2025 23rd European Control Conference (ECC)
June 24-27, 2025. Thessaloniki, Greece

Bayesian Filtering using Galerkin-Methods for Nonlinear Prediction and Measurement Updates

Wolfram Martens[1], Manon Kok[1] and Riccardo Ferrari[1]

Abstract— **This article addresses sequential Bayesian filtering for nonlinear and stochastic dynamical systems. We extend a Galerkin-approach that was previously used for the prediction of non-Gaussian probability density functions, to incorporate linear and non-linear measurement updates. The proposed method results in a linear pipeline of prediction and update steps, which are computed as sparse matrix operations on the finite-dimensional coefficient vector. The performance of our approach is demonstrated in numerical experiments for nonlinear dynamical 2D- and 4D-systems, using results of a standard particle filter as reference, both in terms of accuracy and computational expenses.**

I. INTRODUCTION

In order to safely operate modern technical systems it is essential to maintain accurate and reliable estimates of the system state. Recursive tracking of probability distributions for systems with nonlinear dynamics and measurement models generally results in non-Gaussian posterior distributions and is referred to as *nonlinear filtering* (NLF) [1]–[4]. Example applications of NLF include, but are not limited to, autonomous driving [5], autonomous planetary exploration [6], spacecraft navigation [7], positioning and tracking [8], and automatic detection of faults and cyber-attacks [9].

An important class of NLF approaches tracks parametric representations of the probability density function (pdf), most prominently its first- and second-order moments, as in the various nonlinear extensions of the Kalman Filter (KF). However, linearization and over-simplification due to the pdf parametrization can lead to inaccuracies and, in severe cases, filter inconsistency and divergence for nonlinear, non-Gaussian systems [1]. A fundamental non-parametric alternative for non-Gaussian pdfs and nonlinear systems is given by sequential Monte-Carlo methods [3]: Particle filters (PFs) are broadly applicable and represent the de-facto standard for nonlinear problems where KF extensions are insufficient. However, the point-wise representation of the pdf may lead to sample empoverishment and degeneracy [4].

A key observation we make is that a wide range of stochastic dynamical systems can be modeled by diffusion processes (DPs), for which the pdf can be computed via the Fokker-Planck equation (FPE), a partial differential equation in the pdf [10]. Solving the FPE, in combination with tractable measurement updates, hence represents a promising non-parametric alternative to the previously mentioned NLF techniques. Among the different numerical FPE solution techniques, Galerkin's method represents an elegant way of projecting the infinite-dimensional problem onto a finite-dimensional function space, and has been previously investigated for NLF, both for continuous [11] and discrete measurement updates [12]. However, suffering from the well-known "curse of dimensionality", these methods have since seen only limited application in practice due to the comparably low admissible system dimensions, such as $D = 1$ in [12] and $D = 3$ in [11].

To overcome these limitations we propose the use of a weighted-polynomial expansion, as introduced in [13], [14], and exploit the sparsity and dimension-reduction techniques discussed in [15]. For experimental validation, we consider a nonlinear mechanical system with two degrees of freedom, resulting in a 4D-system (which is considered high in this context [16]), and two variants of a 2D-localization problem with range measurements. We demonstrate the potential of FPE-based filters in scenarios where basic PF approaches are prone to suffer from particle degeneracy. Note that advanced PF methods exist to overcome these issues [17], for which a comparative evaluation is beyond the scope of this paper.

The rest of the paper is structured as follows. Section II provides the mathematical background required for Section III, which describes the prediction and measurement updates in the Galerkin-method. Both sections use 1D-systems for clarity, and the generalization to multivariate systems is provided in Section IV. Section V discusses results from numerical experiments, and Section VI closes with a summary and discussion.

II. MATHEMATICAL BACKGROUND

We formulate the problem of sequential Bayesian filtering for diffusion processes and stochastic measurement models, using a one-dimensional example for clarity.

Notation: Non-bold symbols x are used for scalar expressions, and bold symbols for vectors (lower-case, $\boldsymbol{x}$) and matrices (upper-case, $\boldsymbol{X}$). Upper-case upright letters X_t denote random variables, where t indicates a time index for stochastic processes.

Definition 1 (1D-diffusion process): The stochastic differential equation in Itô-form

$$\mathrm{d}\mathrm{X}_t = f(\mathrm{X}_t)\mathrm{d}t + g(\mathrm{X}_t)\mathrm{d}\mathrm{W}_t, \tag{1}$$

describes a 1D-diffusion process X_t with time-invariant drift and diffusion terms $f(\cdot)$ and $g(\cdot)$, with dW_t denoting the increment of a 1D-Wiener process. Unless stated otherwise,

This work has been partially supported by the EU Horizon program through the project TWAIN, grant id 101122194.

[1]All authors are with Delft Center for Systems and Control, TU Delft, Faculty of Mechanical Engineering, The Netherlands {w.martens,m.kok-1,r.ferrari}@tudelft.nl

the diffusion process and measurement models are assumed time-invariant.

