

Nonlinear Bayesian Estimation with Compactly Supported Wavelets

Achim Hekler, Martin Kiefel, and Uwe D. Hanebeck

Abstract—Bayesian estimation for nonlinear systems is still a challenging problem, as in general the type of the true probability density changes and the complexity increases over time. Hence, approximations of the occurring equations and/or of the underlying probability density functions are inevitable. In this paper, we propose an approximation of the conditional densities by wavelet expansions. This kind of representation allows a sparse set of characterizing coefficients, especially for smooth or piecewise smooth density functions. Besides its good approximation properties, fast algorithms operating on sparse vectors are applicable and thus, a good trade-off between approximation quality and run-time can be achieved. Moreover, due to its highly generic nature, it can be applied to a large class of nonlinear systems with a high modeling accuracy. In particular, the noise acting upon the system can be modeled by an arbitrary probability distribution and can influence the system in any way.

I. INTRODUCTION

In many applications, the state of the system is not completely accessible and consequently has to be estimated. Due to imperfect knowledge, unknown disturbances acting upon the system, and imprecise measurements, it seems reasonable to model the system state as a random variable and to extrapolate it over time by means of a recursive Bayesian estimator.

For linear systems with Gaussian random variables, the well-known Kalman filter provides exact solutions in a computationally feasible manner [1]. However, for non-Gaussian noise or nonlinear systems, a Bayesian estimator is not applicable in general, since the type of the true probability density function changes and its complexity increases over time. Hence, approximations of the occurring equations and/or the underlying true densities are inevitable. For nonlinear systems, well-established variants of the Kalman filter are the Extended Kalman filter [2], the Unscented Kalman filter [3], or other sample-based approaches like the Gaussian estimator [4]. However, due to their Gaussian assumption about the resulting density function, they typically cannot represent the true complex density in an adequate way. In contrast, particle filters, representing the occurring densities by a number of random samples, can approximate arbitrary densities [5]. But one of their main disadvantages is to determine the number of particles for a sufficient representation of the density.

In this situation, wavelets offer a way out. Representing densities with wavelets gives a deterministic approach in

contrast to the random sampling of particle filters. They are known to have excellent properties in the representation of smooth or piecewise smooth functions, too.

Compared to other complex estimation techniques such as Gaussian mixture filters [6], [7], they give a clear description on how the decomposition of the respective densities should be done. Unlike the Gaussian mixture filter, where each component of the describing set of Gaussians influences each point of the density, the number of influencing compactly supported wavelets in a comparable setting is fixed and known in advance.

Taking other orthogonal expansions into account, e.g., a Fourier density approximation [8], [9], the wavelet coefficients do not only carry information about the influence of certain frequencies, but also localize the behavior in space.

Furthermore, the specific hierarchical structure of the wavelet decomposition can be utilized in other research fields, as for example in stochastic nonlinear model predictive control [10]. In particular, states in the remote future could be rated less precisely than states in the near future. Such an approach is very similar to how humans plan.

In this paper, we present an estimator for nonlinear, time-invariant systems, whose complexity is constant over time. Our approach is based on approximations of the conditional densities that represent a probabilistic description of the nonlinear system and measurement equations with wavelet expansions. In doing so, the prediction and measurement step of a wavelet-based Bayesian estimator can be calculated in closed form. By using a wavelet decomposition, we benefit from the preferable characteristics of wavelets mentioned above. Moreover, we show that the corresponding operations for such an estimator only consist of simple summations and multiplications and are linear with respect to the significant coefficients of the wavelet extensions. Besides the excellent approximation capabilities of the wavelet transform, fast algorithms, which operate on sparse vectors, can be applied and an efficient trade-off between approximation quality and run-time can be achieved.

The paper is structured as follows: In the next section, we present the two main processing steps of a Bayesian estimator. Afterwards, a brief review about the properties of compactly supported wavelets is given and the basic idea of multiscale analysis is presented. For all involved operations of a Bayesian estimator, the counterparts in wavelet decomposition are derived in Section V. In Section VI, the capabilities of the proposed approach are illustrated by means of simulations. An outlook to future work concludes the paper.

