Diffusion Maps Particle Filter

Lukas Forster, Alexander Schmidt, Walter Kellermann
Multimedia Communications and Signal Processing
Friedrich-Alexander Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany
alexander.as.schmidt@fau.de

Tal Shnitzer, Ronen Talmon
Viterbi Faculty of Electrical Engineering
Technion – Israel Institute of Technology
Haifa, Israel
ronen@ee.technion.ac.il

***Abstract*—In this paper, we propose a new nonparametric filtering framework combining manifold learning and particle filtering. Diffusion maps, a nonparametric manifold learning method, is applied to obtain a parametric state-space model, inferring the state coordinates, their dynamics, as well as the function that links the state to the noisy observations, in a purely data-driven manner. Then, based on the inferred parametric model, a particle filter is devised, facilitating the processing of high-dimensional noisy observations without rigid prior model assumptions. We demonstrate the performance of the proposed approach in a simulation of a challenging tracking problem with noisy observations and a hidden model.**

***Index Terms*—Manifold learning, nonparametric filtering, nonlinear filtering, sequential Markov chain Monte Carlo**

I. Introduction

In recent years, sequential data sets have become highly evolved with constantly increasing dimensionality. This progress poses significant challenges to classical parametric techniques, such as the seminal Kalman filter (KF) [1], since their prototypical linear and Gaussian models do not capture well the complexity of the data. To address these challenges, one possible approach is to view the high-dimensional sequential data as observations of a dynamical system, which is controlled by latent, low-dimensional driving variables [2]–[7]. Since such dynamical systems could be highly complex and could be lacking a definitive model, nonparametric methods are typically applied. Indeed, nonparametric sequential filtering has recently gained considerable attention, e.g., in [7]–[11]. In general, lacking regularization and constraints, nonparametric methods are prone to perform poorly when the observations are affected by noise. To address this disadvantage, in [9], a nonparametric technique based on Bayesian filtering and Gaussian processes is proposed for filtering observations resulting from nonlinear stochastic dynamical systems. However, modeling based on Gaussian processes might have limited expressivity. In another recent work [11], a Koopman operator-based method is introduced, where a new representation of the data is obtained by relying on a predefined dictionary, followed by nonparametric filtering using an adapted KF. While the dictionary-based representation may provide adequate regularization, it requires prior knowledge. In addition, both [9] and [11] require a training set of hidden state-space samples, which could be hard to obtain in many real-world scenarios.

In this paper, we propose a nonparametric framework based on a hybrid approach, which combines a particle filter (PF), a parametric method, with nonparametric data-driven model inference of the system. Our method leverages a particular nonparametric manifold learning technique, diffusion maps (DM) [12], to obtain a parametric model of the observed dynamical system. We show that, solely from observations in a purely data-driven manner, DM reveals an initial representation of the latent driving variables with an estimate of their dynamics, as well as the function that links the latent variables to the noisy observations. In turn, the inferred model lays the ground for improving the representation by sequential filtering with a PF. As a result, the necessary parametrization of the PF is obtained in a nonparametric data-driven way, thereby adhering to complex sequential data.

The main contribution of this paper is three-fold. First, we devise a data-driven particle filter that can be employed in a completely unknown setting, by recovering the unknown model from observations with only minimal prior assumptions. Second, for improved noise robustness, we propose to first map the observations into a high-dimensional space using echo state representation [13]–[15]. Interestingly, as demonstrated in a large body of work, including support vector machines [16], compressed sensing [17] and sparse representations [18], inflating the initial dimension and obtaining highly redundant representation improves the subsequent dimension reduction. Third, relative to [19] this paper exhibits several advantages: Both [19] and this paper are based on a model inferred from DM. While [19] relies on a KF, which is restricted to Gaussian observations, the present work employs a PF, which supports non-Gaussian observations. Furthermore, we propose a new method for estimating the time-varying state noise, whereas the previous work made use of a time-invariant covariance matrix to represent this noise. In addition to the theoretical justification provided throughout the paper (c.f. Sec. VI), we empirically demonstrate the advantage of the proposed PF compared to the KF-based method. In particular, we show that the performance of the KF-based method deteriorates when the noise level is high. Conversely, we show that the proposed PF equipped with the new robust estimation of the state noise obtains superior performance.

