Universität Karlsruhe (TH), Faculty of Computer Science
Institute of Computer Design and Fault Tolerance
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Prof. Dr.-Ing. Uwe D. Hanebeck

Progressive Bayesian Estimation for Nonlinear Discrete–Time Systems: The Measurement Step

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

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

Abstract

This paper is concerned with estimating the internal state of a dynamic system by processing measurements taken from the system output. An exact analytic representation of the probability density functions characterizing the estimate may not be possible to obtain. Even when available, it may be too complex or not practical because, for example, recursive application is required. Hence, approximations are generally inevitable. 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 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. For polynomial measurement nonlinearities, closed–form analytic expressions for the coefficients of the system of differential equations are derived.

1 Introduction

This paper considers estimating the internal state of a nonlinear dynamic system, where the main focus is on updating a given state estimate by means of measurements of the system output.

An estimate should be made available at every time step and should incorporate all the information gathered so far. Storing all data and reprocessing it at every time step is impractical. Hence, a recursive estimator is required that updates a given estimate based on the current information. We are not interested in point estimates but rather in the entire probability density function characterizing the current estimate. Furthermore, an analytic expression for this density is desired.

In the case of continuous–valued systems, an exact analytic representation of the probability density function characterizing the estimate is generally too complex or not practical for recursive application. In some cases, the exact solution cannot even be explicitly computed. Hence, it is generally required to approximate the exact result even when it is available.

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 [10] 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 [13]. 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 [7, 8]. An alternative approach to guarantee convergence of the EM algorithm is based on modifying the number of mixture components [14, 15].

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. For polynomial measurement nonlinearities, closed–form analytic expression for coefficients of the system of explicit ordinary first–order differential equations are derived. 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 when parametric adaptation is not sufficient to keep the desired measure of deviation within a pre–specified tolerance band.

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 3. Structural adaptation is discussed in Section 4. A numerical example of the new approach is given in [6].

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

2 Problem Formulation

We consider a scalar discrete–time dynamic system, where scalar measurements $\hat{y}_k$ at time $t_k = k\,T_s$ are related to the scalar state x_k via the *measurement equation*

$$\hat{y}_k = h_k(x_k) + v_k \ ,$$

where v_k is additive noise with Gaussian density

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

Given a predicted density $f_k^p(x_k)$, a new measurement is included by means of the *filter step* or *measurement update* according to Bayes' law [12]

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

where $\tilde{f}_k^L(x_k)$ 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 additive measurement noise, the exact likelihood is given by

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

The goal is to approximate the exact density $\tilde{f}_k^e(x_k)$ by means of a Gaussian mixture given by

$$f_k^e(x_k, \underline{\eta}_k) = \sum_{i=1}^{L_k^e} w_k^{(e,i)} \mathcal{N}\left(x_k - \hat{x}_k^{(e,i)}, \sigma_k^{(e,i)}\right) \ ,$$

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 $\sum_{i=1}^{L_k^e} w_k^{(e,i)} = 1$.

An optimal 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 $\underline{\eta}_k^{(i)} = \left[w_k^{(e,i)} \quad \hat{x}_k^{(e,i)} \quad \sigma_k^{(e,i)}\right]^T$ is characterized by the fact that the squared integral distance

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

between the exact density and its approximation is minimized. 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.

3 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 $\tilde{f}^L(x, \gamma = 0)$ is initialized with some kind of density, that is simple to approximate. For $\gamma = 1$, the parameterized likelihood $\tilde{f}^L(x, \gamma = 1)$ attains the exact likelihood $\tilde{f}^L(x)$.

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

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

and an unnormalized true posterior

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

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

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

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

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

$$P_{1,1}^{(i,j)} = 1 \;, \quad P_{1,2}^{(i,j)} = w_j \, \frac{m_i - m_j}{\sigma_i^2 + \sigma_j^2} \;, \quad P_{1,3}^{(i,j)} = w_j \, \sigma_j \, \frac{(m_i - m_j)^2 - (\sigma_i^2 + \sigma_j^2)}{(\sigma_i^2 + \sigma_j^2)^2} \;, \quad P_{2,1}^{(i,j)} = w_i \, \frac{m_j - m_i}{\sigma_i^2 + \sigma_j^2} \;,$$

$$P_{2,2}^{(i,j)} = w_i \, w_j \, \frac{\sigma_i^2 + \sigma_j^2 - (m_i - m_j)^2}{(\sigma_i^2 + \sigma_j^2)^2} \;, \quad P_{2,3}^{(i,j)} = w_i \, w_j \, \sigma_j \, \frac{(m_j - m_i) \, \left((m_i - m_j)^2 - 3\,(\sigma_i^2 + \sigma_j^2)\right)}{(\sigma_i^2 + \sigma_j^2)^3} \;,$$