A. Hekler, M. Kiefel, and U. D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. Achim.Hekler@ieee.org, Martin.Kiefel@stud.uni-karlsruhe.de, uwe.hanebeck@ieee.org

II. PROBLEM FORMULATION

In this paper, we consider a time-invariant, discrete-time system, described by a known stochastic system model

$$\underline{\mathbf{x}}_{k+1} = \underline{a}(\underline{\mathbf{x}}_k, \underline{\mathbf{u}}_k, \underline{\mathbf{w}}_k) \quad (1)$$

and a known stochastic measurement model

$$\underline{\mathbf{y}}_k = \underline{h}(\underline{\mathbf{x}}_k, \underline{\mathbf{v}}_k), \quad (2)$$

where $\underline{\mathbf{x}}_k$ is the not completely accessible system state, $\underline{\mathbf{u}}_k$ the system input, and $\underline{\mathbf{y}}_k$ the measurement. The random variables $\underline{\mathbf{w}}_k$ and $\underline{\mathbf{v}}_k$, characterized by their known underlying probability densities $f_k^w(\underline{\mathbf{w}}_k)$ and $f_k^v(\underline{\mathbf{v}}_k)$, subsume the system and measurement noise and can be *arbitrarily* distributed. Furthermore, $\hat{\underline{\mathbf{y}}}_k$ denotes a specific realization of the measurement process. Throughout the paper, we assume, that the system state $\underline{\mathbf{x}}_k$ is restricted to a bounded set.

Based on the equations (1) and (2), the inaccessible system state $\underline{\mathbf{x}}_k$ can be estimated in an efficient recursive way starting from a known initial prior density f_0^e by applying a Bayesian estimator with its two processing steps.

The *time update* or *prediction step* extrapolates the result of the previous filter step over one time step and can be expressed in terms of probability density functions by the Chapman-Kolmogorov equation

$$f_{k+1}^p(\underline{\mathbf{x}}_{k+1}) := \int f_k^T(\underline{\mathbf{x}}_{k+1}|\underline{\mathbf{x}}_k, \underline{\mathbf{u}}_k) \cdot f_k^e(\underline{\mathbf{x}}_k) d\underline{\mathbf{x}}_k, \quad (3)$$

where $f_k^T(\underline{\mathbf{x}}_{k+1}|\underline{\mathbf{x}}_k, \underline{\mathbf{u}}_k)$ denotes the transition density and can be derived from (1).

In the *filtering step* or *measurement update*, a measurement $\hat{\underline{\mathbf{y}}}_k$ is incorporated into a prior density $f_k^p(\underline{\mathbf{x}}_k)$ in order to obtain an improved estimation $f_k^e(\underline{\mathbf{x}}_k)$, i.e.,

$$f_k^e(\underline{\mathbf{x}}_k) = \frac{f_k(\hat{\underline{\mathbf{y}}}_k|\underline{\mathbf{x}}_k) \cdot f_k^p(\underline{\mathbf{x}}_k)}{\int f_k(\hat{\underline{\mathbf{y}}}_k|\underline{\mathbf{x}}_k) \cdot f_k^p(\underline{\mathbf{x}}_k) d\underline{\mathbf{x}}_k}, \quad (4)$$

where the conditional density $f_k(\hat{\underline{\mathbf{y}}}_k|\underline{\mathbf{x}}_k)$ denotes the likelihood and can be derived from (2).

For nonlinear systems with arbitrarily distributed random variables, both integrals in eq. (3) and (4) cannot be solved in closed form. Thus, adequate approximations are inevitable. Since a repeated approximation of $f_k^p(\underline{\mathbf{x}}_k)$ in every processing step is typically computationally intractable, we approximate the conditional densities in eq. (3) and (4) by their wavelet expansions instead. Due to the assumption that the system state is restricted to a bounded set together with the time-invariant system and measurement models, an offline approximation is possible.

III. KEY IDEA

The key idea of this paper is to enable a closed-form calculation of the prediction and measurement step by means of wavelet expansions of the conditional densities. Compared to similar approaches using axis-aligned Gaussian mixtures as in [6] or [11], wavelets have some significant advantages. Besides their good approximation quality, no complex optimization problem has to be solved in order to obtain

the decomposition of a given density. Furthermore, wavelets are localized in space, so that the number of influencing wavelets of each point of the density is fixed and known in advance. This fact is particularly interesting with respect to a parallel processing of the density. Finally, efficient algorithms operating on wavelet expansions combined with an adequate representation of a wide class of functions results in a good trade-off between run-time and approximation quality. In order to take advantage of all these facts, all operations of the estimator should be executed in the wavelet domain.