II. Signal Model

Consider a set of N observations $\boldsymbol{z}[t] \in \mathbb{R}^n$ governed by some underlying latent variables $\boldsymbol{\theta}[t] \in \mathbb{R}^m$, such that $\boldsymbol{z}[t] = g(\boldsymbol{\theta}[t])$, where $g : \mathbb{R}^m \to \mathbb{R}^n$ is an arbitrary nonlinear measurement function and t denotes the time index. We further assume that $\boldsymbol{\theta}[t]$ evolves in time according to

$$\boldsymbol{\theta}[t+\Delta t] = \boldsymbol{\theta}[t] + \Delta\boldsymbol{\theta}[t], \qquad (1)$$

where Δt is a small time step and $\Delta\boldsymbol{\theta}[t]$ is defined by

$$\Delta\boldsymbol{\theta}[t] = -\nabla_{\boldsymbol{\theta}[t]} U(\boldsymbol{\theta}[t])\,\Delta t + \sqrt{\frac{2}{\beta}}\Delta\boldsymbol{w}[t], \qquad (2)$$

where $\boldsymbol{w}[t] \in \mathbb{R}^m$ is Brownian motion, $\Delta\boldsymbol{w}[t] = \boldsymbol{w}[t+\Delta t] - \boldsymbol{w}[t]$, $U(\boldsymbol{\theta}[t])$ is the potential field of $\boldsymbol{\theta}[t]$ and β is some constant. This propagation model can be viewed as a discretization of a *Langevin equation* with no external force [20].

III. Preliminaries on Particle Filters

In the sequel, we briefly describe a formulation of a PF based on the signal model presented in Sec. II. Assume we are given a set of observations $\boldsymbol{z}[t]$, a measurement likelihood function (MLF) $p(g(\boldsymbol{\theta}[t])|\boldsymbol{z}[t])$ and a state prediction function (SPF) $p(\boldsymbol{\theta}[t+\Delta t]|\boldsymbol{\theta}[t])$.

The PF is a sequential procedure based on a set of K *particles*, $\tilde{\boldsymbol{\theta}}^{(i)}[t], i = 1,\ldots,K$, where each particle represents a possible realization of the state of the latent variables $\boldsymbol{\theta}[t]$. The particles are initially sampled from some distribution at time $t = 0$. At each time $t > 0$, the particles are propagated iteratively in time according to the SPF. Then, an approximation of the current state, denoted by $\hat{\boldsymbol{\theta}}[t]$, is determined by a weighted average over all particles $\tilde{\boldsymbol{\theta}}^{(i)}[t]$ according to $\hat{\boldsymbol{\theta}}[t] = \sum_{i=1}^{K} \omega_i[t]\,\tilde{\boldsymbol{\theta}}^{(i)}[t]$, where $\omega_i[t]$ is the weight of the i-th particle. The weights are equal to the MLF of the associated particles $\omega_i[t] = p\left(\tilde{\boldsymbol{z}}^{(i)}[t]\,|\boldsymbol{z}[t]\right)$, where $\tilde{\boldsymbol{z}}^{(i)}[t] = g(\tilde{\boldsymbol{\theta}}^{(i)}[t])$ describes the expected observation $\tilde{\boldsymbol{z}}^{(i)}[t]$ given particle $\tilde{\boldsymbol{\theta}}^{(i)}[t]$.

In many PF implementations, the particles associated with small weights are resampled using the MLF, which ensures that only the most probable particles are considered in the subsequent propagation step. For more details on PFs, we refer the readers to [21] and references therein.

IV. Proposed Method

The PF in the previous section requires the MLF and SPF, which are unknown in the setting considered in this paper. Particularly, knowledge of the unknown measurement function $g(\boldsymbol{\theta}[t])$ and the unknown evolution in (2) is required for deriving the MLF and SPF, respectively.

In this section, we first show that by applying DM [12], we can obtain new state coordinates with the required knowledge of (i) the SPF, and (ii) an appropriate MLF. Since these required model properties are unknown in many real-world applications, the proposed PF approach is applicable in a broad range of problems. Fig. 1 summarizes the proposed approach.

