USING THE BHATTACHARYYA DISTANCE IN FUNCTIONAL SAMPLING DENSITY OF PARTICLE FILTER

Ondřej Straka and Miroslav Šimandl

Department of Cybernetics and
Research Centre: Data - Algorithms - Decision
University of West Bohemia in Pilsen
Univerzitní 8, 306 14 Plzeň, Czech Republic
straka30@kky.zcu.cz, simandl@kky.zcu.cz

Abstract: The particle filter for nonlinear state estimation of discrete time dynamic stochastic systems is treated. The functional sampling density of the particle filter strongly affecting estimate quality is studied. The density is given by weighted mixture of the transition probability density functions. The weights are calculated using distance of two reference variable probability density functions representing prior and measurement information. The aim is to find a suitable distance that does not suffer from problems with its numerical computation and that can be computed for a large set of systems analytically. It seems that the Bhattacharyya distance is feasible for evaluation of such a distance. Quality and computational demands of the functional particle filter with primary weights computed using the Bhattacharyya distance are illustrated in a numerical example.

Keywords: Nonlinear systems, state estimation, particle filtering, Monte Carlo method, probability density function, Bhattacharyya distance

1. INTRODUCTION

Recursive state estimation of discrete-time nonlinear stochastic dynamic systems from noisy measurement data has been a subject of considerable research interest for the last three decades.The Bayesian approach can be used for general solution of the state estimation problem. The closed form solution of the Bayesian recursive relations (BRR) is available for a few special cases only so some approximative solutions have to be applied.

Since nineties, Monte Carlo (MC) simulation-based methods have been dominating in nonlinear estimation due to their easy implementation in very general settings and cheap and formidable computational power. The particle filter (PF) belongs to the MC simulation-based methods which provide a convenient approach to computing the posterior probability density function (pdf) of the state. The fundamental paper dealing with MC solution of the BRR was published by Gordon *et al.* (1993) where the first effective filtering method in MC framework was proposed. Contribution of the paper was introducing the resampling step to the sequential importance sampling method to eliminate a convergence problem.

Estimate quality of the PF is strongly affected by sample size and sampling density (SD) which represent key design parameters of the PF.

Effective sample size setting has been disregarded for a long time although sample size represents a key parameter of the PF design, Some advances in effective sample size setting were done in Šimandl and Straka (2002) where the Cramér Rao bound was used as a gauge for quality evaluation of the PF and in Fox (2001), Koller and Fratkina (1998), Straka and Šimandl (2004) where some sample size adaptation techniques have been proposed.

SD design can proceed from two approaches. The first one uses the transition and the measurement pdf's directly (e.g. the prior SD, the optimal SD) (Liu *et al.*, 2001). The second approach uses approximative filtering pdf computed by another filtering method (e.g. the sigma point filter, the extended Kalman filter or the Gaussian sum filter) as the SD and results in so called hybrid particle filters (van der Merwe and Wan, 2003).

This paper deals with the first approach to SD design only. An important contribution to this approach was published by Pitt and Shephard (2001) where the concept of prior rating of samples using primary weights was introduced. Šimandl and Straka (2003) proposed the PF with the functional sampling density (FSD) based on comparison of the prior and the measurement information in the form of reference variable pdf's. To compare the two pdf's, the Kullback J-divergence was chosen. The distance can be computed analytically for a few special cases only and it may have problems during its numerical computation.

Goal of the paper is to analyze the functional sampling density, its aspects and to find a suitable distance of the reference variable pdf's that can be computed analytically for large set of cases and that does not suffer from problems during its numerical computation in other cases.

The paper is organized as follows: State estimation using the PF is introduced in Section 2. A brief survey of SD design and the FSD design is described in Section 3. Section 4 deals with detailed discussion about the FSD together with aspects of choice of reference variable pdf's distance. Section 5 contains application of the Bhattacharyya distance in the FSD design. Further, a numerical illustration of the PF with the Bhattacharyya distance in the FSD is provided in Section 6 and finally, the main results are summarized in Section 7.