IV. INTRODUCTION IN WAVELETS

In this section, we would like to briefly introduce the notion of orthogonal and biorthogonal compactly supported wavelets. For clarity reasons, we first consider a one-dimensional state space. The expansion to the multi-dimensional case is presented at the end of this section. Please also refer to [12] for a deeper review.

A. Compactly Supported Wavelets

A wavelet basis of $L^2(\mathbb{R})$ consists of dilated and translated versions of a function $\psi(\cdot)$,

$$\psi_{j,n}(x) = \frac{1}{\sqrt{2^j}} \psi(2^{-j}x - n), \text{ with } j, n \in \mathbb{Z}. \quad (5)$$

The *wavelet function* $\psi(\cdot)$ is strongly connected to the so-called *scaling function* $\phi(\cdot)$ and they satisfy the equations

$$\begin{aligned} \phi(x) &= \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \phi(2x - n) \text{ and} \\ \psi(x) &= \sqrt{2} \sum_{n \in \mathbb{Z}} g_n \phi(2x - n). \end{aligned} \quad (6)$$

The sequences $h = \{h_n\}_{n \in \mathbb{Z}}$ and $g = \{g_n\}_{n \in \mathbb{Z}}$ are called filters as they are used for the fast wavelet transform [13]. It can be shown that there are only finitely many non-zero sequence elements if and only if the corresponding wavelet has compact support [13].

For the rest of the paper, we use the notation $\phi_{j,n}$ analog to the definition of $\psi_{j,n}$ introduced in (5).

In order to understand the following, it is also necessary to know that the scaling function is normalized and the wavelet function has M vanishing moments, such that

$$\begin{aligned} \int \phi(x) dx &= 1 \text{ and} \\ \int x^m \psi(x) dx &= 0, m = 0, \dots, M-1. \end{aligned} \quad (7)$$

To reconstruct a function $f \in L^2(\mathbb{R})$ from its inner products with this basis, a reconstruction basis consisting of $\tilde{\psi}_{j,n}$ and $\tilde{\phi}_{j,n}$ and with it the filters \tilde{h}_n and \tilde{g}_n are needed. In the case of biorthogonal wavelets, the basis functions fulfill the equations

$$\langle \psi_{j,n}, \tilde{\psi}_{j',n'} \rangle = \delta[n - n'] \delta[j - j'],$$

and therefore, the reconstruction can be described as

$$f = \sum_{n,j=-\infty}^{\infty} \langle f, \psi_{j,n} \rangle \tilde{\psi}_{j,n} = \sum_{n,j=-\infty}^{\infty} \langle f, \tilde{\psi}_{j,n} \rangle \psi_{j,n}$$

with the usual inner product of $L^2(\mathbb{R})$

$$\langle f, \psi \rangle = \int f(x) \psi^*(x) dx .$$

ψ^* constitutes the conjugate complex of ψ , but since we are using real-valued wavelets, we drop the conjugate complex for the rest of the paper.

The two possible ways of expressing a function f relative to either of the basis show that this approach is completely symmetric and it does not matter, which filter is used for transformation, and which is used for reconstruction.

In the special case of an orthonormal wavelet basis, the reconstruction basis coincides with the transformation basis, as the basis constitutes a tight frame.

B. Multiscale Analysis

With

$$V_j = \text{span}\{\phi_{j,n} : n \in \mathbb{Z}\} \text{ and } W_j = \text{span}\{\psi_{j,n} : n \in \mathbb{Z}\}$$

the wavelet basis induces separate nested subspace of $L^2(\mathbb{R})$ – a multiresolution analysis on $L^2(\mathbb{R})$

$$\dots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \dots$$

with

$$V_{j-1} = V_j \oplus W_j \text{ and}$$

$$\bigcap_{j=-\infty}^{\infty} V_j = \{0\}, \text{ closure} \left(\bigcup_{j=-\infty}^{\infty} V_j \right) = L^2(\mathbb{R}) .$$

A simple way to think about this multiresolution approach is to build up the function from an approximation (low frequency components) and its details (high frequency components) on top of it.

