Progressive Bayesian Estimation for Nonlinear Discrete–Time Systems: The Filter Step for Scalar Measurements and Multidimensional States

Uwe D. Hanebeck, Olga Feiermann
Institute of Computer Design and Fault Tolerance
Universität Karlsruhe
76128 Karlsruhe, Germany
Uwe.Hanebeck@ieee.org

***Abstract*— This paper is concerned with recursively estimating the internal state sequence of a discrete–time dynamic system by processing a sequence of noisy measurements taken from the system output. Recursive processing requires some kind of sufficient statistic for representing the information collected up to a certain time step. For this purpose, the probability density functions of the state are especially well suited. Once they are available, almost any type of point estimate, e.g. mean, mode, or median, can be derived. In the case of continuous states, however, the exact probability density functions characterizing the state estimate are in general either not feasible or not well suited for recursive processing. Hence, approximations of the true densities are generally inevitable, where Gaussian mixture approximations are convenient for a number of reasons. However, calculating appropriate mixture parameters that minimize a global measure of deviation from the true density is a tough optimization task. Here, we propose a new approximation method that minimizes the squared integral deviation between the true density and its mixture approximation. Rather than trying to solve the original problem, it is converted into a corresponding system of explicit ordinary first–order differential equations. This system of differential equations is then solved over a finite "time" interval, which is an efficient way of calculating the desired optimal parameter values. We focus on the measurement update in the important case of vector states and scalar measurements. In addition, approximation densities with separable kernels are assumed. It will be shown, that if the measurement nonlinearities are also separable, the required multidimensional integrals can be reduced to the product of one–dimensional integrals. For several important types of measurement functions including polynomial measurement nonlinearities, closed–form analytic expressions for the coefficients of the system of differential equations are available.**

I. Introduction

Often, the internal state sequence of a dynamic system is not directly available and has to be reconstructed from the system input sequence and a measurement sequence supplied by sensor devices. Here, we assume discrete–time systems with continuous states, where the measurements and the inputs are nonlinearly related to the states.

The goal is to make an estimate of the state sequence available at every time step. Of course, this estimate should incorporate the information contained in all the input and measurement samples collected up to that time step. Instead of storing all data and reprocessing them at every time step, a recursive estimator is preferred, that uses some kind of sufficient statistic as an *exact* compressed representation of the collected data. For that purpose, the probability density functions of the state are well suited. Once they are available, almost any type of point estimate, e.g. mean, mode, or median, can be derived.

In the case of continuous states, however, the exact probability density functions characterizing the state estimate are in general either not feasible or not well suited for recursive processing. Hence, approximations of the true densities are generally inevitable. Several different choices for representing the density of the state estimate are possible. A Gaussian mixture approximation is especially convenient as its moments can be calculated analytically.

Early approaches to analytic nonlinear estimation used Gaussian mixture approximations together with individual updating of the mixture components [1], which yields suboptimal results. On the other hand, systematically minimizing a measure of distance between the true density and its approximation by calculating appropriate density parameters generally is a tough optimization task. Numerical algorithms such as the Expectation–Maximation (EM) algorithm [8] or gradient based schemes suffer from the local minima problem, i.e., their results strongly depend upon the initialization. In addition, convergence may be slow. In the context of density estimation, a deterministic annealing EM algorithm has been proposed to overcome these problems [11]. Beginning with an unimodal objective function at a high temperature, the objective function gradually approaches the original function as the temperature decreases. This method increases the probability of converging to a global optimum. Similar approaches based on moving from a tractable density to the desired density via a sequence of intermediate densities have been proposed in the context of particle filters, see chapter 12 of [4] and [7]. An alternative approach to guarantee convergence of the EM algorithm is based on modifying the number of mixture components [12], [13].

In this paper, a new estimator is introduced, which minimizes the squared integral deviation between the true density and its Gaussian mixture approximation. It is based on a general framework for estimator design presented in [6]. In order to minimize the given distance measure, *both parametric and structural* adaptations of the approximation density are performed. For that purpose, a parameterized true density is introduced, which starts from a tractable density and *continuously* approaches the exact density to be approximated. Based on this type of progressive processing, the original optimization problem is converted into a corresponding system of explicit ordinary first–order differential equations. The desired optimal density parameters are then

calculated by solving the differential equations over a finite "time" interval. Structural adaptation of the approximation density is performed during the progression in order to modify the local approximation capabilities of the mixture approximation by changing the number of components.

The paper is organized as follows. The next section gives a formulation of the estimation problem. The set of differential equations for parametric adaptation is derived in Section III. Structural adaptation is discussed in Section IV.