$$P_{3,1}^{(i,j)} = w_i \, \sigma_i \, \frac{(m_i - m_j)^2 - (\sigma_i^2 + \sigma_j^2)}{(\sigma_i^2 + \sigma_j^2)^2} \;, \quad P_{3,2}^{(i,j)} = w_i \, w_j \, \sigma_i \, \frac{(m_i - m_j) \, \left((m_i - m_j)^2 - 3\,(\sigma_i^2 + \sigma_j^2)\right)}{(\sigma_i^2 + \sigma_j^2)^3} \;,$$

$$P_{3,3}^{(i,j)} = w_i \, w_j \, \sigma_i \, \sigma_j \, \frac{(m_i - m_j)^4 + 3\,(\sigma_i^2 + \sigma_j^2) \, \left(\sigma_i^2 + \sigma_j^2 - 2\,(m_i - m_j)^2\right)}{(\sigma_i^2 + \sigma_j^2)^4} \;.$$

Table 1. The components $P_{n,m}^{(i,j)}$ for $n = 1,2,3$ and $m = 1,2,3$ in (6).

with

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

with $\underline{\eta}_{i,1} = w_i^e$, $\underline{\eta}_{i,2} = \hat{x}_i^e$, and $\underline{\eta}_{i,3} = \sigma_i^e$ for $n = 1,2,3$ and $m = 1,2,3$ can be obtained analytically according to

$$\mathbf{P}^{(i,j)} = \frac{1}{\sqrt{2\,\pi\,(\sigma_i^2 + \sigma_j^2)}} \, \exp\left(-\frac{1}{2} \, \frac{(m_i - m_j)^2}{\sigma_i^2 + \sigma_j^2} \right) \begin{bmatrix} P_{1,1}^{(i,j)} & P_{1,2}^{(i,j)} & P_{1,3}^{(i,j)} \\ P_{2,1}^{(i,j)} & P_{2,2}^{(i,j)} & P_{2,3}^{(i,j)} \\ P_{3,1}^{(i,j)} & P_{3,2}^{(i,j)} & P_{3,3}^{(i,j)} \end{bmatrix} \tag{6}$$

with $P_{n,m}^{(i,j)}$ for $n = 1,2,3$ and $m = 1,2,3$ given in Table 1.

3.2 Analytic Expression for $\underline{b}(\underline{\eta}, \gamma)$

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

$$\underline{b}(\underline{\eta}, \gamma) = \begin{bmatrix} \underline{b}_1^T(\underline{\eta}, \gamma) & \underline{b}_2^T(\underline{\eta}, \gamma) & \cdots \underline{b}_L^T(\underline{\eta}, \gamma) \end{bmatrix}^T \;,$$

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

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

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

$$\mathbf{C}_i = \begin{bmatrix} \frac{\hat{y}}{w_i^e} & 0 & 0 & -\frac{2\,\hat{y}}{w_i^e} & 0 & 0 & \frac{1}{w_i^e} & 0 & 0 \\ -\frac{m_i^e\,\hat{y}^2}{(\sigma_i^e)^2} & \frac{\hat{y}^2}{(\sigma_i^e)^2} & 0 & \frac{2\,m\,\hat{y}}{(\sigma_i^e)^2} & -\frac{2\,\hat{y}}{(\sigma_i^e)^2} & 0 & -\frac{m}{(\sigma_i^e)^2} & \frac{1}{(\sigma_i^e)^2} & 0 \\ \frac{(m_i^e)^2\,\hat{y}^2}{(\sigma_i^e)^3} - \frac{\hat{y}^2}{\sigma_i^e} & -\frac{2\,m_i^e\,\hat{y}^2}{(\sigma_i^e)^3} & \frac{\hat{y}^2}{(\sigma_i^e)^3} & -\frac{2\,(m_i^e)^2\,\hat{y}}{(\sigma_i^e)^3} + \frac{2\,\hat{y}}{\sigma_i^e} & \frac{24\,m_i^e\,\hat{y}}{(\sigma_i^e)^3} & -\frac{2\,\hat{y}}{(\sigma_i^e)^3} & \frac{(m_i^e)^2}{(\sigma_i^e)^3} - \frac{1}{\sigma_i^e} & -\frac{2\,m_i^e}{(\sigma_i^e)^3} & \frac{1}{(\sigma_i^e)^3} \end{bmatrix}$$

Table 2. An example of the matrix $\mathbf{C}_i$ for $h(x) = x^3$.