2. STATE ESTIMATION BY THE PARTICLE FILTER

Consider the discrete time nonlinear stochastic system given by the transition equation (1) and the measurement equation (2):

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{w}_k), \quad k = 0, 1, 2, \dots \tag{1}$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \quad k = 0, 1, 2, \dots \tag{2}$$

where the vectors $\mathbf{x}_k \in \mathbb{R}^n$, $\mathbf{z}_k \in \mathbb{R}^m$ represent the state of the system and the measurement at time k, respectively, $\mathbf{w}_k \in \mathbb{R}^n$ and $\mathbf{v}_k \in \mathbb{R}^m$ are state and measurement white noises, mutually independent and independent on $\mathbf{x}_0$, with known pdf's $p(\mathbf{w}_k)$ and $p(\mathbf{v}_k)$ respectively, $\mathbf{f}_k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, $\mathbf{h}_k : \mathbb{R}^n \to \mathbb{R}^m$ are known vector functions and the pdf $p(\mathbf{x}_0)$ of the initial state $\mathbf{x}_0$ is known.

The general solution of the state estimation problem is provided by the BRR which produce the filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$ and the predictive pdf $p(\mathbf{x}_{k+1}|\mathbf{z}^k)$, where $\mathbf{z}^k = [\mathbf{z}_0^{\mathrm{T}}, \mathbf{z}_1^{\mathrm{T}}, \dots, \mathbf{z}_k^{\mathrm{T}}]^{\mathrm{T}}$.

The idea of the PF in nonlinear state estimation is to approximate the filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$, $k = 0, 1, 2 \dots$, by the empirical filtering pdf $r_N(\mathbf{x}_k|\mathbf{z}^k)$ which is given by N random samples of the state $\{\mathbf{x}_k^{(i)}, \quad i = 1, \dots, N\}$ and corresponding weights $\{w_k^{(i)}, \quad i = 1, \dots, N\}$. General algorithm of the PF (Liu *et al.*, 2001) can be summarized using the following steps:

Initialization: The samples $\{\mathbf{x}_0^{(i)}\}$, i=1,2,...,N are generated from the prior pdf $p(\mathbf{x}_0|\mathbf{z}^{-1})$. Then the weights $\{\mathfrak{w}_0^{(i)}\}$ are associated to the samples $\{\mathbf{x}_0^{(i)}\}$,

$$\mathfrak{w}_0^{(i)} \propto p(\mathbf{z}_0|\mathbf{x}_0^i), \quad \text{i=1,2,\dots,N}, \tag{3}$$

where $\sum_{i=1}^N \mathfrak{w}_0^{(i)} = 1$. The empirical pdf $r_N(\mathbf{x}_0|\mathbf{z}^0)$ given as

$$r_N(\mathbf{x}_0|\mathbf{z}^0) = \sum_{i=1}^N \mathfrak{w}_0^{(i)} \delta(\mathbf{x}_0 - \mathbf{x}_0^{(i)})$$

approximates the filtering pdf $p(\mathbf{x}_0|\mathbf{z}^0)$. The function $\delta(\cdot)$ is the Dirac function defined as $\delta(\mathbf{x}) = 0$ for $\mathbf{x} \neq 0$ and $\int \delta(\mathbf{x}) \mathrm{d}\mathbf{x} = 1$. Let the time step k be $k = 1$.

Resampling: Resampling serves for rejuvenating the samples $\{\mathbf{x}_{k-1}^{(i)}\}$ according to the weights $\{\mathfrak{w}_{k-1}^{(i)}\}$.

Filtering: The samples $\{\mathbf{x}_k^{(i)}\}$, i=1,2,...,N for the next time instant k are generated from the global sampling pdf $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)}, \mathbf{z}_k)$ where

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)}, \mathbf{z}_k) = \sum_{i=1}^N \mathfrak{v}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k) \pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k). \tag{4}$$

To generate the samples $\mathbf{x}_k^{(i)}$, i=1,2,...,N, firstly N indices j_i, i=1,2,...,N have to be drawn from the multinomial distribution with parameters given by the primary weights $\{\mathfrak{v}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k),\ i{=}1, \dots, N\}$. Then each sample $\mathbf{x}_k^{(i)}$ is generated from the local sampling pdf $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(j_i)}, \mathbf{z}_k)$. The weights $\{\mathfrak{w}_k^{(i)},\ i{=}1, \dots, N\}$ associated to the samples $\{\mathbf{x}_k^{(i)}$, i=1,2,...,N$\}$ are calculated using the following form

$$\mathfrak{w}_k^{(i)} \propto \frac{p(\mathbf{z}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(j_i)})}{\mathfrak{v}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(j_i)}, \mathbf{z}_k)} \mathfrak{w}_{k-1}^{(j_i)}. \quad (5)$$

The empirical pdf $r_N(\mathbf{x}_k|\mathbf{z}^k)$ given by the samples $\{\mathbf{x}_k^{(i)}\}$ and the weights $\{\mathfrak{w}_k^{(i)}\}$ as

$$r_N(\mathbf{x}_k|\mathbf{z}^k) = \sum_{i=1}^{N} \mathfrak{w}_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)})$$

approximates the filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$.