II. Problem Formulation

A discrete–time dynamic system is considered, where scalar measurements $\hat{z}_k$ at time $t_k = k\,T_s$ are related to the state $\underline{x}_k$ via the *measurement equation*

$$\hat{z}_k = h_k(\underline{x}_k) + v_k \ .$$

v_k is additive Gaussian noise with density

$$f_k^v(v) = \frac{1}{\sqrt{2\,\pi}\sigma_k^v} \exp\left\{-\frac{1}{2}\frac{v_k^2}{(\sigma_k^v)^2}\right\} \ .$$

In the context of this paper we focus on two–dimensional states $\underline{x}_k = [x \quad y]^T \in \mathbb{R}^2$ without loss of generality to keep the notation at an acceptable complexity.

Estimators will first be derived for general measurement nonlinearities $h_k(\underline{x}_k)$. In III-B.2, simplified estimators will then be derived for the special case of a sum of separable functions of the form

$$h_k(\underline{x}_k) = h_k^{x,1}(x)\,h_k^{y,1}(y) + h_k^{x,2}(x)\,h_k^{y,2}(y) + \ldots$$

Given a predicted density $f_k^p(\underline{x}_k)$, a new measurement is included by means of the *filter step* or *measurement update* according to Bayes' law [10]

$$\tilde{f}_k^e(\underline{x}_k) = c_k\,f_k^p(\underline{x}_k)\,\tilde{f}_k^L(\underline{x}_k) \ ,$$

where $\tilde{f}_k^L(\underline{x}_k)$ is the so–called *likelihood* and c_k is a normalization constant. Exact densities are denoted by a tilde, e.g. $\tilde{f}_k$. For the case of additive measurement noise considered here, the exact likelihood is given by

$$\tilde{f}_k^L(\underline{x}_k) = f_k^v(\hat{z}_k - h_k(\underline{x}_k)) \ . \tag{1}$$

The goal is to approximate the exact density $\tilde{f}_k^e(\underline{x}_k)$ by means of a Gaussian mixture density. We assume separable approximation densities

$$\begin{aligned} f_k^e(\underline{x}_k, \underline{\eta}_k) &= \sum_{i=1}^{L_k^e} w_k^{(e,i)} f_k^{(e,i)}(\underline{x}_k, \underline{\eta}_k^{(i)}) \\ &= \sum_{i=1}^{L_k^e} w_k^{(e,i)} \mathcal{N}\left(x_k - \hat{x}_k^{(e,i)}, \rho_k^{(e,i)}\right) \\ &\qquad \mathcal{N}\left(y_k - \hat{y}_k^{(e,i)}, \sigma_k^{(e,i)}\right) \end{aligned}$$

for that purpose, where L_k^e is the number of components. $\mathcal{N}(x-m,\sigma)$ denotes a Gaussian density with mean m and standard deviation σ. $w_k^{(e,i)}$ are weighting coefficients with $w_k^{(e,i)} > 0$ and $\sum_{i=1}^{L_k^e} w_k^{(e,i)} = 1$.

For a compact description of the approximation density, we introduce a parameter vector

$$\underline{\eta}_k = \left[\left(\underline{\eta}_k^{(1)}\right)^T \quad \left(\underline{\eta}_k^{(2)}\right)^T \quad \cdots \quad \left(\underline{\eta}_k^{(L)}\right)^T\right]^T$$

with

$$\underline{\eta}_k^{(i)} = \left[w_k^{(e,i)} \quad \hat{x}_k^{(e,i)} \quad \rho_k^{(e,i)} \quad \hat{y}_k^{(e,i)} \quad \sigma_k^{(e,i)}\right]^T \ .$$

An optimal parameter vector is desired, that minimizes the squared integral distance

$$G_k(\underline{\eta}_k) = \frac{1}{2}\int_{\mathbb{R}^2}\left(\tilde{f}_k^e(\underline{x}_k) - f_k^e(\underline{x}_k, \underline{\eta}_k)\right)^2 d\underline{x}_k \tag{2}$$

between the exact density and its approximation. Hence, the estimation problem is reduced to an optimization problem, which consists of calculating the smallest set of parameters collected in a parameter vector $\underline{\eta}_k^{opt}$ for which the distance measure attains its minimum and is below a pre–specified threshold G_k^{max}, i.e., $G_k(\underline{\eta}_k^{opt}) < G_k^{max}$.

The main difficulty in calculating the optimal vector $\underline{\eta}_k^{opt}$ is the existence of local minima. Hence, application of numerical minimization routines generally does not yield the desired optimal parameter vector. In addition, the run time of numerical minimization routines depends upon the specific problem and, in general, is not known a priori.

The following two sections are concerned with a new method for calculating the desired parameter vector $\underline{\eta}_k^{opt}$, that does not rely on numerical search and optimization techniques. For the sake of simplicity, the time index k will be omitted in the corresponding derivations.