Since we are only able to use a finite amount of memory, we have to restrict ourselves to a certain granularity and express the function f with the basis of V_0 with respect to a base space $V_{j'}$. Such a basis is $\{\phi_{j',n}\}_{n \in \mathbb{Z}} \cup \{\psi_{j,n}\}_{j \in \{1, \dots, j'\}, n \in \mathbb{Z}}$. A restriction of the allowed shifts of the wavelet and scaling functions is accomplished by our assumption to only work with densities with compact support.

Therefore, we are able to express $f \in V_0$ with

$$f(x) = \sum_{n \in \mathbb{Z}} \langle f, \tilde{\phi}_{j',n} \rangle \phi_{j',n}(x) + \sum_{j=1}^{j'} \sum_{n \in \mathbb{Z}} \langle f, \tilde{\psi}_{j,n} \rangle \psi_{j,n}(x) .$$

Please refer to [14] for more information on how to actually calculate the wavelet coefficients with the fast wavelet transform.

C. Multi-dimensional Case

In order to expand this approach to the multi-dimensional case, we use the separable wavelet basis (to notation the expression, assume $\psi_{j'+1,k}(x) := \phi_{j',k}(x)$) for a N -dimensional space $\bigotimes_{d=1}^N V_0$

$$\{\psi_{j_1, n_1}(x_0) \cdots \psi_{j_N, n_N}(x_N) \mid d \in \{1, \dots, N\}, j_d \in \{1, \dots, j' + 1\}, n_d \in \mathbb{Z}\} .$$

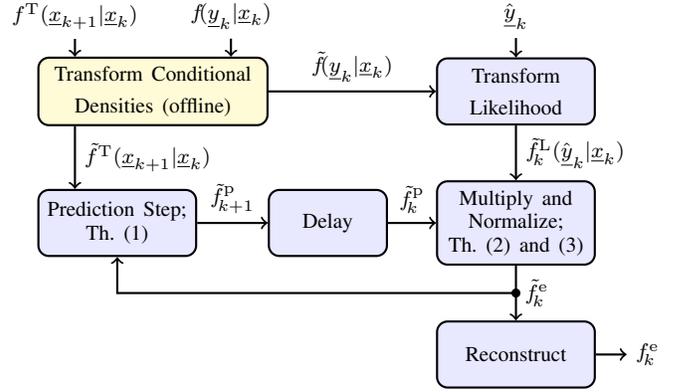


Fig. 1: Bayesian wavelet estimator. All densities that are marked with a tilde are to be understood in wavelet expansion.

Please note that we can choose the wavelet family for each dimension of the density separately. This also includes the symmetric pair of a biorthogonal wavelet basis.

The reason why we use this special multi-dimensional basis in contrast to the more common basis that ends up with the same scale parameters in every dimension is rooted in the prediction operation that will be introduced shortly.

Other more specialized methods exist to build multi-dimensional basis but they are beyond the scope of this paper.

V. BAYESIAN ESTIMATION WITH WAVELETS

As described in Section II, two different abstract operations have to be specified with all involved functions represented in wavelet domain. From now on, we assume that all functions are available in wavelet expansion, calculated via the fast wavelet transform. On the one hand, the prediction step has to implement a marginalization of a product of a transition density with a density function. Our approach solves this in a fast and exact way. On the other hand, the measurement update has to calculate a product of two densities and normalize the result. We show how an adequate approximation for the multiplication can be found. In Fig. 1, the modules of the wavelet-based Bayesian estimator are depicted.

For the proposed technique, both the use of compactly supported orthonormal wavelet basis and of compactly supported biorthogonal wavelet basis is possible.

A. Prediction Step

We make some simplifying assumption in this paragraph to increase the readability of the proofs and theorems. First, we assume that the system state can be expressed by a scalar. Since we are working with separable basis, the discussion can be extended easily to the multi-dimensional case using the multi-dimensional basis presented in the previous chapter. Second, the control input is implicitly given by the transition density. A different input at time step k can be simulated by exchanging the system transition density at time step k .

Third, to simplify the notation of the wavelet expansion of transition densities and likelihoods, we subsume both the scale and shift parameters in one single parameter. The set Λ is used to denote all valid combinations.