Definition 2 (1D-Fokker-Planck equation): The joint pdf $p(t,x)$ of X_t in (1) is described by the 1D-FPE

$$0 = \mathcal{L}^{\mathrm{FP}}(p(t,x)), \tag{2}$$

where $\mathcal{L}^{\mathrm{FP}}(\cdot) = \frac{\partial(\cdot)}{\partial t} + \frac{\partial}{\partial x}(f(x)\,\cdot) - \frac{\partial^2}{\partial x^2}(b(x)\,\cdot)$ denotes the Fokker-Planck operator with $b(x) = g(x)^2$ [10].

Definition 3 (1D-Measurement model): We assume noisy measurements

$$\mathrm{Z}_t = h(\mathrm{X}_t) + j(\mathrm{X}_t)\mathrm{V}_t \tag{3}$$

at discrete times t, where V_t denotes independent Gaussian noise. In practice, we require for a given measurement z the likelihood $p(z|\mathrm{X}_t = x)$ as a function of x, which we will write as $p(z|x)$ for brevity.

Let us assume a dynamics model as in Def. 1, a measurement model as in Def. 3, an initial pdf $p^0(x)$ of the state variable X_0 at $t = 0$ and a sequence of measurements $z_0, \cdots, z_k$ at $t_0 < \cdots < t_k$. In sequential Bayesian filtering we seek for the conditional pdf of X_t, $p(t,x|z_0,\cdots,z_k)$, for $t \geq t_k$, which can be computed recursively by means of a pdf *prediction* and a *measurement update* step:

Definition 4 (Pdf prediction): Given the pdf at time t_{j-1}, $p(t_{j-1},x|z_0,\cdots,z_{j-1})$, the pdf at time t_j is found by integrating (2) between t_{j-1} and t_j.

Definition 5 (Pdf measurement update): Given the pdf at time t_j, $p(t_j,x|z_0,\cdots,z_{j-1})$ and a measurement z_j at time t_j, the updated pdf is given by Bayes rule,

$$p(t_j,x|z_0,\cdots,z_j) = \frac{p(t_j,x|z_0,\cdots,z_{j-1})p(z_j|x)}{p(z_j|z_0,\cdots,z_{j-1})}. \tag{4}$$

The denominator represents the *marginal likelihood*, which is a normalization factor that does not depend on x and is hence usually not required.

Notation: For simplicity, we will use the following short-hand notation for pdfs before (prior) and after (posterior) measurement updates,

$$p^-(t_j,x) := p(t_j,x|z_0,\cdots,z_{j-1}) \quad \text{(Prior)} \tag{5}$$
$$p^+(t_j,x) := p(t_j,x|z_0,\cdots,z_j) \quad \text{(Posterior)}. \tag{6}$$

III. GALERKIN-METHOD

In order to compute approximate solutions to the problem stated in Sec. II, we first provide a brief review of the approach in [15], which is then extended by a Galerkin-solution for the measurement update according to Def. 5, using a 1D-system for clarity as in the previous section.

A. Finite-dimensional pdf approximation

We assume that $p(t,x)$ is an element of a Hilbert space $\mathcal{U}$ with complete basis $\mathcal{U}$, $\{u_n\}_{n=0}^{\infty}$, such that

$$p(t,x) = \sum_{n=0}^{\infty} c_n(t)u_n(x). \tag{7}$$

Given a subspace $\mathcal{U}_N$ of $\mathcal{U}$, referred to as *trial space*, we search for finite-dimensional approximations of $p(t,x)$ [15]

$$\tilde{p}(t,x) = \sum_{n=0}^{N} c_n(t)u_n(x) = \boldsymbol{u}(x)^T\boldsymbol{c}(t) \approx p(t,x), \tag{8}$$

where

$$\boldsymbol{u}(x) : \mathbb{R} \to \mathbb{R}^{N+1},\ \boldsymbol{u}(x) = [u_0(x),\cdots,u_N(x)]^T \tag{9}$$

is the vector-notation of a basis $\{u_n\}_{n=0}^{N}$ of $\mathcal{U}_N$.

B. Finite-dimensional test space

The key idea of Galerkin's method for pdf approximation is to find the linear coefficients $\boldsymbol{c}(t)$ in (8) such that the approximation error is orthogonal to a finite-dimensional *test space* $\mathcal{V}_N$ with basis $\boldsymbol{v}(x)$. We specify the trial and test basis functions $\boldsymbol{u}(x)$ and $\boldsymbol{v}(x)$ to form a bi-orthogonal system,

$$\left\langle \boldsymbol{v}, \boldsymbol{u}^T \right\rangle_{L_2} = \int \boldsymbol{v}(x)\boldsymbol{u}(x)^T \mathrm{d}x = \boldsymbol{I}, \tag{10}$$

where $\boldsymbol{I}$ is the identity matrix, and $\left\langle \boldsymbol{v}, \boldsymbol{u}^T \right\rangle_{L_2}$ and $\int \boldsymbol{v}(x)\boldsymbol{u}(x)^T\mathrm{d}x$ are short-hand notations for an $(N+1)\times(N+1)$ matrix with element-wise evaluation of the L_2-inner product $\langle v_m, u_n \rangle_{L_2}$.

