and measurement takes place. We start from the Bayesian paradigm (as opposed to the linear filter for linear systems), again introducing approximations to make it tractable: For a particle-based prior approximation, samples are re-weighted with the likelihood function in order to obtain the posterior density (in particle form). The major problem of particle filters is particle starvation: the weights of most particles tend to go to zero over time, reducing the spatial expressiveness of the remaining particles. Therefore, a central element of every particle filter is a resampling procedure, where weighted samples are replaced with unweighted ones.

Assumed Density Particle Filters: Assumed density particle filters perform resampling by continually fitting a continuous

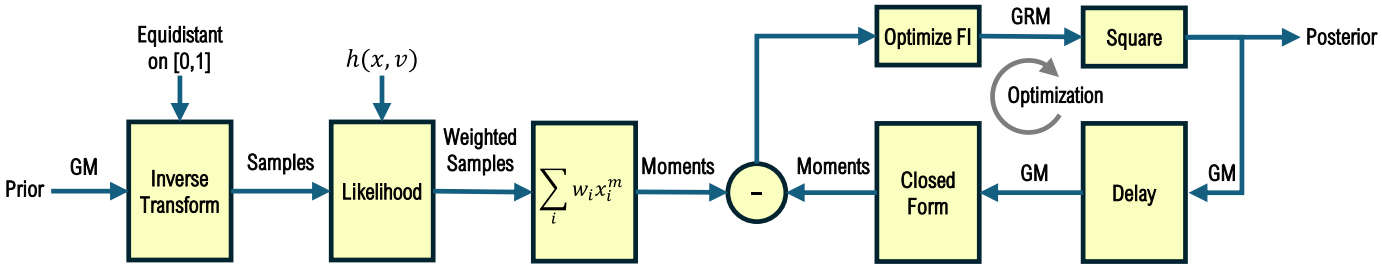

Fig. 1: Block diagram of the proposed filter step.

density to the weighted samples, and then drawing new, unweighted samples from that continuous density. For example, the Gaussian Particle Filter (GPF) [14], [15] re-approximates the weighted samples with a Gaussian, which can be done very efficiently via the empirical mean and covariance. If too many samples are assigned negligible weights, the progressive GPF [16], [17] provides a remedy: The filter step is subdivided into several progression steps where only part of the measurement information is introduced, respectively. The PGF 42 [18] contains an additional twist and can be applied even when the measurement model cannot be transformed to a likelihood function.

Even without the second assumption on the joint density of state and measurement, the assumption of a Gaussian posterior still limits the expressiveness of the posterior to unimodality and Gaussian shape. This restriction can be removed by switching to the Gaussian Sum Particle Filter (GSPF) [19]. A progressive GSPF has been proposed in [20]. It can represent non-Gaussian and multimodal densities. The major problem of the GSPF is fitting a Gaussian mixture (GM) to the weighted samples [21]. Even k-means clustering is NP-hard [22], so finding algorithms that quickly compute a local optimum is very important. In the present work, we explore an alternative method for fitting a GM: The method of moments.

II. Key Idea

In re-approximating probability densities, regularization is a critical component. Of the infinitely many possible densities, we want to select the one that includes all information from the given data but introduces no information beyond that. We propose to regularize via minimizing the Fisher information (FI), an information-theoretic measure for the roughness of densities. A simple expression for the FI is available for root densities [23], [24], and a closed-form expression exists for Gaussian root mixtures (20). We present a GSPF where the GM reapproximation of the posterior Dirac mixture (DM), i.e., the likelihood-reweighted samples, is done via minimizing the FI under the constraints of matching a prespecified set of higher-order moments between posterior DM and the posterior GM approximation.

In particular, the proposed GSPF works as follows: From the prior GM representation, we compute L unweighted optimal deterministic samples x_i via inverse transform sampling. We determine moments (e.g., E_1, E_2, E_3, E_4) of the likelihood-reweighted samples. Then we compute a Gaussian root mixture (GRM) by minimizing the FI, i.e., solving a constrained optimization problem, additionally enforcing the just mentioned moments as constraints on the corresponding GM. The GM is thereby obtained by squaring the GRM. The result of the optimization is the desired posterior GM. See Figure 1 for a block diagram.

For numerical reasons, we can only calculate moments up to a certain (low) order that do not uniquely characterize the posterior GM. Therefore, we employ the regularization via minimizing the FI.

III. Gaussian (Root) Mixtures and Fisher Information

This work focuses on the 1D case. For some of its aspects, higher-dimensional generalization is not straightforward and will be examined in future works. In particular, higher-order moments involve m-dimensional tensors, and optimal deterministic GM sampling is still subject to active research in 2D and higher.