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

Using the relations

$$\frac{\partial\, f^e(x,\underline{\eta})}{\partial\, \underline{\eta}_i} = f_i^e\left(x,\underline{\eta}_i\right) \begin{bmatrix} \frac{1}{w_i^e} \\ \frac{x-\hat{x}_i^e}{(\sigma_i^e)^2} \\ \frac{(x-\hat{x}_i^e)^2-(\sigma_i^e)^2}{(\sigma_i^e)^3} \end{bmatrix}$$

and

$$\frac{\partial\, \tilde{f}^L(x,\gamma)}{\partial\, \gamma} = -\frac{(\epsilon+\gamma)\,(y-h(x))^2}{(1+\epsilon)^2(\sigma^v)^2}\,\tilde{f}^L(x,\gamma) \ ,$$

we obtain

$$\underline{b}_i(\underline{\eta},\gamma) = -\int_{\mathbb{R}} f^p(x)\,\frac{(\epsilon+\gamma)\,(y-h(x))^2}{(1+\epsilon)^2(\sigma^v)^2}\,\tilde{f}^L(x,\gamma) \begin{bmatrix} \frac{1}{w_i^e} \\ \frac{x-\hat{x}_i^e}{(\sigma_i^e)^2} \\ \frac{(x-\hat{x}_i^e)^2-(\sigma_i^e)^2}{(\sigma_i^e)^3} \end{bmatrix} f_i^e\left(x,\underline{\eta}_i\right)\, dx \ .$$

For further simplification, we use (3) and replace $\tilde{f}^e(x)$ by its approximation $f^e(x,\underline{\eta})$. The resulting expression $f^e(x,\underline{\eta})\, f_i^e\left(x,\underline{\eta}_i\right)$ is converted to a Gaussian mixture

$$f_i^s(x) = \sum_{j=1}^{L^e} f_{i,j}^s(x) = \sum_{j=1}^{L^e} w_{i,j}^s \mathcal{N}\left(x-\hat{x}_{i,j}^s, \sigma_{i,j}^s\right) \ ,$$

where the coefficients $w_{i,j}^s$, $\hat{x}_{i,j}^s$ and $\sigma_{i,j}^s$ are collected in coefficient matrices $\mathbf{W}$, $\mathbf{M}$, $\mathbf{S}$. In addition, for polynomial measurement nonlinearities $h(x)$ we set

$$-\frac{(\epsilon+\gamma)\,(y-h(x))^2}{(1+\epsilon)^2(\sigma^v)^2} \begin{bmatrix} \frac{1}{w_i^e} \\ \frac{x-\hat{x}_i^e}{(\sigma_i^e)^2} \\ \frac{(x-\hat{x}_i^e)^2-(\sigma_i^e)^2}{(\sigma_i^e)^3} \end{bmatrix} = \mathbf{C}_i \begin{bmatrix} 1 \\ x \\ \vdots \\ x^N \end{bmatrix} \ ,$$

where N is the highest polynomial order occurring on the left hand side. An example of the matrix $\mathbf{C}_i$ for $h(x) = x^3$ is given in Table 2. These simplifications result in

$$\underline{b}_i(\underline{\eta},\gamma) = \mathbf{C}_i \int_{\mathbb{R}} \begin{bmatrix} 1 \\ x \\ \vdots \\ x^N \end{bmatrix} f_i^s(x)\, dx \ .$$

With

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{C}_2 & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_3 & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

we finally obtain

$$\underline{b}(\underline{\eta},\gamma) = \mathbf{C}\,\mathbf{\Theta}_{L^e,N} \begin{bmatrix} \mathbf{E}^{(0)} \\ \mathbf{E}^{(1)} \\ \vdots \\ \mathbf{E}^{(N)} \end{bmatrix} \underline{1}_{L^e}$$

where $\underline{1}_{L^e}$ is a vector of length L^e with all components equal to one. $\mathbf{\Theta}_{L^e,N}$ is a permutation matrix given in the appendix. $\mathbf{E}^{(k)}$ for $k = 1,\ldots,N$ are moment matrices according to

$$\mathbf{e}_{i,j}^{(k)} = \int_{\mathbb{R}} x^k f_{i,j}^s(x) dx \ .$$

In the case of Gaussian densities $f_{i,j}^s(x)$, higher–order moments can be expressed in terms of $w_{i,j}^s$, $\hat{x}_{i,j}^s$ and $\sigma_{i,j}^s$ [9]. Expressions that directly relate the moment matrices $\mathbf{E}^{(k)}$ to the coefficient matrices $\mathbf{W}$, $\mathbf{M}$, $\mathbf{S}$ are given in Table 3.