C. Weighted polynomial trial and test space

$\boldsymbol{u}(x)$ and $\boldsymbol{v}(x)$ are chosen as weighted polynomials,

$$\boldsymbol{u}(x) = \boldsymbol{a}(x)p_0(x), \quad \boldsymbol{v}(x) = \boldsymbol{a}(x)p_w(x), \tag{11}$$

where $\boldsymbol{a}(x) = [a_0(x),\cdots,a_N(x)]^T$ denotes a polynomial basis up to order N. The weighting functions $p_0(x)$ and $p_w(x)$, denoted *trial* and *test pdf*, are selected a priori for a given problem as stationary pdfs in x. The polynomial basis $\boldsymbol{a}(x)$ is orthogonal with respect to a combined weighting function $p_c(x) \propto p_0(x)p_w(x)$, such that (10) holds. It is important to note that the trial and test pdfs can be any kind of pdfs with corresponding orthogonal polynomials, allowing for customized solutions that incorporate prior knowledge about a particular problem. Due to lack of space, however, we restrict the considerations in this paper to Gaussian $p_0(x), p_w(x)$. By using Gaussian $p_0(x)$, the trial space is guaranteed to be dense in L_2 [18], and hence suitable for approximating a large range of pdf classes, excluding only pdfs with heavy tails that are not L_2-integrable.

D. Pdf prediction

Approximate FPE solutions are found by applying the Galerkin-method to the residual $R(t,x) = \mathcal{L}^{\mathrm{FP}}(\tilde{p}(t,x))$,

$$\boldsymbol{0} = \left\langle \boldsymbol{v}, \mathcal{L}^{\mathrm{FP}}(\tilde{p}) \right\rangle_{L_2}. \tag{12}$$

Due to linearity of the FP-operator, this yields a linear system for the time-evolution of $\boldsymbol{c}(t)$ [15],

$$\frac{\partial \boldsymbol{c}(t)}{\partial t} = \boldsymbol{H}c(t). \tag{13}$$

E. Measurement update

Assume that an approximate prior pdf is given at time t,

$$\tilde{p}^-(t,x) = \boldsymbol{u}(x)^T \boldsymbol{c}^-(t), \tag{14}$$

and that we make a measurement observation z. According to (4), the exact posterior after the measurement reads

$$p^+(t,x) = k\ \tilde{p}^-(t,x) p(z|x) \tag{15}$$

$$= k\ \boldsymbol{u}(x)^T \boldsymbol{c}^-(t) p(z|x), \tag{16}$$

with normalization constant k.

Remark 1: By "exact" we mean here that $p^+(t,x)$ is *exact given* $\tilde{p}^-(t,x)$. Any previous approximation errors in $\tilde{p}^-(t,x)$, will of course propagate into $p^+(t,x)$.

Computing $p^+(t,x)$ is generally difficult and does not result in finite representations of the form (8). We hence seek posterior *approximations* of the form

$$\tilde{p}^+(t,x) = \boldsymbol{u}(x)^T \boldsymbol{c}^+(t) \approx p^+(t,x). \tag{17}$$

The coefficients $\boldsymbol{c}^+(t)$ are again found via orthogonality of the approximation error $e_{\tilde{p}^+}(t,x) = \tilde{p}^+(t,x) - p^+(t,x)$,

$$\mathbf{0} = \left\langle \boldsymbol{v}, e_{\tilde{p}^+} \right\rangle_{L_2} = \int \boldsymbol{v}(x)(\tilde{p}^+(t,x) - p^+(t,x))\mathrm{d}x, \tag{18}$$

and using (16) we obtain

$$\boldsymbol{c}^+(t) = k \underbrace{\int \boldsymbol{v}(x)\boldsymbol{u}(x)^T p(z|x)\mathrm{d}x}_{\boldsymbol{L}(z)} \boldsymbol{c}^-(t). \tag{19}$$

The *measurement update matrix* $\boldsymbol{L}(z) \in \mathbb{R}^{(N+1)\times(N+1)}$ needs to be computed for each measurement z.

Remark 2: Naïve computation of all elements in $\boldsymbol{L}(z)$ requires $O\left(N^2\right)$ integral evaluations over x for each combination of $v_m(x), u_n(x)$. However, the products of $\boldsymbol{u}(x)$ and $\boldsymbol{v}(x)$ result in weighted polynomials in x of order up to $2N$, such that it suffices to compute the integrals of the corresponding monomial terms,

$$m_n(z) = \int_{\mathbb{R}} x^n p_c(x) p(z|x)\mathrm{d}x, \quad n = 0, \ldots, 2N, \tag{20}$$

resulting in $O\left(N\right)$ integral evaluations.

It is worth noting that the proposed filter tracks $\tilde{p}(t,x)$ over time without ever establishing the full joint pdf, which can be prohibitive for higher system dimensions. Instead, the prediction and measurement updates correspond to repeated linear updates according to (13) and (19).

IV. GENERALIZATION FOR MULTIVARIATE SYSTEMS

The generalization of Galerkin-solutions of the FPE to multivariate systems with diffusion process $\mathbf{X}_t \in \mathbb{R}^D$, and the implications for $\boldsymbol{H}$ in (13) required for the prediction step, are discussed in [15]. This section hence focuses on the generalization of the measurement update in Section III-E to higher dimensions.

As in [15], we assume independent trial pdfs $p_0(\boldsymbol{x}) = \prod_{d=1}^D p_0^{(d)}(x_d)$, for which the trial functions can equally be written in decoupled form, $u_{\boldsymbol{n}}(\boldsymbol{x}) = \prod_{d=1}^D u_{n_d}^{(d)}(x_d)$, with multi-index $\boldsymbol{n} \equiv \{n_d\}_{d=1}^D$.