Theorem 1 Let the current density and the transition density of eq. (3) be given by

$$f_k^e(x_k) = \sum_{i \in \Lambda} c_i^{(e)} \tilde{\psi}_i(x_k) \text{ and}$$

$$f_k^T(x_{k+1}|x_k) = \sum_{i, i' \in \Lambda} c_{i, i'}^{(T)} \tilde{\psi}_i(x_{k+1}) \psi_{i'}(x_k), \text{ where}$$

$$c_i^{(e)} = \langle f_k^e, \psi_i \rangle \text{ and}$$

$$c_{i, i'}^{(T)} = \int \int f_k^T(x_{k+1}|x_k) \psi_{i'}(x_{k+1}) \tilde{\psi}_i(x_k) dx_{k+1} dx_k.$$

Then, the predicted density can be expressed by

$$f_{k+1}^p(x_{k+1}) = \sum_{i \in \Lambda} c_i^{(p)} \tilde{\psi}_i(x_{k+1}), \text{ where}$$

$$c_i^{(p)} = \sum_{i' \in \Lambda} c_{i, i'}^{(T)} c_{i'}^{(e)}.$$

Proof: Starting with

$$f_{k+1}^p(x_{k+1}) = \int f_k^T(x_{k+1}|x_k) f_k^e(x_k) dx_k$$

$$= \int f_k^T(x_{k+1}|x_k) \sum_{i \in \Lambda} c_i^{(e)} \tilde{\psi}_i(x_k) dx_k$$

$$= \sum_{i \in \Lambda} c_i^{(e)} \int f_k^T(x_{k+1}|x_k) \tilde{\psi}_i(x_k) dx_k$$

and taking the inner product with the basis element $\psi_{i'}$ on both sides, we get

$$\langle f_{k+1}^p, \psi_{i'} \rangle$$

$$= \sum_{i \in \Lambda} c_i^{(e)} \int \int f_k^T(x_{k+1}|x_k) \psi_{i'}(x_{k+1}) \tilde{\psi}_i(x_k) dx_{k+1} dx_k$$

$$= \sum_{i \in \Lambda} c_i^{(e)} c_{i, i'}^{(T)}.$$

Wavelet-based Bayesian prediction is $\mathcal{O}(|f_k^T|)$ when $|f_k^T|$ is defined as the number of non-zero coefficients of the transition density¹. This can be achieved if the lookup of coefficients is $\mathcal{O}(1)$ and the operation is iterating over all non-zero coefficients of the transition density. ■

B. Filter Step

In contrast to the prediction step, there are two necessary operations for the measurement update in eq. (4). The next few paragraphs cover the multiplication. At the end of this section, there is a discussion how normalization is achieved.

Multiplication of two functions given in their wavelet expansions is a tricky operation (see [15] for a more detailed review), since it is not a linear operation and the product of two elements of a basis is not necessarily an element of that

basis again. Therefore, we have to find a way to express the result of the operation relative to the original basis functions.

To simplify the notation for the next theorem, we introduce two projections $P_j : L^2(\mathbb{R}) \rightarrow V_j$ and $Q_j : L^2(\mathbb{R}) \rightarrow W_j$.

Theorem 2 Given two densities $f, g \in V_0$ and their multiresolution expansions (g is split up in the same way) with respect to a base space $V_{j'}$,

$$f = P_{j'}(f) + \sum_{j=1}^{j'} Q_j(f).$$

Then, the product can be written as

$$f \cdot g = P_{j'}(f) P_{j'}(g)$$

$$+ \sum_{j=1}^{j'} (P_j(f) Q_j(g) + Q_j(f) P_j(g) + Q_j(f) Q_j(g)).$$

Proof: Please see [15] for a variant of this proof. ■

So, the problem of finding the result of the multiplication of two functions boils down to finding the product of basis functions in the same scale.

The missing piece for expressing the product of two basis functions with respect to the basis is given by the *connection coefficients*

$$\int \phi_{0,0}(x) \phi_{0,m}(x) \phi_{0,n}(x) dx,$$

which are fixed for a certain wavelet family. A description, of how to calculate these values, can be found in [14].

From Theorem 2, we know that we only have to consider basis functions on the same scale. This fact drastically reduces the amount of connection coefficients that have to be calculated and enables the reuse of intermediate results.