DM-based Model Inference. DM is a manifold learning technique which relies on eigenvalue decomposition of a normalized affinity matrix. A common affinity measure is based on a variant of the Mahalanobis distance, which was introduced in [22] and is defined by

$$w(\boldsymbol{z}[t], \boldsymbol{z}[\tau]) = \exp\left(-|\tfrac{1}{2}(\Delta\boldsymbol{z})^T\boldsymbol{M}(\Delta\boldsymbol{z})|^2/\varepsilon^2\right), \qquad (3)$$

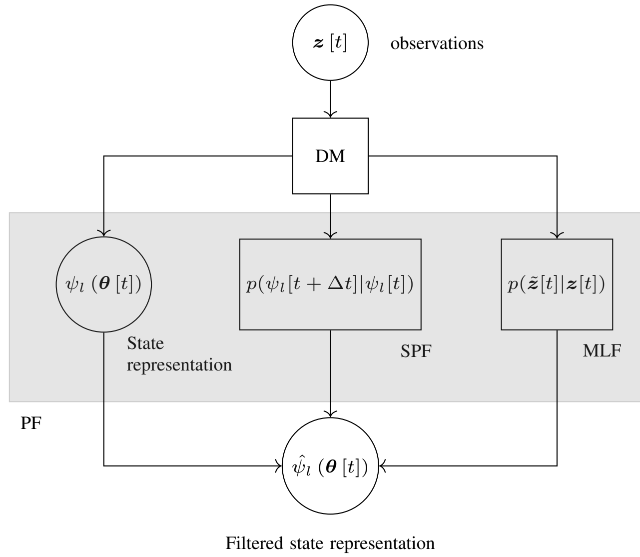

Fig. 1. Scheme of the proposed approach. DM is applied to the observations $\boldsymbol{z}[t]$ to obtain a representation of the latent variables $\boldsymbol{\theta}[t]$, $\psi_l(\boldsymbol{\theta}[t])$, as well as an SPF and an MLF necessary for the application of the PF. These are used as input and parametrisation of the PF which returns a filtered version of $\psi_l(\boldsymbol{\theta}[t])$. Note that only $\boldsymbol{z}[t]$ is known beforehand.

where ε is a scaling factor, which is usually set per application, $\Delta\boldsymbol{z} = \boldsymbol{z}[t] - \boldsymbol{z}[\tau]$, and $\boldsymbol{M} = \boldsymbol{C}^{-1}[t] + \boldsymbol{C}^{-1}[\tau]$, where $\boldsymbol{C}[t]$ is the covariance matrix of $\boldsymbol{z}[t]$. The affinity matrix $\boldsymbol{W} \in \mathbb{R}^{N\times N}$, whose (t,τ) element is $w(\boldsymbol{z}[t], \boldsymbol{z}[\tau])$, is then normalized to be row stochastic, which yields the matrix $\boldsymbol{K} \in \mathbb{R}^{N\times N}$.

Formally, given a set of N observations $\boldsymbol{z}[t]$ in $\mathbb{R}^n$, the normalized pairwise affinity matrix $\boldsymbol{K}$ is of size $N \times N$ having N eigenvalues associated with N eigenvectors $\psi_l \in \mathbb{R}^N, l = 1,\ldots,N$. Using m eigenvectors corresponding to the largest m eigenvalues, each observation can be (nonlinearly) mapped to a new representation by

$$\boldsymbol{z}[t] \mapsto (\psi_1[t], \psi_2[t], \ldots, \psi_m[t]) \in \mathbb{R}^m, \qquad (4)$$

where $\psi_l[t]$ is the t-th element of the eigenvector ψ_l, $t = 1,\ldots,N$. According to [12], setting $m < n$ attains dimension reduction while preserving the affinity between the different samples.

It was shown in [22] that the particular use of the Mahalanobis distance in the affinity (3) gives rise to the new representation in (4), which extracts the 'essence' of the observations and serves as a proxy of the latent variables $\boldsymbol{\theta}[t]$. Namely, the m eigenvectors associated with the m largest eigenvalues constitute a new coordinate system of an inferred latent state-space. For convenience, we denote the implicit dependence of the eigenvectors on the latent state variables by $\psi_l(\boldsymbol{\theta}[t]) := \psi_l[t]$.

State prediction function. It was shown in [19] that for state equations of the form of (2), the eigenvectors obtained by DM evolve according to a discretization of a Langevin equation given by

$$\Delta\psi_l\left(\boldsymbol{\theta}\left[t\right]\right) = -\lambda_l\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)\Delta t + \sqrt{\frac{2}{\beta}}\|\nabla_{\boldsymbol{\theta}[t]}\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)\|_2\Delta\tilde{w}_l\left[t\right], \quad (5)$$

where $\Delta\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)$ is the difference between two consecutive samples of $\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)$, $\tilde{w}_l[t]$ denotes Brownian motion and $-\lambda_l = (2\log\nu_l)/\epsilon$ with ν_l being the l-th eigenvalue and ψ_l the l-th eigenvector obtained by the eigenvalue decomposition of $\boldsymbol{K}$.