Remark 3: We discuss in the following properties of the *trial* basis $\boldsymbol{u}(\boldsymbol{x})$ for decoupled $p_0(\boldsymbol{x})$, which apply equally for decoupled *test* pdfs and *test* functions. For brevity, we will not explicitly restate this duality.

The trial basis for decoupled p_0 is given by the Kronecker-product of the univariate trial functions,

$$\boldsymbol{u}(\boldsymbol{x}) = \bigotimes_{d=1}^{D} \boldsymbol{u}^{(d)}(x_d). \tag{21}$$

For each state variable x_d we select an approximation order N_d, such that the dimension of the multivariate trial basis $\boldsymbol{u}(\boldsymbol{x})$ is $N^{\text{total}} = \prod_{d=1}^D (N_d + 1)$. In practice, we apply a hyperbolic index truncation (HIT) scheme [15] to manage the overall expansion order N^{total}, which would otherwise increase exponentially with the system dimension D. For clarity, we assume no HIT in the following discussion of the multivariate measurement update.

A. Measurement update

For multivariate systems the measurement model reads

$$\mathbf{Z}_t = \boldsymbol{h}(\mathbf{X}_t) + \boldsymbol{J}(\mathbf{X}_t)\mathbf{V}_t, \tag{22}$$

where $\mathbf{Z}_t$ is a D_m-dimensional measurement variable. $\boldsymbol{h} \in \mathbb{R}^{D_m}$ and $\boldsymbol{J} \in \mathbb{R}^{D_m \times D_v}$ denote the multi-dimensional counterparts of h and j in (3), and $\mathbf{V}_t \in \mathbb{R}^{D_v}$ denotes D_v-dimensional independent Gaussian noise. We assume in the following scalar measurements Z_t and denote by $p(z|\boldsymbol{x})$ the measurement likelihood function, which may generally depend on all state variables $\boldsymbol{x}$.

As mentioned in Section III-E, the measurement update (19) involves $O(N)$ integrations over x for each measurement update. This problem is exacerbated for multivariate systems, where the measurement update matrix $\boldsymbol{L}(z) \in \mathbb{R}^{N^{\text{total}} \times N^{\text{total}}}$ reads

$$\boldsymbol{L}(z) = \int \boldsymbol{v}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})^T p(z|\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{23}$$

with $\int(\cdot)\mathrm{d}\boldsymbol{x}$ denoting integration over *all* state variables. Clearly, computing $\boldsymbol{L}(z)$ for arbitrary $p(z|\boldsymbol{x})$ is infeasible in higher dimensions. However, in many real-world applications, the measurement model depends only on a low-dimensional subset of "measured" system states, $\boldsymbol{x}^{\mathrm{m}}$, which allows us to exploit the decoupled structure of the trial functions (21). Without loss of generality we assume $\boldsymbol{x} = [\boldsymbol{x}^{\mathrm{nm}}, \boldsymbol{x}^{\mathrm{m}}]^T$, where $\boldsymbol{x}^{\mathrm{nm}}$ denotes the states that are "not measured", such that $\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{u}^{\mathrm{nm}}(\boldsymbol{x}^{\mathrm{nm}}) \otimes \boldsymbol{u}^{\mathrm{m}}(\boldsymbol{x}^{\mathrm{m}})$. In the following, let N^{m} denote the dimension of the trial basis $\boldsymbol{u}^{\mathrm{m}}$ in the measured states, and N^{nm} the dimension of the trial basis $\boldsymbol{u}^{\mathrm{nm}}$ in the non-measured states, with $N^{\text{total}} = N^{\mathrm{nm}} N^{\mathrm{m}}$.

Basic computation rules for the Kronecker-product yield (omitting the arguments for brevity)

$$\boldsymbol{v}\boldsymbol{u}^T = \left(\boldsymbol{v}^{\mathrm{nm}}(\boldsymbol{u}^{\mathrm{nm}})^T\right) \otimes \left(\boldsymbol{v}^{\mathrm{m}}(\boldsymbol{u}^{\mathrm{m}})^T\right), \tag{24}$$

such that the measurement update matrix can be simplified by separating the integrals over $\boldsymbol{x}^{\text{nm}}$ and $\boldsymbol{x}^{\text{m}}$. We get

$$\boldsymbol{L}(z) = \boldsymbol{I}^{\text{nm}} \otimes \boldsymbol{L}^{\text{m}}(z), \tag{25}$$

where $\boldsymbol{L}^{\text{m}} \in \mathbb{R}^{N^{\text{m}} \times N^{\text{m}}}$ reads

$$\boldsymbol{L}^{\text{m}}(z) = \int \boldsymbol{v}^{\text{m}}(\boldsymbol{x}^{\text{m}})(\boldsymbol{u}^{\text{m}}(\boldsymbol{x}^{\text{m}}))^T p(z|\boldsymbol{x}^{\text{m}}) \mathrm{d}\boldsymbol{x}^{\text{m}}, \tag{26}$$

and $\boldsymbol{I}^{\text{nm}} \in \mathbb{R}^{N^{\text{nm}} \times N^{\text{nm}}}$ is the identity matrix due to bi-orthogonality of $\boldsymbol{u}^{\text{nm}}$ and $\boldsymbol{v}^{\text{nm}}$. The *reduced* measurement update matrix $\boldsymbol{L}^{\text{m}}(z)$ requires integration only over $\boldsymbol{x}^{\text{m}}$ rather than the full state space.