III. Parametric Adaptation

The key idea of the new approach is to perform progressive processing. Hence, instead of directly approximating the true density, we start with a tractable density that continuously approaches the true density via intermediate densities. This is achieved by parameterizing the exact likelihood. For that purpose, a progression parameter γ is introduced, which varies between zero and one. For $\gamma = 0$, the parameterized likelihood $\tilde{f}^L(\underline{x}, \gamma = 0)$ is initialized with some kind of density, that is simple to approximate. For $\gamma = 1$, the parameterized likelihood $\tilde{f}^L(\underline{x}, \gamma = 1)$ attains the exact likelihood $\tilde{f}^L(\underline{x})$.

A convenient type of progression schedule is obtained by starting with large measurement noise, which is continuously reduced until the desired standard deviation σ^v is obtained. For that purpose, we define a parameterized noise density $f^v(v,\gamma)$ with a standard deviation according to

$$\bar{\sigma}^v(\gamma) = \frac{1+\epsilon}{\gamma+\epsilon}\sigma^v = \begin{cases} \text{large} & \text{for } \gamma = 0 \\ \sigma^v & \text{for } \gamma = 1 \end{cases} \ ,$$

where ϵ is a small constant and $\gamma \in [0,1]$.

Since Gaussian mixture densities can easily be normalized, the following derivations will be conducted for an unnormalized likelihood

$$\tilde{f}^L(\underline{x}, \gamma) = \exp\left\{-\frac{1}{2}\frac{(\hat{z} - h(\underline{x}))^2}{(\bar{\sigma}^v(\gamma))^2}\right\}$$

and an unnormalized true posterior

$$\tilde{f}^e(\underline{x},\gamma) = f^p(\underline{x})\,\tilde{f}^L(\underline{x},\gamma) \ . \tag{3}$$

The distance measure between the parameterized exact density $\tilde{f}^e(\underline{x},\gamma)$ and its approximation $f^e(\underline{x},\underline{\eta})$ now depends on γ

$$G(\underline{\eta},\gamma) = \frac{1}{2}\int_{\mathbb{R}^2}\left(\tilde{f}^e(\underline{x},\gamma) - f^e(\underline{x},\underline{\eta})\right)^2 d\underline{x} \ .$$

We assume a nominal parameter vector $\bar{\underline{\eta}}$ to be given and consider only small deviations $\Delta\underline{\eta}(\gamma)$ according to $\underline{\eta}(\gamma) = \bar{\underline{\eta}} + \Delta\underline{\eta}(\gamma)$. Around the nominal parameter vector $\bar{\underline{\eta}}$, the approximation density is replaced by a Taylor series expansion up to first order

$$f^e(\underline{x},\underline{\eta}) \approx f^e(\underline{x},\bar{\underline{\eta}}) + \underline{F}^T(\underline{x},\bar{\underline{\eta}})\,\Delta\underline{\eta}(\gamma)$$

with

$$\underline{F}(\underline{x},\bar{\underline{\eta}}) = \left.\frac{\partial\, f^e(\underline{x},\underline{\eta})}{\partial\,\underline{\eta}}\right|_{\underline{\eta}=\bar{\underline{\eta}}} \ . \tag{4}$$

The distance measure $G(\underline{\eta},\gamma)$ can now be rewritten accordingly

$$G(\underline{\eta},\gamma) \approx \frac{1}{2}\int_{\mathbb{R}^2}\Big(\tilde{f}^e(\underline{x},\gamma) - f^e(\underline{x},\bar{\underline{\eta}}) - \underline{F}^T(\underline{x},\bar{\underline{\eta}})\,(\underline{\eta}-\bar{\underline{\eta}})\Big)^2 d\underline{x} \ .$$

Taking the partial derivative of the distance measure $G(\underline{\eta},\gamma)$ with respect to the parameter vector $\underline{\eta}$ and setting the result to zero, i.e., $\partial G/\partial\,\underline{\eta} \stackrel{!}{=} 0$, gives

$$\int_{\mathbb{R}^2}\left(\tilde{f}^e(\underline{x},\gamma) - f^e(\underline{x},\bar{\underline{\eta}}) + \underline{F}^T(\underline{x},\bar{\underline{\eta}})\,\bar{\underline{\eta}}\right)\underline{F}(\underline{x},\bar{\underline{\eta}})\,d\underline{x} = \left(\int_{\mathbb{R}^2}\underline{F}(\underline{x},\bar{\underline{\eta}})\,\underline{F}^T(\underline{x},\bar{\underline{\eta}})\,d\underline{x}\right)\underline{\eta}(\gamma) \ .$$