Broadly, the evolution of the eigenvectors in (5) was obtained by Itô stochastic calculus [23] relying on the evolution of $\boldsymbol{\theta}[t]$ according to a Langevin equation (2), assuming a sufficiently small time step between consecutive samples. We note that the use of Mahalanobis distance in (3) simplifies the evolution, enforcing $\beta = 2$.

The evolution in (5) gives rise to a parametric model of the dynamics of the new coordinates obtained by DM, even if the state-space and its, possibly nonlinear, dynamics are unknown. We propose to derive the SPF of the PF from (5). By the properties of Brownian motion [24] and (5), we have

$$\Delta\psi_l\left(\boldsymbol{\theta}\left[t\right]\right) \sim \mathcal{N}\left(\mu\left[t\right], \sigma^2\left[t\right]\right), \quad (6)$$

where

$$\mu\left[t\right] = -\lambda_l\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)\Delta t,$$
$$\sigma^2\left[t\right] = \frac{2}{\beta}\|\nabla_{\boldsymbol{\theta}[t]}\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)\|_2^2\Delta t. \quad (7)$$

The parameters of this Gaussian model are completely determined by the eigenvalues ν_l and eigenvectors ψ_l of $\boldsymbol{K}$, with the exception of the term

$$\|\nabla_{\boldsymbol{\theta}[t]}\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)\|_2 = \sqrt{\sum_{i=1}^{m}\left(\frac{\partial\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)}{\partial\theta_i\left[t\right]}\right)^2} \approx \sqrt{\sum_{i=1}^{m}\left(\frac{\Delta\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)}{\Delta\theta_i\left[t\right]}\right)^2}, \quad (8)$$

where m, which is the unknown dimensionality of the hidden parameters $\boldsymbol{\theta}\left[t\right]$, is required. Here, $\Delta\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)$ is used to approximate $\partial\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)$ in this discrete setting where only N samples are available. In the same manner, $\Delta\theta_i\left[t\right]$ is the discrete approximation of $\partial\theta_i\left[t\right]$.

One of the contributions of this paper is an estimation of $\sigma^2[t]$. We note that previous work [19] was based on the same evolution of the eigenvectors (5), where the drift $\mu[t]$ was directly determined by the eigenvalues ν_l and eigenvectors ψ_l, but the diffusion term $\sigma^2[t]$ was set to be a constant covariance matrix. In the remainder of this section, we propose a new method for approximating this diffusion term, which improves the accuracy of the model and facilitates the employment of the PF in a significantly larger set of problems, not restricted only to i.i.d Gaussian noise.

If $m = 1$, (8) can be recast as

$$\|\nabla_{\theta[t]}\psi_l\left(\theta\left[t\right]\right)\|_2 \approx \frac{|\Delta\psi_l\left(\theta\left[t\right]\right)|}{|\Delta\theta\left[t\right]|}. \quad (9)$$

To obtain an approximation of this gradient, we analyze how small changes in $\theta[t]$ affect the coordinates $\psi_l(\theta[t])$. For this purpose, we first approximate $|\Delta\theta[t]|$ by calculating $|\Delta\theta[t]| \approx \min_\tau |\theta[t]-\theta[\tau]|$. Then, we compute $|\Delta\psi_l\left(\theta\left[t\right]\right)| \approx |\psi_l\left(\theta\left[t\right]\right) - \psi_l\left(\theta\left[\tau\right]\right)|$ using the minimizing τ. Since we do not have access to $\theta[t]$, we rely on [22], which showed that the Euclidean distance between samples of the hidden state $\|\theta[t] - \theta[\tau]\|_2$ can be approximated using a variant of the classical Mahalanobis distance between the observations $\boldsymbol{z}[t]$ presented in (3). We omit the details of this - essentially linear - approximation and refer the readers to [22]. For improved stability, the k smallest distances are considered, yielding

$$\|\nabla_{\theta[t]}\psi_l\left(\theta\left[t\right]\right)\|_2 \approx \frac{1}{k}\sum_{\tau\in T}\left(\frac{|\Delta\psi_l|}{|\theta\left[t\right] - \theta\left[\tau\right]|}\right), \quad (10)$$

where $\Delta\psi_l = \psi_l\left(\theta\left[t\right]\right) - \psi_l\left(\theta\left[\tau\right]\right)$, and T is the set of the time indices of the k nearest neighbors of $\theta\left[t\right]$. Since $\psi_l\left(\theta\left[t\right]\right)$ could be noisy due to model mismatches and estimation errors, we further improve the approximation by using linear regression instead of averaging. Concretely, for each $\theta\left[t\right]$, the k smallest distances $|\theta\left[t\right] - \theta\left[\tau\right]|$ and the associated distances $|\psi_l\left(\theta\left[t\right]\right) - \psi_l\left(\theta\left[\tau\right]\right)|$ are used to linearly fit