4 Structural Adaptation

While performing the progression, the measure of deviation between the true density and its approximation is continuously checked. For that purpose, a *normalized* distance measure

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

is used, which 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

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

for $i = 1,\ldots,L^e$. The most critical component is then replaced by a Gaussian mixture with smaller individual variances. For that purpose, a precalculated splitting library is employed. Numerical values of a typical splitting library for

$$\mathbf{E}^{(0)} = \mathbf{W}$$

$$\mathbf{E}^{(1)} = \mathbf{W} \odot \mathbf{M}$$

$$\mathbf{E}^{(2)} = \mathbf{W} \odot (\mathbf{M}^{\odot 2} + \mathbf{S}^{\odot 2})$$

$$\mathbf{E}^{(3)} = \mathbf{W} \odot \mathbf{M} \odot (\mathbf{M}^{\odot 2} + 3\,\mathbf{S}^{\odot 2})$$

$$\mathbf{E}^{(4)} = \mathbf{W} \odot (\mathbf{M}^{\odot 4} + 6\,\mathbf{M}^{\odot 2} \odot \mathbf{S}^{\odot 2} + 3\,\mathbf{S}^{\odot 4})$$

$$\mathbf{E}^{(5)} = \mathbf{W} \odot \mathbf{M} \odot (\mathbf{M}^{\odot 4} + 10\,\mathbf{M}^{\odot 2} \odot \mathbf{S}^{\odot 2} + 15\,\mathbf{S}^{\odot 4})$$

$$\mathbf{E}^{(6)} = \mathbf{W} \odot (\mathbf{M}^{\odot 6} + 15\,\mathbf{M}^{\odot 4} \odot \mathbf{S}^{\odot 2} + 45\,\mathbf{M}^{\odot 2} \odot \mathbf{M}^{\odot 4} + 15\,\mathbf{S}^{\odot 6})$$

$$\mathbf{E}^{(7)} = \mathbf{W} \odot \mathbf{M} \odot (\mathbf{M}^{\odot 6} + 21 \odot \mathbf{M}^{\odot 4} \odot \mathbf{S}^{\odot 2} + 105\,\mathbf{M}^{\odot 2} \odot s^{\odot 4} + 105\,s^{\odot 6})$$

$$\mathbf{E}^{(8)} = \mathbf{W} \odot (\mathbf{M}^{\odot 8} + 28 \odot \mathbf{M}^{\odot 6} \odot \mathbf{S}^{\odot 2} + 210\,\mathbf{M}^{\odot 4} \odot \mathbf{S}^{\odot 4} + 420\,\mathbf{M}^{\odot 2} \odot \mathbf{S}^{\odot 6} + 105 \odot \mathbf{S}^{\odot 8})$$

Table 3. Moment matrices $\mathbf{E}^{(0)}$ to $\mathbf{E}^{(8)}$ where $\mathbf{A} \odot \mathbf{B}$ denotes the element–by–element multiplication of two matrices $\mathbf{A}$ and $\mathbf{B}$ with identical dimensions. $\mathbf{A}^{\odot k}$ denotes the element–wise power.

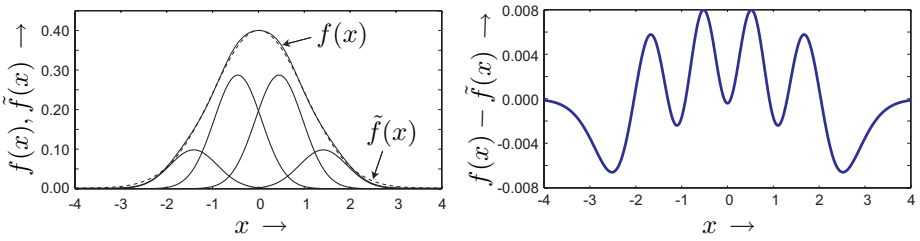

Figure 1. Approximating the standard Gaussian density by means of a Gaussian mixture with $L = 4$ mixture components. Left: True density $\tilde{f}(x)$, its Gaussian mixture approximation $f(x)$, and the individual mixture components $f_i(x)$, $i = 1, \ldots, 4$. Right: The deviation $f(x) - \tilde{f}(x)$ between true and approximate density.

i	w_i	m_i	σ_i
1	0.35690452	−1.41312052	0.51751260
2	0.61042539	−0.44973060	0.51751260
3	0.61042539	0.44973060	0.51751260
4	0.35690453	1.41312052	0.51751260

Table 4. Parameters of a Gaussian mixture with 4 mixture components for approximating the standard Gaussian density.

replacing a standard Gaussian density by a Gaussian mixture with four components of equal variance is given in Table 4 and visualized in Figure 1. The results are applied to splitting arbitrary Gaussian densities by means of suitable shifting and scaling.