The partial derivative with respect to γ gives the desired system of explicit ordinary first–order differential equations

$$\int_{\mathbb{R}^2}\frac{\partial\,\tilde{f}^e(\underline{x},\gamma)}{\partial\,\gamma}\,\underline{F}(\underline{x},\bar{\underline{\eta}})\,d\underline{x} = \left(\int_{\mathbb{R}^2}\underline{F}(\underline{x},\bar{\underline{\eta}})\,\underline{F}^T(\underline{x},\bar{\underline{\eta}})\,d\underline{x}\right)\frac{\partial\,\underline{\eta}}{\partial\,\gamma} \ ,$$

which upon replacing $\bar{\underline{\eta}}$ by $\underline{\eta}$ can be written as

$$\underline{b}(\underline{\eta},\gamma) = \mathbf{P}(\underline{\eta})\,\dot{\underline{\eta}}$$

with

$$\underline{b}(\underline{\eta},\gamma) = \int_{\mathbb{R}^2}\frac{\partial\,\tilde{f}^e(\underline{x},\gamma)}{\partial\,\gamma}\,\underline{F}(\underline{x},\underline{\eta})\,d\underline{x} \tag{5}$$

$$\mathbf{P}(\underline{\eta}) = \int_{\mathbb{R}^2}\underline{F}(\underline{x},\underline{\eta})\,\underline{F}^T(\underline{x},\underline{\eta})\,d\underline{x} \ ,$$

and $\dot{\underline{\eta}} = \partial\,\underline{\eta}/\partial\,\gamma$. We will now derive analytic expressions for $\underline{b}(\underline{\eta},\gamma)$ and $\mathbf{P}(\underline{\eta})$.

A. Analytic Expression for $\mathbf{P}(\underline{\eta})$

$\mathbf{P}(\underline{\eta})$ is composed of $(L^e)^2$ five–by–five block matrices according to

$$\mathbf{P}(\underline{\eta}) = \int_{\mathbb{R}^2}\underline{F}(\underline{x},\underline{\eta})\,\underline{F}^T(\underline{x},\underline{\eta})\,d\underline{x} = \begin{bmatrix}\mathbf{P}^{(1,1)} & \mathbf{P}^{(1,2)} & \dots & \mathbf{P}^{(1,L^e)}\\ \mathbf{P}^{(2,1)} & \mathbf{P}^{(2,2)} & \dots & \mathbf{P}^{(2,L^e)}\\ \vdots & \vdots & & \vdots\\ \mathbf{P}^{(L^e,1)} & \mathbf{P}^{(L^e,2)} & \dots & \mathbf{P}^{(L^e,L^e)}\end{bmatrix} \ .$$

The individual block matrices $\mathbf{P}^{(i,j)}$ for $i = 1,\dots,L^e$ and $j = 1,\dots,L^e$ with

$$P^{(i,j)}_{n,m} = \int_{\mathbb{R}^2}\frac{\partial\, f_i(\underline{x},\underline{\eta}_i)}{\partial\,\underline{\eta}_{i,n}}\frac{\partial\, f_j(\underline{x},\underline{\eta}_j)}{\partial\,\underline{\eta}_{j,m}}\,d\underline{x}$$

and $\underline{\eta}_{i,1} = w^e_i$, $\underline{\eta}_{i,2} = \hat{x}^e_i$, $\underline{\eta}_{i,3} = \rho^e_i$, $\underline{\eta}_{i,2} = \hat{y}^e_i$, $\underline{\eta}_{i,3} = \sigma^e_i$ for $n = 1,\dots,5$ and $m = 1,\dots,5$ can be obtained analytically according to the expression in Figure 1 with

$$R^{i,j}_{n,m} = P^{i,j}_{n,m}(w_i, w_j, x_i, x_j, r_i, r_j) \ ,$$
$$S^{i,j}_{n,m} = P^{i,j}_{n,m}(w_i, w_j, y_i, y_j, s_i, s_j) \ ,$$

and $w_i = w^e_i$, $w_j = w^e_j$, $x_i = \hat{x}^e_i$, $x_j = \hat{x}^e_j$, $r_i = \rho^e_i$, $r_j = \rho^e_j$, $y_i = \hat{y}^e_i$, $y_j = \hat{y}^e_j$, $s_i = \sigma^e_i$, $s_j = \sigma^e_j$. The corresponding elements are given in Figure 2.

B. Simplified Expressions for $\underline{b}(\underline{\eta},\gamma)$

$\underline{b}(\underline{\eta},\gamma)$ from (5) can be decomposed into subvectors according to

$$\underline{b}(\underline{\eta},\gamma) = \begin{bmatrix}\underline{b}^T_1(\underline{\eta},\gamma) & \underline{b}^T_2(\underline{\eta},\gamma) & \cdots & \underline{b}^T_L(\underline{\eta},\gamma)\end{bmatrix}^T \ ,$$

where a subvector $\underline{b}_i(\underline{\eta},\gamma)$ is obtained with (3) and (4) as

$$\underline{b}_i(\underline{\eta},\gamma) = \int_{\mathbb{R}^2} f^p(\underline{x})\,\frac{\partial\,\tilde{f}^L(\underline{x},\gamma)}{\partial\,\gamma}\,\frac{\partial\, f^e(\underline{x},\underline{\eta})}{\partial\,\underline{\eta}_i}\,d\underline{x} \ .$$