$$|\psi_l\left(\theta\left[t\right]\right) - \psi_l\left(\theta\left[\tau\right]\right)| = a|\theta\left[t\right] - \theta\left[\tau\right]|_2, \quad (11)$$

where $|\cdot|_2$ denotes the ℓ_2-norm. The fitted variable a can subsequently be used to approximate $\|\nabla_{\theta[t]}\psi_l\left(\theta\left[t\right]\right)\|_2$. The fitted variable a can subsequently be used to approximate $\|\nabla_{\theta[t]}\psi_l\left(\theta\left[t\right]\right)\|_2$.

If $m > 1$, by similar considerations, the following approximation is used

$$\|\nabla_{\boldsymbol{\theta}[t]}\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)\|_2 \approx \frac{\|\Delta\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)\|_2}{\|\Delta\boldsymbol{\theta}\left[t\right]\|_2}, \quad (12)$$

at the expense of an induced squared estimation error

$$E = \Delta\psi_l^2\left(\boldsymbol{\theta}\left[t\right]\right)\left(\sum_{i=1}^{m}\left(\frac{1}{\Delta\theta_i\left[t\right]}\right)^2 - \frac{1}{\sum_{i=1}^{d}\left(\Delta\theta_i\left[t\right]\right)^2}\right).$$

Note that this error is not bounded, i.e., $E \to \infty$ if $m \to \infty$, yet, $\boldsymbol{\theta}\left[t\right]$ is assumed to be low dimensional, reducing the influence of E.

Measurement likelihood function. Since the eigenvectors $\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)$ form a set of basis elements, any real function defined on the latent variables $\boldsymbol{\theta}\left[t\right]$, including the measurement function $g\left(\cdot\right)$, can be written as a linear combination of the eigenvectors $\psi_l\left(\boldsymbol{\theta}\left[t\right]\right)$. Thus, the i-th coordinate of the observation $\boldsymbol{z}[t]$ is given by

$$z_i\left[t\right] = g_i\left(\boldsymbol{\theta}\left[t\right]\right) = \sum_{l=1}^{N}\gamma_{i,l}\psi_l\left(\boldsymbol{\theta}\left[t\right]\right), \quad (13)$$

where $\gamma_{i,l}$ are the expansion weights given by inner products [19]. Typically, it is sufficient to use only the first d eigenvectors retaining the most important information

$$z_i\left[t\right] \approx \sum_{l=1}^{d}\gamma_{i,l}\psi_l\left(\boldsymbol{\theta}\left[t\right]\right) \triangleq \tilde{z}_i[t], \quad (14)$$

where the number of considered eigenvectors d is typically determined heuristically, e.g., using the eigenvalue gap [25].

To account for noise and model mismatches, the MLF is modeled as a Gaussian with diagonal covariance matrix $\mathbf{\Sigma}[t]$, where the i-th diagonal element is the variance of $z_i[t]$, i.e., the i-th coordinate of the n-dimensional observation $\boldsymbol{z}[t]$. Note that if there is some prior knowledge regarding the measurement noise, the MLF can be adapted to the specific noise distribution, e.g., Poisson noise.

We summarize the proposed PF algorithm (cf. Fig. 1):

1) Given the system observations $\boldsymbol{z}[t]$, a new set of coordinates representing the system state, $\psi_l(\boldsymbol{\theta}[t])$, is constructed by applying DM to $\boldsymbol{z}[t]$.
2) Based on the theoretical properties of the DM coordinates, the SPF and MLF are defined as

$$\begin{aligned} &\text{SPF:} \quad p(\psi_l[t+\Delta t]\psi_l[t]) \sim \mathcal{N}\left(-\lambda_l\psi_l[t], \sigma^2[t]\right), \\ &\text{MLF:} \quad p(\tilde{\boldsymbol{z}}[t]|\boldsymbol{z}[t]) \sim \mathcal{N}\left(\mathbf{\Gamma\Psi}, \mathbf{\Sigma}[t]\right), \end{aligned}$$