A. Gaussian Density

The Gaussian density is defined as

$$\mathcal{N}(x; \mu, \Sigma) = \frac{1}{\sqrt{2\pi\Sigma}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\Sigma}\right\} , \tag{1}$$

with mean μ and variance Σ. The cumulative density function (CDF) F of a density f is defined as

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t , \tag{2}$$

and for the Gaussian density, it is

$$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2\,\Sigma}}\right)\right] , \tag{3}$$

with the error function

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp\{-t^2\}\,\mathrm{d}t . \tag{4}$$

We will also need the Gaussian derivative, which is

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{N}(x; \mu, \Sigma) = -\frac{x-\mu}{\Sigma}\mathcal{N}(x; \mu, \Sigma) . \tag{5}$$

B. Gaussian Root Mixture

We define the GRM $r(x)$

$$r(x) = \sum_{i=1}^{R} v_i \, \mathcal{N}(x; \rho_i, P_i) \tag{6}$$

with weights v_i, mean values ρ_i, and variances P_i. As it can be negative and does not sum to ones, the GRM r is not itself a density, but its square f

$$f(x) = r^2(x) \tag{7}$$

is a valid density, the corresponding GM.

C. Gaussian Mixture From Gaussian Root Mixture

The conversion of a GRM $r(x)$ to its corresponding GM $f(x)$ is

$$f(x) = r^2(x) = \sum_{i=1}^{R} \sum_{j=1}^{R} w_{i,j} \, \mathcal{N}(x; \mu_{i,j}, \Sigma_{i,j}) \ , \tag{8}$$

with weights $w_{i,j}$

$$w_{i,j} = v_i v_j \, \mathcal{N}(0; \rho_i - \rho_j, P_i + P_j) \ , \tag{9}$$

mean values $\mu_{i,j}$

$$\mu_{i,j} = \frac{P_j \rho_i + P_i \rho_j}{P_i + P_j} \ , \tag{10}$$

and covariances $\Sigma_{i,j}$

$$\Sigma_{i,j} = \frac{P_i P_j}{P_i + P_j} \ . \tag{11}$$

In summary, the GM f corresponding to GRM r can be represented as

$$f(x) = \sum_{k=1}^{K} w_k \, \mathcal{N}(x; \mu_k, \Sigma_k) \ , \tag{12}$$

with $K = R^2$. Exploiting symmetry in (9) to (11), it can be reduced to $K = R \cdot (R+1)/2$ components. Furthermore, components with small weights w_k can be removed. See Figure 2 for a visualization of a GRM and corresponding GM.

D. Moments

The central moments of a Gaussian density $\mathcal{N}(x; \mu, \Sigma)$ are given by

$$C_m = \mathrm{E}\{(\boldsymbol{x} - \mu)^m\} = \begin{cases} 0 \ , & m \text{ odd} \ , \\ \Sigma^{\frac{m}{2}} \cdot \prod\limits_{\substack{j=1, \\ j \text{ odd}}}^{m-1} j \ , & m \text{ even} \ , \end{cases} \tag{13}$$

and the non-central moments by

$$E_m = \mathrm{E}\{x^m\} = \sum_{i=0}^{m} \binom{m}{i} C_{m-i} \, \mu^i \ , \tag{14}$$

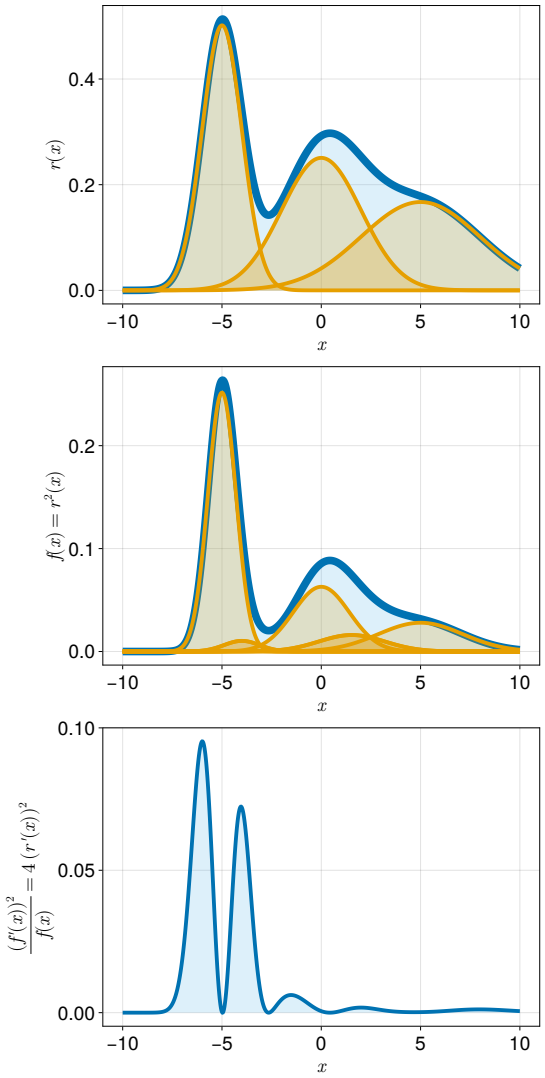

Fig. 2: GRM $r(x)$ with $R = 3$ components (top), and its corresponding GM $f(x) = r^2(x)$ with 9 components (middle). In addition, we show the "score function" (bottom) whose integral is the FI (17).