For two–dimensional states $\underline{x}_k \in \mathbb{R}^2$ we have

$$\frac{\partial\, f^e(\underline{x},\underline{\eta})}{\partial\,\underline{\eta}_i} = f^e_i\left(\underline{x},\underline{\eta}_i\right)\underline{\Theta}^x_i(x)\odot\underline{\Theta}^y_i(y)$$

with

$$\underline{\Theta}^x_i(x) = \begin{bmatrix}\frac{1}{\sqrt{w^e_i}}\\ \frac{x-\hat{x}^e_i}{\left(\rho^e_i\right)^2}\\ 1\\ \frac{(x-\hat{x}^e_i)^2-(\rho^e_i)^2}{\left(\rho^e_i\right)^3}\\ 1\end{bmatrix} \ , \ \underline{\Theta}^y_i(y) = \begin{bmatrix}\frac{1}{\sqrt{w^e_i}}\\ 1\\ \frac{y-\hat{y}^e_i}{\left(\sigma^e_i\right)^2}\\ 1\\ \frac{(y-\hat{y}^e_i)^2-(\sigma^e_i)^2}{\left(\sigma^e_i\right)^3}\end{bmatrix} \ ,$$

where $\odot$ denotes the element–by–element product, and

$$\frac{\partial\,\tilde{f}^L(\underline{x},\gamma)}{\partial\,\gamma} = c^L\left(\hat{z} - h(\underline{x})\right)^2\tilde{f}^L(\underline{x},\gamma)$$

with

$$c^L = -\frac{(\epsilon+\gamma)}{(1+\epsilon)^2(\sigma^v)^2} \ .$$

$$\mathbf{P}^{(i,j)} = \frac{1}{\sqrt{2\,\pi\,(r_i^2+r_j^2)}}\,\exp\left(-\frac{1}{2}\,\frac{(x_i-x_j)^2}{r_i^2+r_j^2}\right)\,\frac{1}{\sqrt{2\,\pi\,(s_i^2+s_j^2)}}\,\exp\left(-\frac{1}{2}\,\frac{(y_i-y_j)^2}{s_i^2+s_j^2}\right)$$

$$\begin{bmatrix} \begin{bmatrix} 1 & R_{1,2}^{i,j} & R_{1,3}^{i,j} \\ R_{2,1}^{i,j} & R_{2,2}^{i,j} & R_{2,3}^{i,j} \\ R_{3,1}^{i,j} & R_{3,2}^{i,j} & R_{3,3}^{i,j} \end{bmatrix} & \begin{bmatrix} 1 \\ R_{2,1}^{i,j} \\ R_{3,1}^{i,j} \end{bmatrix} \cdot \begin{bmatrix} S_{1,2}^{i,j} & S_{1,3}^{i,j} \end{bmatrix} \\ \left(\begin{bmatrix} 1 \\ R_{2,1}^{i,j} \\ R_{3,1}^{i,j} \end{bmatrix} \cdot \begin{bmatrix} S_{1,2}^{i,j} & S_{1,3}^{i,j} \end{bmatrix} \right)^T & \begin{bmatrix} S_{2,2}^{i,j} & S_{2,3}^{i,j} \\ S_{3,2}^{i,j} & S_{3,3}^{i,j} \end{bmatrix} \end{bmatrix}$$

Fig. 1. The matrices $\mathbf{P}^{(i,j)}$ for $i = 1, \ldots, L^e$ and $j = 1, \ldots, L^e$.

$$\begin{aligned}
P_{1,2}^{i,j}(w_i, w_j, x_i, x_j, r_i, r_j) &= w_j\,\frac{x_i - x_j}{r_i^2 + r_j^2}\ , \\
P_{1,3}^{i,j}(w_i, w_j, x_i, x_j, r_i, r_j) &= w_j\,r_j\,\frac{(x_i - x_j)^2 - (r_i^2 + r_j^2)}{(r_i^2 + r_j^2)^2}\ , \\
P_{2,1}^{i,j}(w_i, w_j, x_i, x_j, r_i, r_j) &= w_i\,\frac{x_j - x_i}{r_i^2 + r_j^2}\ , \\
P_{2,2}^{i,j}(w_i, w_j, x_i, x_j, r_i, r_j) &= w_i\,w_j\,\frac{r_i^2 + r_j^2 - (x_i - x_j)^2}{(r_i^2 + r_j^2)^2}\ , \\
P_{2,3}^{i,j}(w_i, w_j, x_i, x_j, r_i, r_j) &= w_i\,w_j\,r_j\,\frac{(x_j - x_i)((x_i - x_j)^2 - 3(r_i^2 + r_j^2))}{(r_i^2 + r_j^2)^3}\ , \\
P_{3,1}^{i,j}(w_i, w_j, x_i, x_j, r_i, r_j) &= w_i\,r_i\,\frac{(x_i - x_j)^2 - (r_i^2 + r_j^2)}{(r_i^2 + r_j^2)^2}\ , \\
P_{3,2}^{i,j}(w_i, w_j, x_i, x_j, r_i, r_j) &= w_i\,w_j\,r_i\,\frac{(x_i - x_j)((x_i - x_j)^2 - 3(r_i^2 + r_j^2))}{(r_i^2 + r_j^2)^3}\ , \\
P_{3,3}^{i,j}(w_i, w_j, x_i, x_j, r_i, r_j) &= w_i\,w_j\,r_i\,r_j\,\frac{(x_i - x_j)^4 + 3(r_i^2 + r_j^2)(r_i^2 + r_j^2 - 2(x_i - x_j)^2)}{(r_i^2 + r_j^2)^4}\ .
\end{aligned}$$