V. EXPERIMENTAL RESULTS

In this section we consider two physical examples to illustrate the capabilities and limitations of the proposed filter. In both examples we use decoupled Gaussian trial and test pdfs, $p_0(\boldsymbol{x}), p_w(\boldsymbol{x})$, resulting in Gaussian $p_c(\boldsymbol{x})$ and generalized Hermite polynomials $\boldsymbol{a}(\boldsymbol{x})$ [15].

A. Coupled Van-der-Pol oscillator

As a first numerical example we consider the 2-degree-of-freedom system described in [15] consisting of two coupled Van-der-Pol oscillators. The drift and diffusion terms read

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{pmatrix} x_2 \\ \beta x_1 + \rho(1+\varepsilon x_1^2)x_2 + \gamma x_3 + \kappa x_4 \\ x_4 \\ \beta x_3 + \rho(1+\varepsilon x_3^2)x_4 + \gamma x_1 + \kappa x_2 \end{pmatrix}$$
$$\boldsymbol{B}(\boldsymbol{x}) = \text{diag}\left(0, \sigma^2, 0, \sigma^2\right), \tag{27}$$

where x_1, x_3 represent the displacements and x_2, x_4 the velocities of the two oscillators. β, ρ and ε describe the linear restoring and nonlinear damping of a single Van-der-Pol oscillator, whereas γ and κ introduce coupling between the oscillators via linear spring and damping elements.

We assume that measurements are available only of the *second* oscillator's displacement, x_3, while we are interested in the probabilistic behavior of the *first* oscillator, namely the marginal pdf of its displacement, $p(t, x_1)$. Assuming a Gaussian measurement model,

$$\mathrm{Z}_t = \mathrm{X}_{3,t} + \sigma_{\text{m}} \mathrm{V}_t, \tag{28}$$

the likelihood function $p(z|\boldsymbol{x}) = p(z|x_3)$ is Gaussian in x_3, and according to (26) we have

$$\begin{aligned} \boldsymbol{L}^{\text{m}}(z) &= \int \boldsymbol{v}^{(3)}(x_3)(\boldsymbol{u}^{(3)}(x_3))^T p(z|x_3) \mathrm{d}x_3 \\ &\propto \int \boldsymbol{a}^{(3)}(x_3)(\boldsymbol{a}^{(3)}(x_3))^T p^*(x_3) \mathrm{d}x_3, \end{aligned} \tag{29}$$

where $\boldsymbol{a}^{(3)}(x_3)$ denotes the Hermite polynomial basis with weighting function $p_c^{(3)} = p_0^{(3)} p_w^{(3)}$, and

$$p^*(x_3) \propto p_c^{(3)}(x_3) p(z|x_3). \tag{30}$$

Due to Gaussian $p_0^{(3)}, p_w^{(3)}$ and $p(z|x_3)$, $p^*(x_3)$ is also Gaussian (with non-zero mean), such that the moments in (20) can be computed analytically, and no numerical integration is required for the computation of $\boldsymbol{L}^{\text{m}}(z)$.

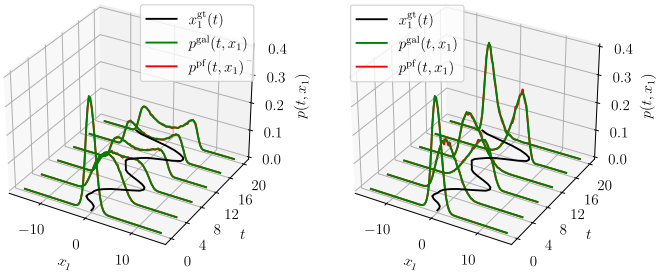

Fig. 1. Marginal pdf evolutions $p(t, x_1)$ for 2DOF Van-der-Pol oscillator (27) with $N_d^{\text{gal}} = 30$, without (left) and with measurements (right) of the second oscillator's displacement x_3 at a frequency of 1Hz. The ground truth trajectory of x_1 is shown in black. Bootstrap filter results are shown for comparison using $N^{\text{pf}} = 10^5$ particles. The numerical parameter values used were $\rho = \gamma = \kappa = 0.2$, $\beta = -0.8$, $\varepsilon = -0.2$, $\sigma = 1$ in (27)

We inspect the marginal pdf of the first oscillator's displacement x_1, compared with the results that would be obtained without any measurements. For both scenarios we provide, as reference, solutions of a basic PF implementation that uses the dynamics model as importance distribution, known as the *bootstrap filter* [3], with a large number of particles $N^{\text{pf}} = 10^5$. Fig. 1 shows the time-evolution of $p(t, x_1)$ with and without measurements. In the absence of measurements (left plot), the marginal pdf $p(t, x_1)$ of the first oscillator's displacement is unaffected by the true movement and approaches a stationary bi-modal distribution, corresponding to the characteristic limit-cycle behavior of the noisy Van-der-Pol oscillator [15]. Using measurements of the second oscillator's displacement x_3 (right plot), the movement of x_1 is indeed reflected in $p(t, x_1)$.