for $m \in \mathbb{N}_0$, respectively. The first central moments C_m and non-central moments E_m are therefore given by

$$\begin{aligned} C_0 &= 1 \ , & E_0 &= 1 \ , \\ C_1 &= 0 \ , & E_1 &= \mu \ , \\ C_2 &= \Sigma \ , & E_2 &= \mu^2 + \Sigma \ , \\ C_3 &= 0 \ , & E_3 &= \mu^3 + 3\mu\Sigma \ , \\ C_4 &= 3\,\Sigma^2 \ , & E_4 &= \mu^4 + 6\mu^2\,\Sigma + 3\Sigma^2 \ , \\ C_5 &= 0 \ , & E_5 &= \mu^5 + 10\mu^3\Sigma + 15\mu\Sigma^2 \ , \\ C_6 &= 15\,\Sigma^3 \ , & E_6 &= \mu^6 + 15\mu^4\Sigma + 45\mu^2\Sigma^2 + 15\Sigma^3 \ . \end{aligned}$$

Moments of a GM (8) are

$$E_m = \sum_{k=1}^{K} w_k \, E_m^{(k)} \ , \tag{15}$$

where the $E_m^{(k)}$ are the moments of the individual mixture components $\mathcal{N}(x; \mu_k, \Sigma_k)$.

E. Fisher Information

The FI is defined as [23]

$$I_F(f) = \int_{x\in\mathbb{R}, f>0} \frac{\big(f'(x)\big)^2}{f(x)}\,\mathrm{d}x \tag{16}$$

$$= I_F^{\mathcal{R}}(r) = 4\int_{x\in\mathbb{R}} \big(r'(x)\big)^2\,\mathrm{d}x \ , \tag{17}$$

where its representation (17) in terms of the GRM $r = \sqrt{f}$ is much more convenient and allows for a closed-form solution. The integrand of the FI is called the "score function", see Figure 2 for a visualization.

We obtain for the derivative of the GRM (6)

$$r'(x) = \sum_{i=1}^{R}\left(-\frac{x-\rho_i}{P_i}\right) v_i\,\mathcal{N}(x; \rho_i, P_i) \ , \tag{18}$$

and the square of that

$$\begin{aligned}&\big(r'(x)\big)^2\\ &= \sum_{i=1}^{R}\sum_{j=1}^{R} \frac{x-\rho_i}{P_i}\frac{x-\rho_j}{P_j} w_{i,j}\,\mathcal{N}(x; \mu_{i,j}, \Sigma_{i,j}) \qquad (19)\\ &= \sum_{i=1}^{R}\sum_{j=1}^{R} \frac{x^2 - (\rho_i+\rho_j)x + \rho_i\rho_j}{P_iP_j} w_{i,j}\,\mathcal{N}(x; \mu_{i,j}, \Sigma_{i,j}) \ .\end{aligned}$$

Note the mixed notation using both GRM parameters (v, ρ, P) as well as GM parameters (w, μ, Σ) from the corresponding GM, see Section III-B. The FI (17) can now be written in closed form

$$\begin{aligned}&I_F(f) = I_F^{\mathcal{R}}(r) \qquad (20)\\ &= 4\sum_{i=1}^{R}\sum_{j=1}^{R} \frac{w_{i,j}}{P_iP_j}\Big(E_2^{(i,j)} - (\rho_i+\rho_j)E_1^{(i,j)} + \rho_i\rho_j\Big) \ ,\end{aligned}$$

again using mixed notation employing parameters from the GRM and the corresponding GM. Additionally, we need the GM moments

$$E_m^{(i,j)} = \int_{-\infty}^{\infty} x^m\,\mathcal{N}(x; \mu_{i,j}, \Sigma_{i,j})\,\mathrm{d}x \tag{21}$$

that are available in closed form as well, see (14).

F. Gaussian Mixture Fitting

We can now specifically compute a GRM r that, if squared, yields a GM $\hat{f}$ that exhibits certain given properties but contains, loosely speaking, as little additional (Fisher) information as possible. We do this by solving the nonlinear constrained optimization problem

$$r = \underset{\{v_i,\rho_i,P_i\}_{i=1}^{R}}{\arg\min}\ I_F^{\mathcal{R}}(r) \tag{22}$$

$$\text{subject to} \quad \sum_{i=1}^{R}\sum_{i=1}^{R} w_{i,j} = 1 \ , \tag{23}$$

$$w_{i,j} \geq 0 \ , \tag{24}$$

$$P_i \geq 0 \ , \tag{25}$$

where the desired properties of f are then enforced by additional equality constraints like moments E_m of f. In other words, this method fits a GM (via its corresponding GRM) precisely to the given constraints while otherwise keeping it as uninformative as possible. We employed the implementation of the interior point Newton method (IPNewton) in Julia's Optim library [25].