Fig. 2. The elements of the matrices $\mathbf{P}^{(i,j)}$ for $i = 1, \ldots, L^e$ and $j = 1, \ldots, L^e$.

Hence, we obtain

$$\underline{b}_i(\underline{\eta}, \gamma) = c^L \int_{\mathbb{R}^2} f^p(\underline{x})\ (\hat{z} - h(\underline{x}))^2\ \tilde{f}^L(\underline{x}, \gamma)\ f_i^e\left(\underline{x}, \underline{\eta}_i\right)\ \underline{\Theta}_i^x(x) \odot \underline{\Theta}_i^y(y)\, d\underline{x}\ .$$

An alternative to numerical integration for solving this type of integrals is given in [9].

1) First Simplification: True Density Replaced $\tilde{f}^e(\underline{x})$ by its Approximation $f^e(\underline{x}, \underline{\eta})$: For simplifying the expression, we assume that our current approximation $f^e(\underline{x}, \underline{\eta})$ of the true posterior density $\tilde{f}^e(\underline{x})$ is good enough for replacing $\tilde{f}^e(\underline{x}) = f^p(\underline{x})\,\tilde{f}^L(\underline{x}, \gamma)$. The resulting expression $f^e(\underline{x}, \underline{\eta})\,f_i^e\left(\underline{x}, \underline{\eta}_i\right)$ is converted to a Gaussian mixture given as a sum of separable components

$$\begin{aligned}
f_i^s(\underline{x}) &= \sum_{j=1}^{L^e} f_i^{s,x}(x)\,f_i^{s,y}(y) \\
&= \sum_{j=1}^{L^e} w_{i,j}^s\,\mathcal{N}\left(x - \hat{x}_i^s, \rho_i^s\right)\,\mathcal{N}\left(y - \hat{y}_i^s, \sigma_i^s\right)\ ,
\end{aligned}$$

with weighting coefficients $w_{i,j}^s$, mean vectors $\underline{\hat{x}}_{i,j}^s$, and covariance matrices $\mathbf{C}_{i,j}^s$.

Finally, we obtain

$$\underline{b}_i(\underline{\eta}, \gamma) = c^L \int_{\mathbb{R}^2} (\hat{z} - h(\underline{x}))^2 f_i^{s,x}(x)\,f_i^{s,y}(y)\ \underline{\Theta}_i^x(x) \odot \underline{\Theta}_i^y(y)\, d\underline{x}\ .$$

2) Second Simplification: Separable Measurement Functions: In addition to approximation densities given by a sum of separable densities, we now assume that the measurement functions are also sums of separable functions of the form

$$h(\underline{x}) = h^{x,1}(x)\, h^{y,1}(y) + h^{x,2}(x)\, h^{y,2}(y) + \dots \ ,$$

where considering only one term according to

$$h(\underline{x}) = h^{x}(x)\, h^{y}(y)$$

is sufficient for the following derivations. Hence, $\underline{b}_i(\underline{\eta}, \gamma)$ can be rewritten as

$$\underline{b}_i(\underline{\eta}, \gamma) = c^L \int_{\mathbb{R}^2} \left(\hat{z} - h^x(x)\, h^y(y)\right)^2 f_i^{s,x}(x)\, f_i^{s,y}(y)\, \underline{\Theta}_i^x(x) \odot \underline{\Theta}_i^y(y)\, d\underline{x} \ ,$$

which gives

$$\begin{aligned}
\underline{b}_i(\underline{\eta}, \gamma) = c^L \hat{z}^2 & \int_{\mathbb{R}} f_i^{s,x}(x)\, \underline{\Theta}_i^x(x)\, dx \\
& \odot \int_{\mathbb{R}} f_i^{s,y}(y)\, \underline{\Theta}_i^y(y)\, dy \\
-2c^L \hat{z} & \int_{\mathbb{R}} h^x(x)\, f_i^{s,x}(x)\, \underline{\Theta}_i^x(x)\, dx \\
& \odot \int_{\mathbb{R}} h^y(y)\, f_i^{s,y}(y)\, \underline{\Theta}_i^y(y)\, dy \\
+c^L & \int_{\mathbb{R}} \left(h^x(x)\right)^2\, f_i^{s,x}(x)\, \underline{\Theta}_i^x(x)\, dx \\
& \odot \int_{\mathbb{R}} \left(h^y(y)\right)^2\, f_i^{s,y}(y)\, \underline{\Theta}_i^y(y)\, dy \ .
\end{aligned}$$

Analytic solutions for these integrals are available for a wide variety of interesting measurement nonlinearities $h^x(x)$ and $h^y(y)$. The corresponding formulae for polynomial measurement nonlinearities are given in [5].