Working with compactly supported wavelets renders most of the shift combinations to zero. Furthermore, the estimate

$$\int \psi_{0,0}(x) \psi_{j,m}(x) \psi_{j,n}(x) dx$$

$$= 2^{-j} \int \psi(x) \psi(2^{-j}x - m) \psi(2^{-j}x - n) dx$$

$$\leq 2^{-j} c$$

leads to the fact that the error of not calculating the “spill-over” effect into finer scales can be seen as negligible after a certain upscaling. The decision, of how many finer scales are taken into account, allows to set the accuracy of the multiplication operation.

Also note that after the multiplication is finished a partial run of the fast wavelet transformation is necessary to rebuild a proper density in a wavelet expansion.

Considering multi-dimensional state spaces, the number of the multi-dimensional connection coefficients unfortunately is exploding due to the Cartesian product, which is necessary to get all combinations. Additionally, all combinations between the different multi-dimensional projections on the same scale need to be calculated.

The time complexity of the multiplication is $\mathcal{O}(\max(|f|, |g|) \cdot s^2)$ for the worst case, where again

¹ $|\cdot|$ is not a valid norm as it violates $a \in \mathbb{R} : |a \cdot f| = |a| \cdot |f|$.

$|f|$ constitutes the number of non-zero coefficients of the density f and s represents the finest scale. To explain this, we have to split up operations needed for the multiplication of two densities into the amount of operations needed for the multiplication of the different scales and those needed to transform the density back into a valid wavelet expansion.

During the multiplication operation, the product between $Q_j(f)P_j(g)$ and vice versa on each scale is the heavy weight due to the reconstruction of every single needed approximation coefficient. Certainly, only those coefficients of $P_j(g)$ have to be restored where the corresponding coefficient of $Q_j(f)$ is non-zero.

This leads to the question of the cost of the reconstruction of one single approximation coefficient on a high scale, or the question of how many other coefficients have to be touched and calculated. If l is the length of the filter corresponding to the used wavelet, then every approximation coefficient is backed up by $\frac{l}{2}$ approximation and detail coefficients on the next finer scale. Continuing this recursive argument leads to $\frac{l}{2} \sum_{i=1}^s i$ touched and calculated values, which is $\mathcal{O}(\frac{l}{2}s^2)$.

To estimate the worst-case behavior for the transformation of the resulting density into a valid wavelet expansion, we need to know on which scales how many of the newly created approximation coefficients lie. Since the multiplication on each scale yields $\mathcal{O}(\max(|f|, |g|))$ coefficients in total, it is permitted to assume them all on the finest scale, which corresponds to the worst-case. Now $\mathcal{O}(\max(|f|, |g|))$ operations are needed to rebuild the density.

As a result, the multiplication is linear in time with respect to the significant coefficients of the densities.

The next theorem covers the problem of normalization of a density. Since wavelet functions and scaling function are behaving differently when the integral is calculated, we distinguish between them in the next paragraphs.

Theorem 3 *Given a density in its wavelet expansion*

$$f = \sum_{n \in \mathbf{Z}} \langle f, \tilde{\phi}_{j',n} \rangle \phi_{j',n} + \sum_{j=1}^{j'} \sum_{n \in \mathbf{Z}} \langle f, \tilde{\psi}_{j,n} \rangle \psi_{j,n} .$$

Then $g := f/c$ satisfies $\int g(x) dx = 1$ with

$$c = \sqrt{2}^{j'} \sum_{n \in \mathbf{Z}} \langle f, \tilde{\phi}_{j',n} \rangle .$$

Proof: Straightforward calculation leads to

$$\begin{aligned} & \int f(x) dx \\ &= \int \sum_{n \in \mathbf{Z}} \langle f, \tilde{\phi}_{j',n} \rangle \phi_{j',n}(x) + \sum_{j=0}^{j'} \sum_{n \in \mathbf{Z}} \langle f, \tilde{\psi}_{j,n} \rangle \psi_{j,n}(x) dx \\ &= \sqrt{2}^{j'} \sum_{n \in \mathbf{Z}} \langle f, \tilde{\phi}_{j',n} \rangle . \end{aligned}$$

The last step is supported by (7) and the linearity of integration. ■

The worst-case time complexity of the normalization process belongs to the set $\mathcal{O}(|f|)$. Please note that the