where $\psi_l[t] = \psi_l(\boldsymbol{\theta}[t])$ denotes the l-th coordinate of the new state representation at time t, $\sigma^2[t]$ is calculated according to (7) and the procedure described thereafter, $\mathbf{\Gamma}$ is a diagonal matrix where the i-th diagonal element is $\gamma_{i,l}$ and $\mathbf{\Psi}$ is a matrix whose l-th row is $(\psi_l[1], \psi_l[2], ..., \psi_l[N])$.
3) At each time step t and for each coordinate l, new particles, denoted by $\tilde{\psi}_l^{(i)}[t],\ i = 1, \ldots, K$, are drawn from the SPF and the state estimate is then calculated according to $\hat{\psi}_l[t] = \sum_{i=1}^{K}\omega_i[t]\tilde{\psi}_l^{(i)}[t]$, where $\omega_i[t] = p\left(\tilde{\boldsymbol{z}}^{(i)}[t]\,|\boldsymbol{z}[t]\right)$ is given by the MLF.

V. Echo State Representation

Samples $\boldsymbol{z}[t]$ recorded in real-world scenarios are usually affected by noise which is currently not taken into account in our signal model. To address this issue, we represent the accessible observations in the *echo state* [13]–[15] which was shown to improve the noise robustness if the noise samples are statistically independent over time.

Concretely, $\boldsymbol{z}[t] \in \mathbb{R}^n$ is transformed to a higher-dimensional echo state $\boldsymbol{x}[t] \in \mathbb{R}^o$ by

$$\boldsymbol{x}[t] = \tanh\left(\boldsymbol{W}_x\boldsymbol{x}[t-\Delta t] + \boldsymbol{W}_i\boldsymbol{z}[t]\right), \tag{15}$$

where $o \gg n$ and $\tanh(\cdot)$ denotes the hyperbolic tangent. While $\boldsymbol{W}_i \in \mathbb{R}^{o\times n}$ maps the current sample into the echo state space, $\boldsymbol{W}_x \in \mathbb{R}^{o\times o}$ propagates the echo state one step forward in time. Matrices $\boldsymbol{W}_x$ and $\boldsymbol{W}_i$ are *randomly* initialized and remain constant. In Sec. VI, we empirically show the benefit of the echo state representation, when the observations $\boldsymbol{z}[t]$ are replaced by $\boldsymbol{x}[t]$ as input to our algorithm (cf. Fig. 1).

VI. Experimental Results

The proposed PF is evaluated on a nonlinear tracking problem: the hidden state $\boldsymbol{\theta}[t] \in \mathbb{R}^2$ evolves by

$$\begin{aligned} \Delta\theta_1[t] &= (3.0 - 0.1\theta_1[t])\,\Delta t + \Delta w_1[t] \\ \Delta\theta_2[t] &= (0.5 - 0.1\theta_2[t])\,\Delta t + \Delta w_2[t] \end{aligned} \tag{16}$$

and can be interpreted as the Cartesian coordinates of an object moving in a two-dimensional space. Nonlinear observations are given in polar coordinates

$$\boldsymbol{z}[t] = \begin{pmatrix}\phi[t] \\ r[t]\end{pmatrix} = \begin{pmatrix}\arctan\left(\frac{\theta_1[t]}{\theta_2[t]}\right) \\ \sqrt{(\theta_1[t])^2 + (\theta_2[t])^2}\end{pmatrix} + \boldsymbol{v}[t], \tag{17}$$

with radius r, azimuth ϕ (both with respect to a reference position) and additive Gaussian noise $\boldsymbol{v}[t]$. This typical simultaneous localization and mapping (SLAM) problem poses two challenges for filtering. First, the observations are nonlinear, rendering many classical filtering methods useless. Second, we assume here that the model of the problem, including the state coordinates, the dynamics, and the measurement function, is unknown and needs to be inferred from the noisy observations. In our simulation, we set $\Delta t = 0.01$ and $N = 1000$. For each noise level, 100 different realizations of latent variables are evaluated to reduce the dependence on particular values. For the same reason, we calculate the average estimation error over 8 different realizations of the echo state representation. The hyperparameters of the algorithm are set to achieve the best empirical performance as follows. The affinity scale ε in (3) is set to 4 times the median distance between the observations. The Mahalanobis distance is computed in short time windows of 20 samples [5]. The dimensionality of the state is $m = 2$, and $k = 8$ in the estimation of $\|\nabla_{\boldsymbol{\theta}[t]}\psi_l(\boldsymbol{\theta}[t])\|_2$ (see (10)). The number of particles is $K = 5000$. For the echo state representation, $o = 100$ dimensions are used, and 60% of the entries of $\boldsymbol{W}_x$ are set to zero.