B. Planar localization with nonlinear measurement model

Second, we consider a nonlinear localization problem, in which the 2D-coordinates of a moving target are tracked by a mobile range sensor. Let $\boldsymbol{x}^{\text{target}}$ and $\boldsymbol{x}^{\text{sensor}}$ denote the positions of the target and the sensor, where $\boldsymbol{x}^{\text{sensor}}(t)$ is assumed known, as uncertainty in the sensor location could readily be integrated in the measurement noise.

Assuming little prior knowledge about the movement of the target we model its dynamics by planar Brownian motion with intensity σ_{target}, and further assume noisy range-measurements,

$$Z_t = d\left(\mathbf{X}_t^{\text{target}}, \boldsymbol{x}^{\text{sensor}}(t)\right) + \sigma_{\text{sensor}} \mathrm{d}V_t, \tag{31}$$

where $d(\cdot, \cdot)$ computes the Euclidian distance between the target and the sensor location, and $\mathrm{d}V_t$ denotes independent Gaussian noise. The measurement likelihood is now a function of time due to the (known) sensor location $\boldsymbol{x}^{\text{sensor}}(t)$, and can be written as a 1D-Gaussian in z with mean $d\left(\boldsymbol{x}^{\text{target}}, \boldsymbol{x}^{\text{sensor}}(t)\right)$ and variance σ^2_{sensor}. Note, however, that the likelihood as a function of $\boldsymbol{x}^{\text{target}}$ is *non-Gaussian*.

We first consider a scenario in which the mobile sensor makes noisy range-measurements of the target from different locations, as shown in Fig. 2, together with the resulting joint (posterior) pdfs and the latest sensor location at each time-step. Note that the "target"-superscript in the state variables is omitted for simplicity.

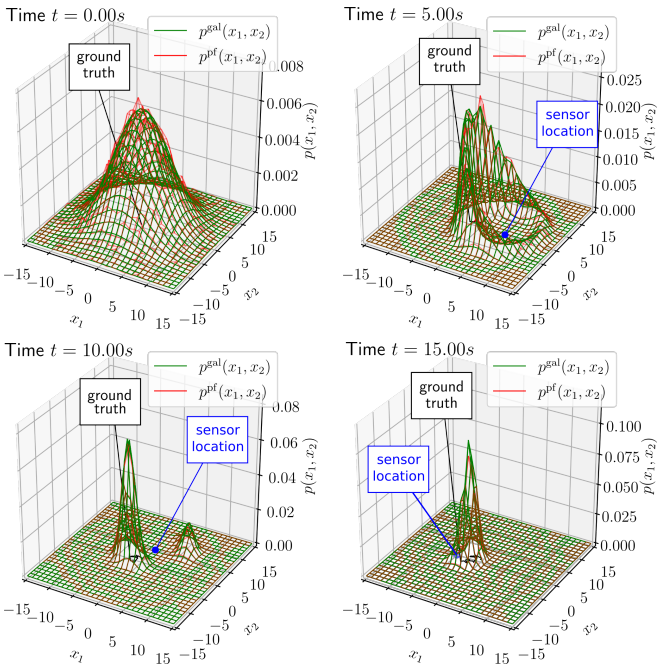

Fig. 2. Joint pdfs for sequential distance measurements with moving sensor, with PF result as reference solution. The ground truth trajectory of the target is plotted in black at the bottom for reference.

The filter is initialized with a zero-mean Gaussian with large standard deviation $\sigma_{t=0} = 5$ both in x_1- and x_2-direction (top left). After the first measurement at $t = 5$, the pdf corresponds to a skewed crater, due to the circular shape of the likelihood function in $\boldsymbol{x}^{\text{target}}$. The combination of measurements at $t = 5$ and $t = 10$ results in a bi-modal distribution, according to the intersection of the corresponding regions with maximum likelihood (bottom left). The last measurement at $t = 15$ resolves the remaining ambiguity, and the probability mass concentrates around the true target location (bottom right). As in the previous section, we provide reference results from a bootstrap filter with $N^{\text{pf}} = 10^5$, denoted p^{pf}.

C. Planar localization with limited observability

We now consider a degenerate version of the planar-localization problem, in which the range sensor makes repeated observations from a *fixed* location. Clearly, the target location becomes unobservable, as there is no way to discern the azimuth angle from range measurements alone. The resulting problem can be efficiently analyzed in polar coordinates, which allows us to compute an accurate reference solution for evaluation.

1) Polar-coordinates reference solution: The diffusion process can be cast to polar coordinates,

$$r(x_1, x_2) = \sqrt{x_1^2 + x_2^2}, \quad \theta(x_1, x_2) = \text{arctan2}(x_2, x_1),$$

where r and θ denote the target's distance from the origin and its azimuth angle, respectively. We assume for simplicity that the sensor is located at the origin, and that the filter is initialized with an isotropic zero-mean Gaussian pdf.