Before restarting the progression again after modifying the number of mixture components, an additional correction step is required. This is due to the fact that the structural adaptation of the parameter vector is performed for a fixed value of the progression parameter γ and cannot be compensated by continuing the progression. The appropriate correction step will not be derived here for the sake of brevity.

A block diagram of the progressive Bayesian estimator including the parametric adaptation discussed in Section 3 and the structural modification discussed in this section is shown in Figure 2.

5 Discussion and Future Work

A new type of estimator for performing the measurement update in the case of nonlinear dynamic systems with continuous–valued states has been introduced, which is based on a general framework for progressive Bayesian estimators given in [6].

The first contribution is concerned with obtaining optimal parameters of an 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

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

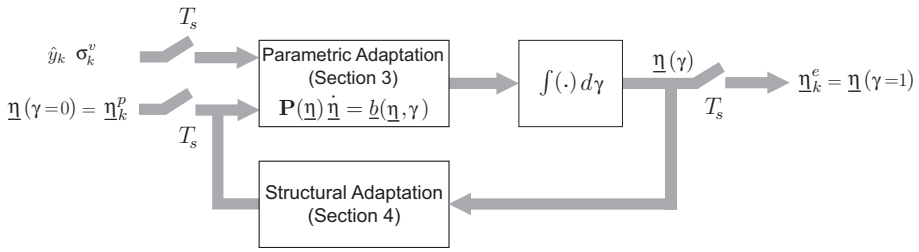

Figure 2. Block diagram of the progressive Bayesian estimator including the parametric adaptation discussed in Section 3 and the structural modification discussed in Section 4.

differential equations. The desired optimal density parameters are then calculated by solving the differential equations over a finite time interval. The second contribution is concerned with adapting the number of components of the Gaussian mixture density approximation. For that purpose, a novel approach for mixture density adaptation based on splitting libraries and on the efficient identification of critical mixture components has been proposed.

For the special case of polynomial measurement nonlinearities, closed–form analytic expression have been derived for the coefficients of the system of explicit ordinary first–order differential equations. In the case of arbitrary measurement nonlinearities [6], some of the coefficients must be calculated numerically. For that purpose, an alternative to numerical integration is given in [11].

The resulting estimators fill the gap between simple estimators based on, e.g. linearization, and complex numerical approaches like particle filters [2] or grid–based estimators [3]. Since the estimator performance is adjusted by specifying the maximum tolerable deviation between the true density and its approximation, the designer can trade accuracy for computational power required. As a result, economic estimators can be designed that are adequate for the given application.

In the context of this paper, white Gaussian measurement noise was assumed. The derived estimator also handles measurement noise characterized by a Gaussian mixture density. Future work will be concerned with colored measurement noise based on the results for estimation in the presence of uncertain correlations derived in [4, 5].

For the sake of simplifying the corresponding derivations, the paper is limited to the case of scalar states. However, a generalization to vector–valued states and measurements is available.

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] U. D. Hanebeck and K. Briechle. New Results for Stochastic Prediction and Filtering with Unknown Correlations. In *Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI'2001), Baden–Baden*, pages 147–152, 2001.

[5] U. D. Hanebeck, K. Briechle, and J. Horn. A Tight Bound for the Joint Covariance of Two Random Vectors with Unknown but Constrained Cross–Correlation. In *Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI'2001), Baden–Baden*, pages 85–90, 2001.

[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] C. Musso, N. Oudjane, and F. LeGland. Improving Regularized Particle Filters. Technical Report CUED/F-INFENG/TR 380, Cambridge University Engineering Department, 2000.

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

[9] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw–Hill, 1984.

[10] 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.

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

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

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

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

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

Appendix

The permutation matrix $\mathbf{\Theta}_{nm}$ is implicitly defined according to

$$\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_n \end{bmatrix} = \mathbf{\Theta}_{nm} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_m \end{bmatrix}$$

with $\underline{x}_i \in \mathbb{R}^n$ for $i = 1, 2, \ldots, m$, $\underline{y}_j \in \mathbb{R}^m$ for $j = 1, 2, \ldots, n$, and

$$\begin{bmatrix} \underline{y}_1 & \underline{y}_2 & \cdots & \underline{y}_n \end{bmatrix} = \begin{bmatrix} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_n \end{bmatrix}^T \quad .$$