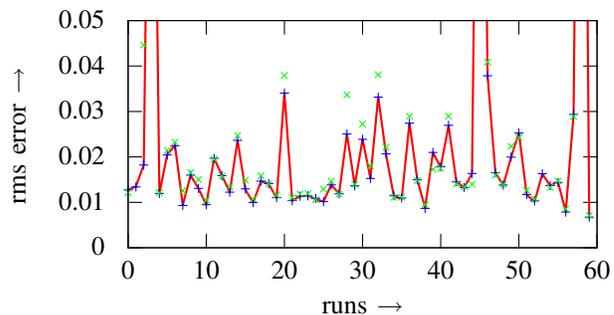


Fig. 3: Root mean square error (rmse) of the proposed estimator (+) measured against a particle filter (x) and the true Bayesian estimator (-).

complexity of describing the densities increases with every multiplication, but it is implicitly reduced in every prediction step.

There are more ways to reduce the complexity of the densities like thresholding and cutting off higher scales, but they are beyond the scope of this paper.

VI. SIMULATION RESULTS

For realization of (1), we employ the specific system model

$$\mathbf{x}_{k+1} = \sin(\mathbf{x}_k) + \mathbf{x}_k + \mathbf{w}_k ,$$

where the random variable \mathbf{w}_k follows a normal distributions with expected value $\mu_k^w = 0$ and standard deviation $\sigma_k^w = 0.8$.

As measurement equation, we use

$$\mathbf{y}_k = \mathbf{x}_k^3 + \mathbf{v}_k ,$$

where \mathbf{v}_k is normally distributed with parameters $\mu_k^v = 0$ and $\sigma_k^v = 0.3$. The initial state of the estimator is also described by a normal distribution with mean $\mu_0^x = 1.5$ and standard deviation $\sigma_0^x = 1.2$. Please note that the wavelet-based estimator can handle much more complex simulation settings, such as an arbitrarily distributed multiplicative noise.

We conducted 60 Monte Carlo simulation runs and each run consisted of 20 alternating prediction and measurement update steps. Throughout the simulations, we use Daubechies wavelets with two vanishing moments to approximate the system and state densities and consider three additional finer scales for the multiplication to cover the spill-over effect.

In Fig. 2, two exemplary results of these operations are shown. The wavelet-based Bayesian estimator is able to follow the densities with a very good quality. Not only does it express the same expected value, but also higher moments are retained in a precise manner. Even high frequency singularities, like the falling edge of the estimated density in Fig. 2b, are reproduced without any ringing, although the calculation of the spill-over effect was stopped after the 3th scale.

To additionally emphasize the success of our method, we compare it to a true Bayesian estimator and a common non-linear filter: the particle filter [5]. We use two instances of

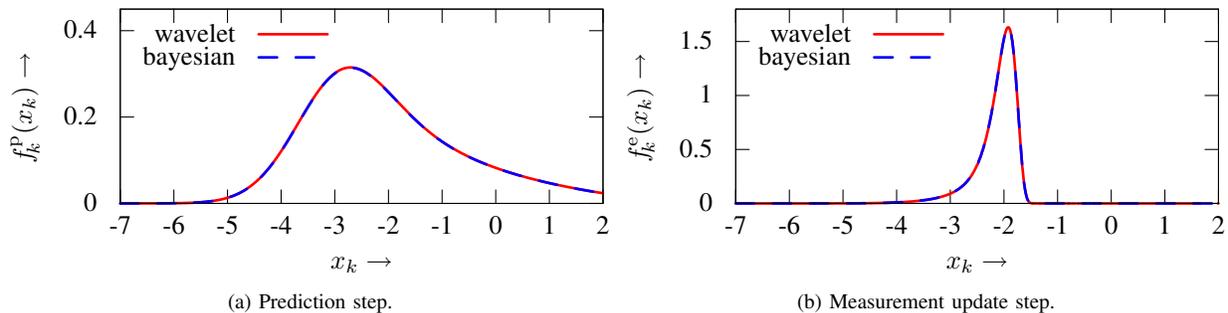


Fig. 2: Exemplary plot of a prediction and a measurement update step of the wavelet-based Bayesian estimator in comparison to the true density.

	bayesian	pf 200	pf 1000	wavelet
rmse	0.024556	0.118244	0.042934	0.024558

TABLE I: Average root mean square error (rmse) over 60 runs.

the particle filter, the first with 200 and the second with 1000 particles. Figure 3 depicts the root mean square error (rmse) for the estimated expected values of the true Bayesian estimator, the wavelet-based Bayesian estimator as well as the larger particle filter instance. The smaller particle filter instance is left out to retain readability of the figure. In Table I, the average root mean square errors over the runs are shown.