IV. Structural Adaptation

During the progression from $\gamma = 0$ to $\gamma = 1$, a continuous validation of the deviation between the true density and its approximation is performed. For that purpose a *normalized* distance measure

$$G_N(\underline{\eta}, \gamma) = \frac{\int_{\mathbb{R}^2} \left(\tilde{f}^e(\underline{x}, \gamma) - f^e(\underline{x}, \underline{\eta})\right)^2 d\underline{x}}{\int_{\mathbb{R}^2} \left(\tilde{f}^e(\underline{x}, \gamma)\right)^2 d\underline{x} + \int_{\mathbb{R}^2} \left(f^e(\underline{x}, \underline{\eta})\right)^2 d\underline{x}}$$

is used instead of the unnormalized distance measure in (2). The normalized distance measure is more appropriate for specifying deviation tolerances as it ranges between 0 and 1. A perfect match is indicated by $G_N(\underline{\eta}, \gamma) = 0$, the maximum deviation between the true density and its approximation is indicated by $G_N(\underline{\eta}, \gamma) = 1$.

As long as $G_N(\underline{\eta}, \gamma)$ is within a prespecified tolerance band, i.e., $G_N^L < G_N(\underline{\eta}, \gamma) < G_N^U$, the progression is continued. Once the normalized distance measure is larger than a pre–specified threshold, i.e., $G_N(\underline{\eta}, \gamma) > G_{N,max}$, the most critical mixture component responsible for the deviation is identified by evaluating L^e individual distance measures according to

$$G_i(\underline{\eta}, \gamma) = \int_{\mathbb{R}^2} \left(\tilde{f}^e(\underline{x}, \gamma) - f^e(\underline{x}, \underline{\eta})\right)^2 f_i^e\left(\underline{x}, \underline{\eta}_i\right) d\underline{x}$$

for $i = 1, \dots, L^e$. The most critical component is then replaced in each dimension by a mixture of two components according to

$$\begin{aligned}
& w\, \exp\left\{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right\} \\
& \overset{!}{\approx} w_1\, \exp\left\{-\frac{1}{2}\frac{(x-m_1)^2}{\sigma_1^2}\right\} + w_2\, \exp\left\{-\frac{1}{2}\frac{(x-m_2)^2}{\sigma_2^2}\right\} \ .
\end{aligned}$$

In order to minimize the deviation of the mixture from the original component, the parameters are selected as

$$\begin{aligned}
w_1 = w_2 &= \frac{w}{2} \ , \\
m_1 = m - \epsilon \, , \ m_2 &= m + \epsilon \ , \\
\sigma_1 = \sigma_2 &= \sigma \ ,
\end{aligned} \tag{6}$$

where ϵ is a "small" constant.

An alternative replacement method based on a precalculated library of Gaussian mixtures is described in [5].

When the number of mixture components becomes too large, i.e., the normalized distance measure is smaller than the prespecified threshold according to $G_N(\underline{\eta}, \gamma) < G_N^L$, merging of mixture components is performed. For that purpose, two Gaussian densities in one dimension are merged into a single one according to

$$\begin{aligned}
& w_1\, \exp\left\{-\frac{1}{2}\frac{(x-m_1)^2}{\sigma_1^2}\right\} + w_2\, \exp\left\{-\frac{1}{2}\frac{(x-m_2)^2}{\sigma_2^2}\right\} \overset{!}{\approx} \\
& w\, \left\{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right\}
\end{aligned}$$

with

$$\begin{aligned}
w &= w_1 + w_2 \ , \\
m &= \frac{w_1\, m_1 + w_2\, m_2}{w_1 + w_2} \ , \\
\sigma &= \frac{w_1\, \sigma_1 + w_2\, \sigma_2}{w_1 + w_2} \ .
\end{aligned}$$

when the corresponding parameters are close.

In addition, components are removed from the mixture once the corresponding weight becomes less than or equal to zero.

The block diagram of the progressive Bayesian estimator including the parametric adaptation discussed in Section III and the structural adaptation discussed in this section is shown in Figure 3.