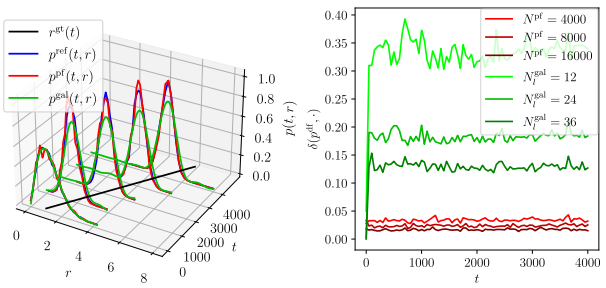

Fig. 3. Left: Marginal distance pdf $p(t, r)$ for a single realization of the experiment with fixed sensor location, with deterministic target movement plotted in black at the bottom, with expansion orders $N_1 = N_2 = 40$ in the Galerkin-filter and $N^{\text{pf}} = 2000$ in the bootstrap filter. Right: TVD with respect to reference solution for distance pdfs for different Galerkin- and PF-solutions, averaged over 10 realizations each.

The random process R_t that describes the distance of the target for planar Brownian motion is described by a *Bessel-Process* [19], and the initial Gaussian distribution in $\mathbf{X}_t$ transforms to a *Rayleigh distribution* [20] in r. R_t's posterior pdf $p(t, r)$ can now be efficiently computed with a simple grid-based filter [3]. The isotropic initial Gaussian distribution results in a uniform marginal pdf for the azimuth angle $p(t = 0, \theta) = \mathcal{U}_{[0,2\pi]}(\theta)$, which is unaffected by the range measurements.

2) Numeric evaluation on controlled scenario: To reduce the variance in our numerical experiments we let the target follow a predefined deterministic trajectory. The target is initialized at $\boldsymbol{x}^{\text{target}}(t = 0) = [1, 0]^T$, thus close to the sensor location at the origin, and moves away in positive x-direction with small, constant velocity $v_{x_1} = 0.001\text{m/s}$. We assume range-measurements with $\sigma_{\text{sensor}} = 1$ at a constant frequency of 1Hz. Fig. 3 (left) shows the marginal pdf in the radial distance $p(t, r)$ at select times as computed by the reference solution ("ref") in blue, our Galerkin-method ("gal") in green, and a bootstrap filter ("pf") in red.

The general trend of the true pdf is consistently tracked by our method, while its sharpness is not fully reproduced due to the finite-dimensional trial space. The bootstrap filter, although using a comparably small number of particles, gives a good recollection of the true pdf. For a quantitative evaluation, we compute the *total variation distance* (TVD) $\delta(p_1, p_2)$ between two probability measures p_1, p_2 [21], shown in Fig. 3 (right) for different expansion orders and numbers of particles. As expected from Fig. 3 (left), the bootstrap filter provides a more accurate representation of $p^{\text{ref}}(t, r)$ than our method, for which the approximation error can be decreased by increasing the expansion orders N_d.

Second, Fig. 4 (left) shows the marginal azimuth pdf $p(t, \theta)$ for the same realization as in Fig. 3, illustrating how the uniform distribution $p(t, \theta)$ is consistently reproduced by our approach. The figure also shows $p^{\text{pf}}(t, \theta)$ as computed by the bootstrap filter. Over the course of repeated measurements and particle resampling, particles are discarded due to their low weight and replaced by duplicates of particles with higher weight, stimulated by the *range measurements*. The azimuth marginal distribution (Fig. 4, left), becomes more

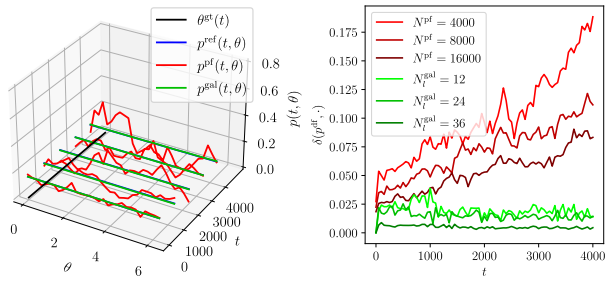

Fig. 4. Left: Marginal azimuth pdf $p(t, \theta)$ for the same experiment as in Fig. 3. Right: TVD with respect to reference solution for azimuth pdfs for different Galerkin- and PF-solutions.

and more clustered, as particles get discarded and resampled, which can be alleviated - but not suppressed entirely - by increasing N^{pf} (Fig. 4, right) in the bootstrap filter.

3) Computation times: The planar-localization example is used for an insight into the computational cost of pdf prediction and measurement updates compared with the bootstrap filter, as implemented on a Dell computer with a 12th Gen Intel(R) Core(TM) i7-1265U processor @ 1.8 GHz and 16GB RAM. Neither of the implementations have been tweaked for maximum performance, hence this comparison represents a qualitative, rather than quantitative, evaluation.

TABLE I

AVERAGE COMPUTATION TIMES.

Method	Parameters	Prediction	Meas. update
Galerkin-filter:	$N_d^{\mathrm{gal}} = 12$	9.87e-04s	2.54e-03s
	$N_d^{\mathrm{gal}} = 24$	9.42e-04s	1.44e-02s
	$N_d^{\mathrm{gal}} = 36$	1.10e-03s	7.54e-02s
Bootstrap filter:	$N^{\mathrm{pf}} = 4000$	2.22e-04s	6.32e-03s
	$N^{\mathrm{pf}} = 8000$	4.07e-04s	1.27e-02s
	$N^{\mathrm{pf}} = 16000$	7.84e-04s	2.53e-02s

Table I shows the average computation times for the pdf prediction and measurement update steps. As expected, the computation times for the measurement update in the Galerkin-filter increase with polynomial expansion order, and are more expensive than the updates in the bootstrap filter. The overall computation times, combining prediction and measurement update steps, however, indicate that our method is competitive in terms of computational cost.