G. Numerical Improvements

The positivity constraint for the GM weights can be imposed more efficiently than enforcing (24) by the optimizer. Note that the GM $w_{i,j}$ weights are computed as products of all pairs of the GRM weights v_i, times a positive factor (9). Therefore, for all $w_{i,j}$ to be positive, the v_i have to be either all positive or all negative. Thus, we can replace $w_{i,j} \geq 0$ with $v_i \geq 0$ which is a box constraint that is much easier for the solver than an inequality constraint on a nonlinear equation. Furthermore, we can avoid the constraint altogether by equivalently replacing it with a transformation: optimize surrogate variables $\tilde{v}_i$ and take $v_i = |\tilde{v}_i|$. In the same way, we can replace the box constraint (25) via optimizing over surrogates $\tilde{P}_i$ and taking $P_i = \left|\tilde{P}_i\right|$.

The equality constraint for the density normalization (24) can be fulfilled in closed form by proper scaling as follows: optimize the unconstrained surrogate parameters $\tilde{v}_i$, then normalize them according to

$$v_i = \frac{\tilde{v}_i}{\sqrt{\sum_{i=1}^{R}\sum_{i=1}^{R} \tilde{v}_i\tilde{v}_j\,\mathcal{N}(0; \rho_i-\rho_j, \Sigma_i+\Sigma_j)}} \ . \tag{26}$$

Then one of the $\tilde{v}_i$ can be excluded from optimization and set to an arbitrary constant value.

Usually, one of the additional custom constraints is a first-moment constraint

$$\int_{-\infty}^{\infty} x\cdot f(x)\,\mathrm{d}x \overset{!}{=} E_1^f \ . \tag{27}$$

It can be replaced equivalently with a closed-form transformation as follows: optimize the surrogate parameters $\tilde{\rho}_i$ and translate them according to

$$\rho = \tilde{\rho} + E_1^f - \sum_{i=1}^{R}\sum_{i=1}^{R} w_{i,j}\frac{P_j\tilde{\rho}_i + P_i\tilde{\rho}_j}{P_i+P_j} \ . \tag{28}$$

Again, one of the $\tilde{\rho}_i$ can then be set to an arbitrary constant value. Note that the GRM weights normalization (26) is not affected by this transform, since $\rho_i - \rho_j = \tilde{\rho}_i - \tilde{\rho}_j$.

Finally, a possible second-order non-central moment constraint

$$\int_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{d}x \stackrel{!}{=} E_2^f \tag{29}$$

can also be replaced with a suitable closed-form transformation. We have (15)

$$E_2^f = \sum_{i=1}^{R}\sum_{i=1}^{R} w_{i,j} \cdot \left(\mu_{i,j}^2 + \Sigma_{i,j}\right) \ , \tag{30}$$

and from this, we find that the transformation

$$P = \tilde{P} \cdot \frac{E_2^f - \sum_{i=1}^{R}\sum_{i=1}^{R} w_{i,j} \cdot \left(\frac{\tilde{P}_j\rho_i + \tilde{P}_i\rho_j}{\tilde{P}_i+\tilde{P}_j}\right)^2}{\sum_{i=1}^{R}\sum_{i=1}^{R} w_{i,j} \cdot \frac{\tilde{P}_i\tilde{P}_j}{\tilde{P}_i+\tilde{P}_j}} \ , \tag{31}$$

with P being the GRM component variance and $\tilde{P}$ its optimization surrogate, can be used instead of the constraint. Again, one of the $\tilde{P}_i$ can then be set to an arbitrary constant value. However, note that this E_2^f-normalization changes the $\Sigma_i + \Sigma_j$ in (26), so another normalization of the v_i and in turn the ρ_i and in turn the P_i becomes necessary and so on, resulting in a gradient-free, expectation–maximization (EM) style optimization procedure to fulfill the E_0^f, E_1^f, and E_2^f constraints simultaneously.

IV. Optimal Deterministic Gaussian Mixture Sampling

In this section, we describe the computation of moments

$$E_m(f^{\mathrm{e}}) = \int_{-\infty}^{\infty} x^m f^{\mathrm{e}}(x)\,\mathrm{d}x \tag{32}$$

from the posterior densitiy $f^{\mathrm{e}}(x)$ that are used to propagate information from the posterior DM to the posterior GM. The posterior is defined due to Bayes' theorem

$$f^{\mathrm{e}}(x) = c \cdot \Lambda(x) \cdot f^{\mathrm{p}}(x) \ , \tag{33}$$

with prior density $f^{\mathrm{p}}(x)$, the likelihood

$$\Lambda(x) = f(\hat{y} \mid x) \ , \tag{34}$$

where $\hat{y}$ is the measurement, and normalization factor c. Here, we assume that the likelihood can be derived from the measurement equation $y = h(x, v)$ in closed form. Calculating the analytic likelihood function Λ corresponding to the measurement equation is always possible in the additive noise case. Otherwise, it depends on the specific form of $h(\cdot,\cdot)$.