A comparison of the number of significant coefficients of the wavelet basis that are determined by a hard thresholding of 0.00001 (average of 129 coefficients) with the constant number of particles for each test run shows that the proposed estimator uses less coefficients and yet, outrules both of the other particle filters in most of all cases (small: 52 of 60, large: 43 of 60).

VII. CONCLUSIONS AND FUTURE WORK

We presented a nonlinear estimator operating on wavelet expansions with constant complexity over time. For the two alternating processing steps of a Bayesian estimator, the counterparts in the wavelet domain were derived, whose time complexities are linear with respect to the non-zero coefficients. In particular, the prediction step can be calculated in an exact way, whereas for the filtering step, a fast and adequate approximation is employed. Together with the approximation quality of wavelets, a good performance is achieved.

To further compress the representation of densities, future work will take other kinds of wavelets into consideration. Extensions like curvelets or contourlets seem to be promising [16]. More enhancements concerning the run-time could also be achieved by thresholding the coefficients or neglecting finer scales. Additionally, the presented estimator will be integrated in a nonlinear model predictive control framework. In particular, if the value function is represented by its wavelet expansion, an effective and adaptive evaluation of the Bellman equation can be achieved.

REFERENCES

- [1] R. E. Kalman, "A new Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME, Journal of Basic Engineering*, no. 82, pp. 35–45, 1960.
- [2] D. Simon, *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*, 1st ed. Wiley & Sons, 2006.
- [3] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," in *Proceedings of the IEEE*, vol. 92, no. 3, 2004, pp. 401–422.
- [4] M. F. Huber and U. D. Hanebeck, "Gaussian Filter based on Deterministic Sampling for High Quality Nonlinear Estimation," in *Proceedings of the 17th IFAC World Congress (IFAC 2008)*, vol. 17, no. 2, Seoul, Korea, Jul. 2008.
- [5] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, February 2002.
- [6] M. F. Huber, D. Brunn, and U. D. Hanebeck, "Efficient Nonlinear Measurement Updating based on Gaussian Mixture Approximation of Conditional Densities," in *Proceedings of the 2007 American Control Conference (ACC 2007)*, New York, New York, Jul. 2007, pp. 4425–4430.
- [7] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian Estimation using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, Aug. 1972.
- [8] R. Kronmal and M. Tarter, "The Estimation of Probability Densities and Cumulatives by Fourier Series Methods," *Journal of the American Statistical Association*, vol. 63, no. 323, pp. 925–952, 1968.
- [9] D. Brunn, F. Sawo, and U. D. Hanebeck, "Nonlinear Multidimensional Bayesian Estimation with Fourier Densities," in *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, Dec. 2006, pp. 1303–1308.
- [10] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Athena Scientific, Belmont, Massachusetts, 2000, vol. 1.
- [11] E. Driver and D. Morrell, "Implementation of Continuous Bayesian Networks Using Sums of Weighted Gaussians," in *Proceedings of the Proceedings of the Eleventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)*. Morgan Kaufmann, 1995, pp. 134–14.
- [12] I. Daubechies, *Ten Lectures on Wavelets*, ser. CBMS-NSF Reg. Conf. Series in Applied Math. SIAM, 1992.
- [13] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 2008.
- [14] A. Latto, H. L. Resnikoff, and E. Tenenbaum, "The Evaluation of Connection Coefficients of Compactly Supported Wavelets," in *In Proceedings of the Princeton Conference on Wavelets and Turbulence*. Springer-Verlag, 1991.
- [15] B. Program and G. Beylkin, "On The Fast Algorithm for Multiplication of Functions in the Wavelet Bases," in *Proceedings of the International Conference "Wavelets and Applications"*. Editions Frontiers, 1992, pp. 259–273.
- [16] E. J. Candes, E. J. C., and D. L. Donoho, "Curvelets - a surprisingly effective nonadaptive representation for objects with edges." Tech. Report Statistics Depart.