We compare the performance of the proposed PF to a DM-based KF [19] and to a naïve solution, where no filtering is applied. For brevity, we do not include comparisons to other methods. However, it was shown in [19] that the DM-based KF outperformed a number of competing nonparametric methods [9], [11] in the considered scenario and performed similarly to parametric methods, e.g., extended KF, a classical PF and a work by Tan et al. [26] having knowledge of the hidden unknown model. Note that DM-based KF uses a similar SPF as our approach but ignores the variance term $\sigma^2[t]$ which is estimated in our approach based on (12).

The results are evaluated by the correlation between the obtained representation and $\boldsymbol{\theta}[t]$ and the root mean squared error (RMSE) between the true state $\boldsymbol{\theta}[t]$ and the estimated state $\hat{\boldsymbol{\theta}}[t]$. Here, $\hat{\boldsymbol{\theta}}[t]$ is obtained by mapping $\hat{\psi}_1(\boldsymbol{\theta}[t])$ and $\hat{\psi}_2(\boldsymbol{\theta}[t])$ to the observation space using (13), and then from the observation space to the true space of the latent variables using g^{-1}. Note that the existence of g^{-1} is not a requirement of the method, and would in any case be unknown to it. It was used only to establish a reference for our evaluation.

In Fig. 2, results are illustrated for varying SNRs. Especially in the presence of low noise, the proposed method outperforms the KF and the naïve null filter in terms of the described evaluation criteria for both dimensions of $\boldsymbol{\theta}[t]$. Applying the proposed PF to the echo state representation further improves the results compared to using the presented method with raw observations only (c.f. Fig. 3).

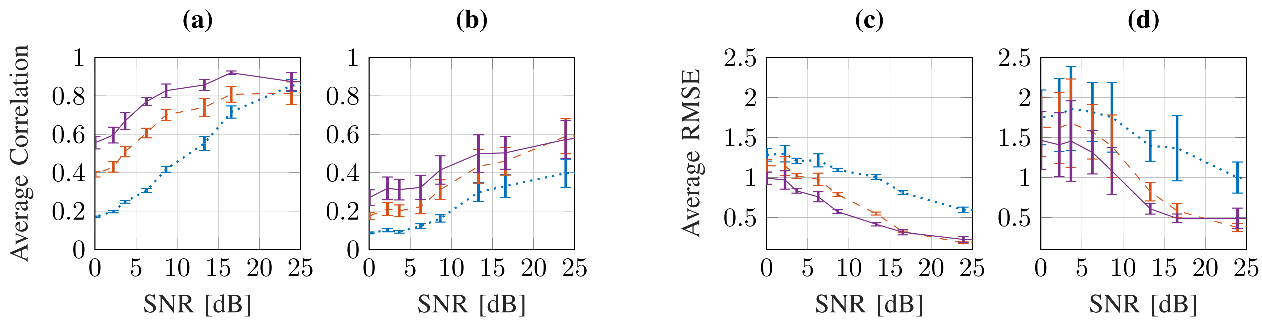

Fig. 2. Results of proposed PF (purple) compared to DM-based KF (red) and initial representation $\psi_l(\boldsymbol{\theta}[t])$ (blue): correlation between (a) $\theta_1[t]$ and $\hat{\psi}_1(\boldsymbol{\theta}[t])$, (b) $\theta_2[t]$ and $\hat{\psi}_2(\boldsymbol{\theta}[t])$, RMSE between (c) $\theta_1[t]$ and $\hat{\theta}_1[t]$, (d) $\theta_2[t]$ and $\hat{\theta}_2[t]$.

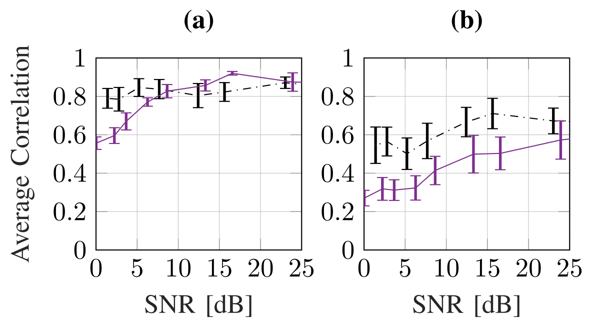

Fig. 3. Results of proposed PF with raw observations (purple) and echo state-represented data (black) as input: correlation between (a) $\theta_1[t]$ and $\hat{\psi}_1(\boldsymbol{\theta}[t])$, (b) $\theta_2[t]$ and $\hat{\psi}_2(\boldsymbol{\theta}[t])$.