In summary, we have for the posterior moments

$$E_m(f^{\mathrm{e}}) = c \cdot \int_{-\infty}^{\infty} x^m \cdot \Lambda(x) \cdot f^{\mathrm{p}}(x)\,\mathrm{d}x \ . \tag{35}$$

In the Assumed Density Particle Filter approach, we first draw L unweighted samples x_i from the prior density f^{p} and then assign weights w_i from the likelihood function

$$w_i = \Lambda(x_i) \ , \quad i \in \{1, 2, \ldots, L\} \ , \tag{36}$$

resulting in a DM representation of the posterior

$$f^{\mathrm{e,DM}}(x) = c \sum_{i=1}^{L} w_i\,\delta(x - x_i) \ , \tag{37}$$

with $c = \left(\sum_{i=1}^{L} w_i\right)^{-1}$ and the Dirac delta function $\delta(x)$. This gives a tractable approximation of (35)

$$E_m\left(f^{\mathrm{e,DM}}\right) = c \sum_{i=1}^{L} w_i\, x_i^m \ . \tag{38}$$

The quadrature error

$$\epsilon = \left|E_m\left(f^{\mathrm{e,DM}}\right) - E_m(f^{\mathrm{e}})\right| \tag{39}$$

now depends on the way the prior samples x_i are chosen, which we will examine in the following.

To analyze the integration error (39), we transform the moment-computing integral (35) via the substitution

$$F(x) \to u \ , \qquad x \in \mathbb{R} \ , \quad u \in [0, 1] \ , \tag{40}$$

where F is the CDF of f. This gives us the equivalent problem

$$E_m(f^{\mathrm{e}}) = c \cdot \int_0^1 \left(F^{-1}(u)\right)^m \cdot \Lambda\left(F^{-1}(u)\right)\mathrm{d}u \ . \tag{41}$$

The corresponding sample-based quadrature is

$$E_m\left(f^{\mathrm{e,DM}}\right) = \sum_{i=1}^{L} g(u_i) \ , \tag{42}$$

where the u_i are uniformly distributed in $[0, 1]$, and

$$g(u) = c \cdot \left(F^{-1}(u)\right)^m \cdot \Lambda\left(F^{-1}(u)\right) \ , \tag{43}$$
$$c = \left(\sum_{i=1}^{L} \Lambda\left(F^{-1}(u_i)\right)\right)^{-1} \ .$$

Since we now integrate over the unit interval $[0, 1]$, we can directly apply standard methods to analyze the integration error.

A. Monte Carlo Quadrature

For samples that are independent identically distributed (iid), the sample-based approximate integration (42) is called a Monte Carlo method. According to the Central Limit Theorem (CLT) [26, Sec. 2.1], [27, p. 244], the integration error

$$\epsilon = \left|\int_{-\infty}^{\infty} g(u)\,\mathrm{d}u - \sum_{i=1}^{L} g(u_i)\right| \tag{44}$$

is characterized via

$$\epsilon \sim \mathcal{N}\left(\epsilon; 0, \Sigma \cdot L^{-1/2}\right) \ , \tag{45}$$

with $\Sigma = \int_0^1 \left(g(u) - \left(\int_0^1 g(\tilde{u})\,\mathrm{d}\tilde{u}\right)\right)^2 \mathrm{d}u$. What depends on L is the standard deviation of the integration error (45). This tells us that the moment approximations converge at a rate of only $L^{-1/2}$.