VI. SUMMARY AND DISCUSSION

We have proposed a nonlinear filter that uses weighted-polynomial pdf approximations and a Galerkin-method for pdf prediction and measurement updates. The pdf representation is non-parametric and continuous, and benefits from the sparsity and dimension-reduction techniques discussed in [15], allowing for comparably high-dimensional problems at a manageable computational cost. The presented results confirm the validity of our approach for nonlinear dynamics and measurement models, and suggest potential for problems that are challenging for basic nonlinear filtering alternatives.

To fully study the potential of our method, a thorough comparison with more advanced versions of the particle filter will be important. In addition, it will be interesting to analyze the performance in more realistic and practical scenarios. In this paper, we considered only fixed Gaussian trial pdfs, although the approach can readily be extended to other pdfs. For general trial pdfs and measurement models, the measurement updates may necessitate numerical integration, which will need to be investigated for validity and numerical stability. A further possibility to increase the flexibility in our approach is to investigate *adaptive* Galerkin methods that facilitate online updates of the trial and test function space.

REFERENCES

[1] Y. Bar-Shalom, X.R. Li, T. Kirubarajan, "Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software," Wiley, 2004.

[2] Z. Chen, "Bayesian filtering: From Kalman filters to particle filters, and beyond," Statistics, 2003.

[3] S. Särkkä, "Bayesian Filtering and Smoothing," Cambridge University Press, 2013.

[4] F. Daum, J. Huang, "Particle degeneracy: root cause and solution," in Proceedings of SPIE - The International Society for Optical Engineering 8050, 2011.

[5] W. Schwarting, J. Alonso-Mora, D. Rus, "Planning and Decision-Making for autonomous vehicles," Annu. Rev. Control Robot. Auton. Syst., 1(1), pp. 187-210, 2018.

[6] M.J. Schuster, S.G. Brunner, K. Bussmann, et al., "Towards autonomous planetary exploration," Journal of Intelligent & Robotic Systems, 93(3), pp. 461–494, 2019.

[7] S. Boone and J. McMahon, "An Efficient Approximation of the Second-Order Extended Kalman Filter for a Class of Nonlinear Systems," Proceedings of the European Control Conference (ECC), 2024.

[8] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P. Nordlund, "Particle filters for positioning, navigation, and tracking," IEEE Trans. Signal Process., 50(2), pp. 425–437, 2002.

[9] K. Pan, A. Teixeira, M. Cvetkovic, and P. Palensky, "Cyber risk analysis of combined data attacks against power system state estimation," IEEE Transactions on Smart Grid, 10(3), pp. 3044–3056, 2018.

[10] C.W. Gardiner, "Handbook of Stochastic Methods," Third Edition, Springer, Berlin, 2004.

[11] N.U. Ahmed, S.M. Radaideh, "A Powerful Numerical Technique Solving Zakai Equation for Nonlinear Filtering," Dynamics and Control, 7, pp. 293–308, 1997.

[12] J. Gunther, J. Wilson, T. Oliphant, W.C. Stirling, "Fast nonlinear filtering via Galerkin's method," Proceedings of the American Control Conference (ACC), 1997, pp. 2815–2819.

[13] U. von Wagner, W.V. Wedig, "On the Calculation of Stationary Solutions of Multi-Dimensional Fokker-Planck Equations by Orthogonal Functions," in Nonlinear Dynamics 21, 2000, pp. 289–306.

[14] W. Martens, U. von Wagner, V. Mehrmann, "Calculation of high-dimensional probability density functions of stochastically excited nonlinear mechanical systems," in Nonlinear Dynamics 67(3), 2012, pp. 2089–2099.

[15] W. Martens, R. Ferrari, "Transient Solutions of the Fokker-Planck Equation using a Galerkin-Method with Weighted Test Functions," Proceedings of the European Control Conference (ECC), 2024.

[16] M. Duffy, S. Chung, L. Bergman, "A general Bayesian nonlinear estimation method using resampled Smooth Particle Hydrodynamics solutions of the underlying Fokker-Planck Equation," International Journal of Non-Linear Mechanics, 146, 2022.

[17] A. Doucet, N. De Freitas, N.J. Gordon, "Sequential Monte Carlo methods in practice,", Springer, New York, 2001.

[18] G. Szego, "Orthogonal Polynomials," 4th edition, Colloquium Publications, Vol. XXIII, American Mathematical Society, Providence, RI.

[19] I. Karatzas, S.E. Shreve, "Brownian Motion and Stochastic Calculus," 2nd ed., Springer-Verlag, 1991.

[20] M.H. DeGroot, M.J. Schervish, "Probability and Statistics," Addison-Wesley, 2012.

[21] D.A. Levin, Y. Peres, E.L. Wilmer, J.G. Propp, D.B. Wilson, "Markov Chains and Mixing Times," American Mathematical Society, 2017