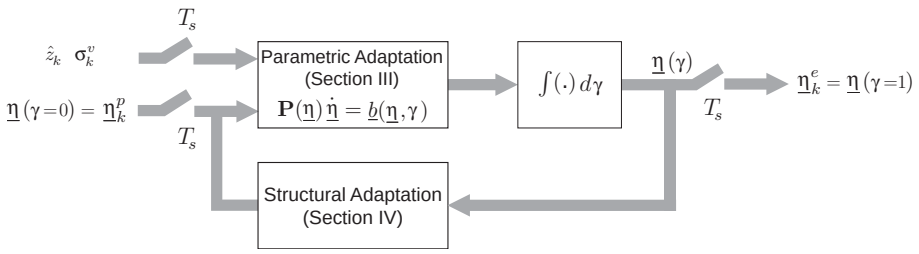

Fig. 3. Block diagram of the progressive Bayesian estimator including the parametric adaptation discussed in Section III and the structural modification discussed in Section IV.

V. Conclusions

A new type of stochastic state estimator for performing the filter step for nonlinear dynamic systems in the case of scalar measurements and multidimensional continuous–valued states has been introduced, which is based on a general framework for progressive Bayesian estimators given in [6]. Without loss of generality, the case of two–dimensional states has been assumed during the derivations. Generalization to states of arbitrary dimensions is straightforward.

The key aspect of the proposed new class of estimators is a new approach for obtaining optimal parameters of a Gaussian mixture approximation, that minimize the squared integral deviation from the true posterior density. Instead of applying numerical search and optimization techniques, which may suffer from local minima, bad convergence, and unpredictable run time, the problem is *exactly* converted to a system of explicit ordinary first–order differential equations. The desired optimal density parameters are then calculated by solving the ordinary differential equations over a finite time interval.

For deriving concrete estimator equations, approximation densities with separable kernels have been assumed. It has been shown, that if the measurement nonlinearities are also separable, the required multidimensional integrals can be reduced to the product of one–dimensional integrals. For the special case of polynomial measurement nonlinearities, closed–form analytic expressions have been derived for the coefficients of the system of explicit ordinary first–order differential equations in [5].

An additional advantage of the new estimation framework is the simplicity of altering the local approximation capabilities of the mixture density by structural adaptations, i.e., by modifying the number of mixture components, during the progression.

The resulting estimators provide an analytic representation of the state densities and fill the gap between simple linear estimators and complex numerical approaches like particle filters [2] or grid–based estimators [3]. They are simple to implement and provide a tradeoff between accuracy and computational complexity, since the estimator performance can be adjusted by specifying the maximum tolerable deviation between the true density and its approximation.

VI. References

[1] D. L. Alspach and H. W. Sorenson. Nonlinear Bayesian Estimation Using Gaussian Sum Approximation. *IEEE Transactions on Automatic Control*, AC–17(4):439–448, 1972.

[2] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A Tutorial on Particle Filters for On–line Non–linear/Non–Gaussian Bayesian Tracking. *IEEE Transactions of Signal Processing*, 50(2):174–188, 2002.

[3] N. Bergman, L. Ljung, and F. Gustafsson. Terrain Navigation Using Bayesian Statistics. *IEEE Control Systems Magazine*, 19(3):33–40, 1999.

[4] A. Doucet, N. De Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.

[5] U. D. Hanebeck. Progressive Bayesian Estimation for Nonlinear Discrete–Time Systems: The Measurement Step. In *Proceedings of the 2003 IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI'03), Tokyo, Japan*, 2003.

[6] U. D. Hanebeck, K. Briechle, and A. Rauh. Progressive Bayes: A New Framework for Nonlinear State Estimation. In *Proceedings of SPIE Vol. 5099, AeroSense Symposium, Orlando, Florida*, 2003.

[7] R. Neal. Annealed Importance Sampling. Technical Report No. 9805, Department of Statistics, University of Toronto, September 1998.

[8] W. B. Poland and R. D. Shachter. Mixtures of Gaussians and Minimum Relative Entropy Techniques for Modeling Continuous Uncertainties. In *Uncertainty in Artificial Intelligence: Proceedings of the Ninth Conference (UAI-1993)*, San Francisco, CA, 1993. Morgan Kaufmann Publishers.

[9] A. Rauh and U. D. Hanebeck. Calculating Moments of Exponential Densities Using Differential Algebraic Equations. *IEEE Signal Processing Letters*, 10(5):144–147, 2003.

[10] F. C. Schweppe. *Uncertain Dynamic Systems*. Prentice–Hall, 1973.

[11] N. Ueda and R. Nakano. Deterministic Annealing EM Algorithm. *Neural Networks*, 11(2):271–282, 1998.

[12] N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton. SMEM Algorithm for Mixture Models. *Neural Computation*, 12(9):2109–2128, 2000.

[13] N. Vlassis and A. Likas. A Greedy EM Algorithm for Gaussian Mixture Learning. *Neural Processing Letters*, 15(1):77–87, 2002.