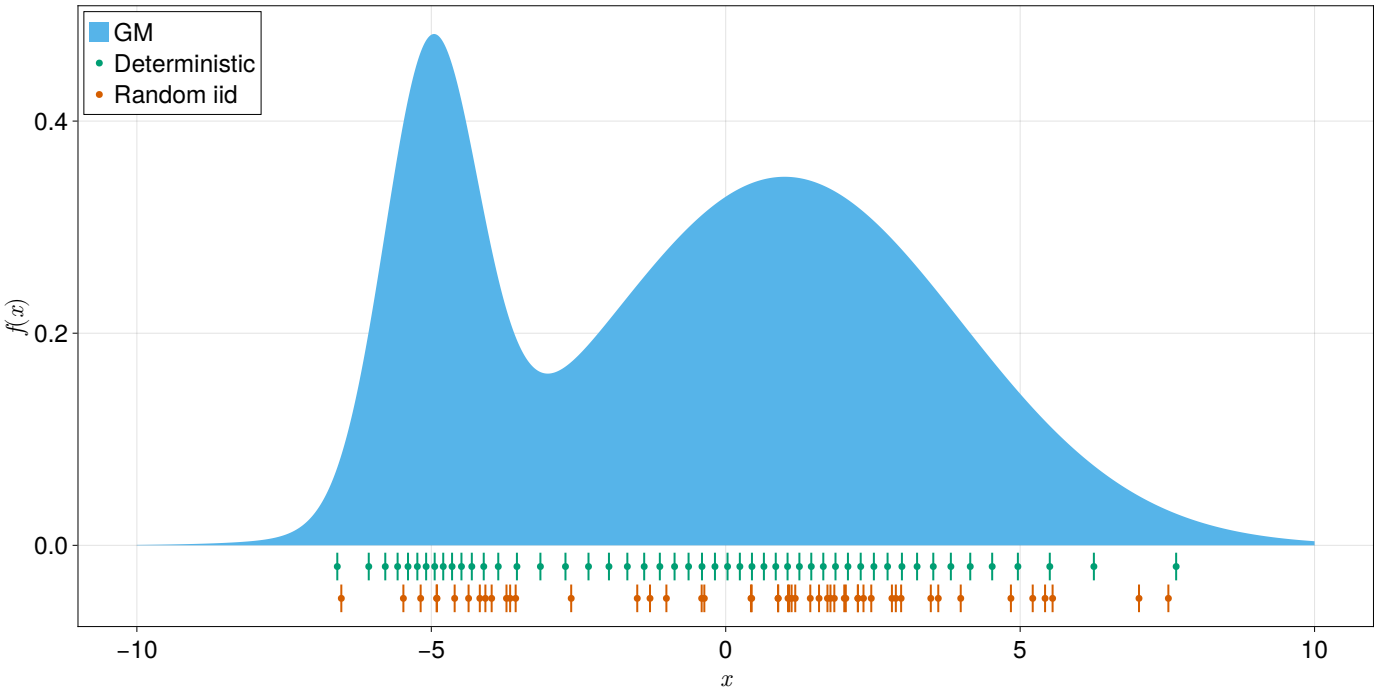

Fig. 3: Sampling of GM (blue) with two components. Proposed deterministic samples (green) and conventional random iid samples (red). $L = 50$ samples are drawn, respectively.

B. Quasi-Monte Carlo Quadrature

We propose not to use uniform unweighted iid, but uniform unweighted equidistant samples

$$u_i = \frac{2i-1}{2L} \quad , \quad i \in \{1, 2, \dots, L\} \quad . \tag{46}$$

The benefit of these deterministic samples is quantified by the Koksma-Hlawka inequality that gives an upper bound on the integration error [28, Sec. 2], [29]

$$\epsilon \leq \operatorname{discr}(\{u_i\}) \cdot V(g) \quad , \tag{47}$$

where

$$V(g) = \int_0^1 |g'(t)| \, \mathrm{d}t \quad , \tag{48}$$

$$\operatorname{discr}(\{u_i\}) = \sup_{t \in [0,1]} \left| x - F_{\{u_i\}}(x) \right| \quad , \tag{49}$$

with $F_{\{u_i\}}$ being the CDF of the samples, a staircase function. For equidistant samples as proposed, we have

$$\operatorname{discr}(\{u_i\}) = L^{-1} \quad . \tag{50}$$

Therefore, the moment approximations now converge at a rate of L^{-1}.

C. Inverse Transform Sampling

We can formulate the quasi-Monte Carlo quadrature equivalently via inverse-transforming the equidistant samples u_i, yielding optimal deterministic non-uniform samples

$$x_i^f = F^{-1}(u_i) \quad \sim f(x) \quad . \tag{51}$$

The integrand $g(u)$ from (43) then simplifies to

$$g(x) = c \cdot (x)^m \cdot \Lambda(x) \quad , \qquad c = \left(\sum_{i=1}^{L} \Lambda(x_i) \right)^{-1} \quad . \tag{52}$$

Thus, just by switching from random samples to optimal deterministic samples, we improved the convergence rate from $L^{-1/2}$ to L^{-1}.

D. Optimal Deterministic Gaussian Sum Sampling

Applying this to GM densities, we begin with equidistant uniform samples u_i (46) and find $x_i^f \sim f$ such that

$$u_i = F\left(x_i^f\right) \quad . \tag{53}$$

The CDF F of a GM f is available in closed form

$$F(x) = \int_{-\infty}^{x} \sum_{k=1}^{K} \frac{w_k}{\sqrt{2\pi\Sigma_k}} \exp\left\{ -\frac{1}{2} \frac{(t-\mu_k)^2}{\Sigma_k} \right\} \mathrm{d}t \tag{54}$$

$$= \sum_{k=1}^{K} \frac{w_k}{2} \left[1 + \operatorname{erf}\left(\frac{x-\mu_k}{\sqrt{2\,\Sigma_k}} \right) \right] \quad . \tag{55}$$

In total, we have

$$\frac{2i-1}{2L} - \sum_{k=1}^{K} \frac{w_k}{2} \left[1 + \operatorname{erf}\left(\frac{x_i^f - \mu_k}{\sqrt{2\,\Sigma_k}} \right) \right] \stackrel{!}{=} 0 \quad . \tag{56}$$

We solve this equation for x_i^f, $i = 1, \dots, L$, with an efficient bisection-based root finding algorithm [30]. This is easily possible because F is monotonous. See Figure 3 for a visual example.