VII. Conclusions

We proposed a new PF based on a parametric model that is inferred from observations using (nonparametric) DM. We showed that the proposed PF outperforms the KF presented in [19] in a challenging tracking problem including noisy observations and a completely hidden underlying model. This shows that the inherent nonlinearity of the PF as well as the improved exploration of the state-space are essential. Future work will include the application of the proposed method to measured data. In addition, the presented experimental results suggest that the combination of echo state representation and manifold learning could be very powerful; we will further explore this direction.

References

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans.ASME – J. Basic Eng.*, vol. 82, no. Series D, pp. 35–45, 1960.

[2] W. T. Coffey and Y. P. Kalmykov, *The Langevin equation: with applications to stochastic problems in physics, chemistry and electrical engineering*, World Scientific, 2004.

[3] J. H. Macke, L. Buesing, J. P. Cunningham, B. M. Yu, K. V. Shenoy, and M. Sahani, "Empirical models of spiking in neural populations," in *Advances Neural Inform. Process. Syst. (NIPS)*, 2011, pp. 1350–1358.

[4] R. Talmon, I. Cohen, S. Gannot, and R. R. Coifman, "Diffusion maps for signal processing: A deeper look at manifold-learning techniques based on kernels and graphs," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 75–86, 2013.

[5] R. Talmon and R. R. Coifman, "Empirical intrinsic geometry for nonlinear modeling and time series filtering," *Proc. Nat. Academy of Sciences*, vol. 110, no. 31, pp. 12535–12540, 2013.

[6] R. Talmon, S. Mallat, H. Zaveri, and R. R. Coifman, "Manifold learning for latent variable inference in dynamical systems," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3843–3856, 2015.

[7] T. Berry, D. Giannakis, and J. Harlim, "Nonparametric forecasting of low-dimensional dynamical systems," *Physical Review E*, vol. 91, no. 3, pp. 032915, 2015.

[8] Y. Wang, M. A. Brubaker, B. Chaib-draa, and R. Urtasun, "Bayesian filtering with online gaussian process latent variable models," in *Proc. 13th Conf. Uncertainty Artificial Intelligence*, 2014, pp. 849–857.

[9] M. P. Deisenroth, R. D. Turner, M. F. Huber, U. D. Hanebeck, and C. E. Rasmussen, "Robust filtering and smoothing with gaussian processes," *IEEE Trans. Automat. Control*, vol. 57, no. 7, pp. 1865–1871, 2012.

[10] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, "Chaos as an intermittently forced linear system," *Nature Commun.*, vol. 8, no. 1, pp. 19–28, 2017.

[11] A. Surana and A. Banaszuk, "Linear observer synthesis for nonlinear systems using Koopman operator framework," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 716–723, 2016.

[12] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. and Computational Harmonic Anal.*, vol. 21, no. 1, pp. 5–30, 2006.

[13] H. Jaeger, "Adaptive nonlinear system identification with echo state networks," in *Advances Neural Inform. Proc. Sys.*, 2003, pp. 609–616.

[14] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *science*, vol. 304, no. 5667, pp. 78–80, 2004.

[15] M. Han and M. Xu, "Laplacian echo state network for multivariate time series prediction," *IEEE Trans. Neural Networks and Learning Sys.*, 2018.

[16] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep 1995.

[17] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[18] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[19] T. Shnitzer, R. Talmon, and J.-J. Slotine, "Diffusion maps Kalman filter," *arXiv preprint arXiv:1711.09598*, 2017.

[20] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, vol. 1, Elsevier, 1992.

[21] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, 2002.

[22] A. Singer and R. R. Coifman, "Non-linear independent component analysis with diffusion maps," *Appl. and Computational Harmonic Anal.*, vol. 25, no. 2, pp. 226–239, 2008.

[23] B. Øksendal, "Stochastic differential equations," in *Stochastic differential equations*, pp. 65–84. Springer, 2003.

[24] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, vol. 113, Springer Science & Business Media, 2012.

[25] B. Nadler, S. Lafon, I. Kevrekidis, and R. R. Coifman, "Diffusion maps, spectral clustering and eigenfunctions of fokker-planck operators," in *Advances Neural Inform. Process.Syst.*, 2006, pp. 955–962.

[26] F. Tan, W. Lohmiller, and J.-J. Slotine, "Analytical SLAM without linearization," *Int. J. Robotics Research*, vol. 36, no. 13-14, pp. 1554–1578, 2017.