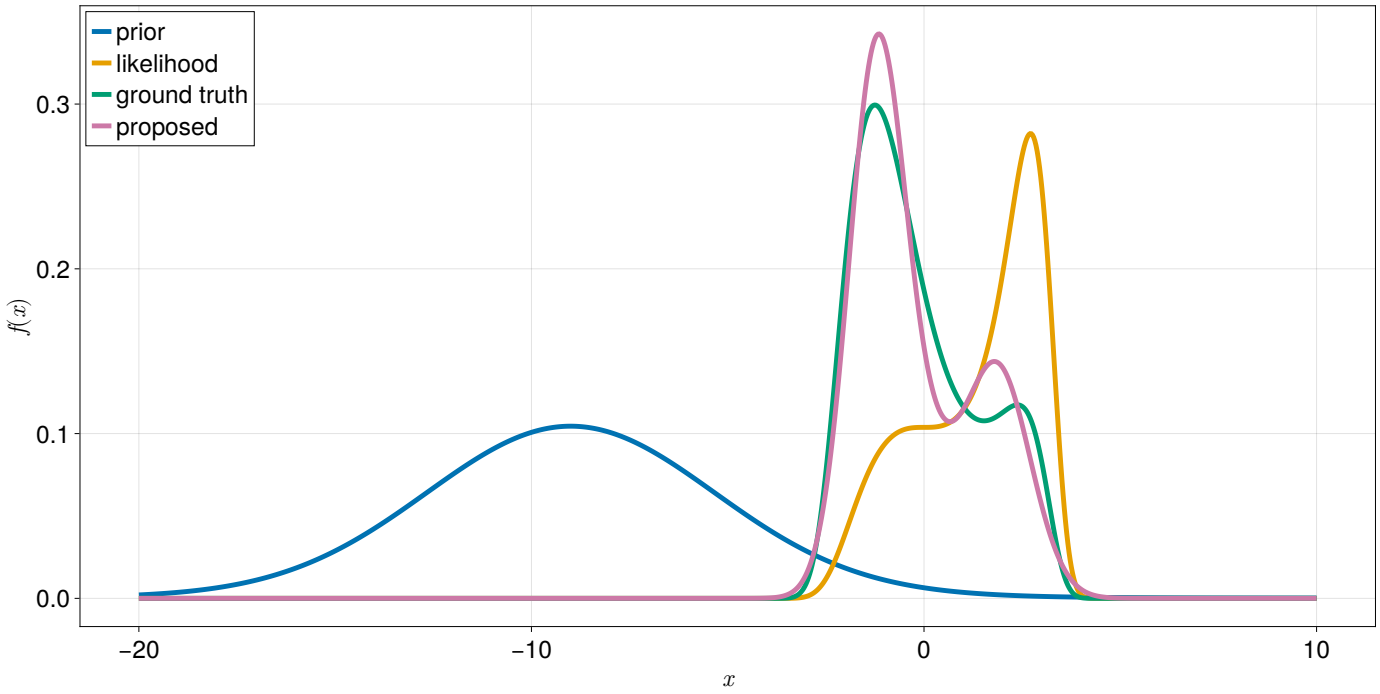

Fig. 4: Proposed filter step based on the nonlinear measurement model described in Section V. Shown are the prior (blue), the likelihood (yellow), the ground truth posterior (green), and the proposed GRM-GM posterior (pink).

V. Demonstration

We define a nonlinear measurement model

$$y = h(x) + v \ , \qquad h(x) = \frac{1}{10}x^3 - x \tag{57}$$

with measurement noise

$$v \sim \mathcal{N}(v; \mu_v = 0, \Sigma_v = 2) \ . \tag{58}$$

This results in the likelihood

$$\mathcal{N}(x; \hat{y} - h(x), \Sigma_v) \ , \tag{59}$$

where we choose $\hat{y} = 2$ in the example. Furthermore, we choose some prior GRM with three components and perform the proposed filter step. The result is shown in Figure 4.

VI. Conclusion

We present a Bayesian update step for GMs. It computes posterior moments via optimal deterministic samples. We show that the choice of optimal deterministic samples over iid samples improves the Monte Carlo (MC) convergence of e.g. moment approximations from $L^{-1/2}$ to L^{-1}.

Several posterior moments serve as constraints to fit a posterior GM in the next step. As the moments do not contain enough information to characterize the posterior GM fully, we regularize by minimizing the FI. As the FI can be expressed more easily in terms of the GRM [23] (and not in terms of the GM), we introduce an internal GRM representation that is used for FI optimization. We provide closed-form solutions for the FI of GRMs. Constrained optimization of the FI provides an elegant way to "interpolate" a continuous mixture density from a few given constraints like moments.

Source code of our implementation in the Julia programming language will be published alongside this paper on IEEE Xplore through Code Ocean.

References

[1] Daniel Fink. *A Compendium of Conjugate Priors*. Tech. rep. Bozeman, MT 59717: Montana State Univeristy, May 1997.

[2] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Journal of Basic Engineering* 82.1 (Mar. 1960), pp. 35–45. ISSN: 0021-9223.

[3] W. M. Wonham. "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering". In: *Journal of the Society for Industrial and Applied Mathematics Series A Control* 2.3 (1964), pp. 347–369. DOI: 10.1137/0302028.

[4] S. J. Julier and J. K. Uhlmann. "Reduced Sigma Point Filters for the Propagation of Means and Covariances through Nonlinear Transformations". In: *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*. Vol. 2. 2002, pp. 887–892. DOI: 10.1109/ACC.2002.1023128.

[5] S. J. Julier. "The Scaled Unscented Transformation". In: *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*. Vol. 6. May 2002, pp. 4555–4559.

[6] Jannik Steinbring and Uwe D. Hanebeck. "S2KF: The Smart Sampling Kalman Filter". In: *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*. Istanbul, Turkey, July 2013.

[7] 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.

[8] Jannik Steinbring, Martin Pander, and Uwe D. Hanebeck. "The Smart Sampling Kalman Filter with Symmetric Samples". In: *Journal of Advances in Information Fusion* 11.1 (June 2016), pp. 71–90.

[9] Uwe D. Hanebeck and Vesa Klumpp. "Localized Cumulative Distributions and a Multivariate Generalization of the Cramér-von Mises Distance". In: *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)*. Seoul, Republic of Korea, Aug. 2008, pp. 33–39. DOI: 10.1109/MFI.2008.4648104.

[10] Uwe D. Hanebeck, Marco F. Huber, and Vesa Klumpp. "Dirac Mixture Approximation of Multivariate Gaussian Densities". In: *Proceedings of the 2009 IEEE Conference on Decision and Control (CDC 2009)*. Shanghai, China, Dec. 2009. DOI: 10.1109/CDC.2009.5400649.

[11] Igor Gilitschenski and Uwe D. Hanebeck. "Efficient Deterministic Dirac Mixture Approximation of Gaussian Distributions". In: *Proceedings of the 2013 American Control Conference (ACC 2013)*. Washington D.C., USA, June 2013. DOI: 10.1109/acc.2013.6580197.

[12] 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. DOI: 10.23919/FUSION49465.2021.9626975.

[13] 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.

[14] J.H. Kotecha and P.M. Djuric. "Gaussian Particle Filtering". In: *IEEE Transactions on Signal Processing* 51.10 (2003), pp. 2592–2601. DOI: 10.1109/TSP.2003.816758.

[15] Monica F. Bugallo and Petar M. Djurić. "Gaussian Particle Filtering in High-Dimensional Systems". In: *2014 IEEE Workshop on Statistical Signal Processing (SSP)*. 2014, pp. 129–132. DOI: 10.1109/SSP.2014.6884592.

[16] 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.

[17] Jannik Steinbring, Antonio Zea, and Uwe D. Hanebeck. "Semi-Analytic Progressive Gaussian Filtering". In: *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2016)*. Baden-Baden, Germany, Sept. 2016. DOI: 10.1109/MFI.2016.7849549.

[18] 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, July 2013.

[19] J. H. Kotecha and P. M. Djuric. "Gaussian Sum Particle Filtering". In: *IEEE Transactions on Signal Processing* 51.10 (Oct. 2003), pp. 2602–2612.

[20] 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. DOI: 10.23919/FUSION45008.2020.9190540.

[21] Daniel Frisch and Uwe D. Hanebeck. "Gaussian Mixture Estimation from Weighted Samples". In: *Proceedings of the 2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2021)*. Karlsruhe, Germany, Sept. 2021. DOI: 10.1109/MFI52462.2021.9591161.

[22] Sanjoy Dasgupta. "The Hardness of k-Means Clustering". In: (2008).

[23] Uwe D. Hanebeck, Daniel Frisch, and Dominik Prossel. "Closed-Form Information-Theoretic Roughness Measures for Mixture Densities". In: *Proceedings of the 2024 American Control Conference (ACC 2024)*. Toronto, Canada, July 2024.

[24] Dominik Prossel and Uwe D. Hanebeck. "Spline-Based Density Estimation Minimizing Fisher Information". In: *Proceedings of the 27th International Conference on Information Fusion (FUSION 2024)*. Venice, Italy, July 2024.

[25] Patrick K. Mogensen and Asbjørn N. Riseth. "Optim: A Mathematical Optimization Package for Julia". In: *Journal of Open Source Software* 3.24 (2018), p. 615. DOI: 10.21105/joss.00615.

[26] Russel E. Caflisch. "Monte Carlo and Quasi-Monte Carlo Methods". en. In: *Acta Numerica* 7 (Jan. 1998). Publisher: Cambridge University Press, pp. 1–49. ISSN: 1474-0508, 0962-4929. DOI: 10.1017/S0962492900002804.

[27] Willliam Feller. *An Introduction to Probability Theory and Its Applications: Volume I*. John Wiley & Sons, 1968.

[28] Fred J. Hickernell. "Koksma-Hlawka Inequality". In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2014. ISBN: 9781118445112. DOI: https://doi.org/10.1002/9781118445112.stat03070.

[29] J. F. Koksma. "The Theory of Asymptotic Distribution Modulo One". en. In: *Compositio Mathematica* 16 (1964), pp. 1–22.

[30] John Verzani. *Roots.jl: Root Finding Functions for Julia*. https://github.com/JuliaMath/Roots.jl. 2020.