Let $k \leftarrow k+1$ and continue with **Resampling**.

3. FUNCTIONAL SAMPLING DENSITY

The global SD $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)}, \mathbf{z}_k)$ is one of two crucial design parameters directly affecting quality of the PF estimate.

One of the first particle filters proposed in Gordon *et al.* (1993) as the bootstrap filter (BF) uses the global SD in the following form:

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)}, \mathbf{z}_k) = \sum_{i=1}^{N} \frac{1}{N} p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}). \quad (6)$$

It can be seen that the pdf (6) does not take into account knowledge of the measurement $\mathbf{z}_k$. The local SD has the form of mixture of the transition pdf's with equal primary weights $\mathfrak{v}(\mathbf{x}_{k-1}^{(j_i)}, \mathbf{z}_k)$, $\mathfrak{v}(\mathbf{x}_{k-1}^{(j_i)}, \mathbf{z}_k) = \frac{1}{N}$. Thus the global SD (6) can be called *the prior sampling density.*

In some cases it is possible to find an explicit form of the pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(j_i)}, \mathbf{z}_k)$ and then the global SD can be proposed in the form

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)}, \mathbf{z}_k) = \sum_{i=1}^{N} \frac{1}{N} p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k). \quad (7)$$

The pdf (7) does not increase variance of weights $\mathsf{E}\{\mathfrak{w}_k^2\}$ which is closely related to estimate quality and thus the pdf (7) can be called *the optimal sampling density.* Drawback of the optimal SD lies in a relatively small set of systems for which such a pdf can be found explicitly (e.g. systems with gaussian transition pdf, linear measurement equation with additive gaussian measurement noise).

Pitt and Shephard (2001) proposed the auxiliary particle filter (APF) and introduced concept of primary weights $\mathfrak{v}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ into SD which has the following form

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)}, \mathbf{z}_k) = \sum_{i=1}^{N} \mathfrak{v}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k) p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}). \quad (8)$$

Each primary weight $\mathfrak{v}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ evaluates corresponding sample $\mathbf{x}_{k-1}^{(i)}$ according to the current measurement $\mathbf{z}_k$. They proposed the primary weight in the form

$$\tilde{\mathfrak{v}}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k) = p_{\mathbf{z}_k|\mathbf{x}_k}(\mathbf{z}_k|\mu_k^{(i)}) \quad (9)$$

where the variable $\mu_k^{(i)}$ represents mean, mode or another likely value of the random variable $\mathbf{x}_k$ given by the transition pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)})$. Note that the primary weight $\mathfrak{v}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ is a normalized version of $\tilde{\mathfrak{v}}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ so that $\mathfrak{v}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k) = \tilde{\mathfrak{v}}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)/[\sum_{j=1}^{N} \tilde{\mathfrak{v}}(\mathbf{x}_{k-1}^{(j)}, \mathbf{z}_k)]$.

It was shown in Pitt and Shephard (2001) that the variance of weights $\mathsf{E}\{\mathfrak{w}_k^2\}$ of the APF with the primary weights (9) is usually lower than that of the PF with the prior SD.

The form (9) of primary weight can be seen as comparison of prior and measurement information. While the measurement information is given by the measurement pdf $p(\mathbf{z}_k|\mathbf{x}_k)$, the prior information is given by the point estimate $\mu_k^{(i)}$ only. To compare full information given by the measurement and the transition pdf's, the PF with the FSD has been introduced in Šimandl and Straka (2003).

The FSD has the form (8). Evaluation of the primary weight $\mathfrak{v}(\mathbf{x}_{k-1}^{(i)}|\mathbf{z}_k)$ gets out of comparison of the measurement information given by $p(\mathbf{z}_k|\mathbf{x}_k)$ and the prior information given by $p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)})$. Thus the primary weight takes into account complete available information.

To compare the pdf's it is necessary to introduce a reference variable $\mathbf{y}_k$ defined as

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k), \quad (10)$$

The comparison procedure can be achieved through the pdf $p(\mathbf{y}_k|\mathbf{z}_k)$ and the pdf $p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)})$. The pdf $p(\mathbf{y}_k|\mathbf{z}_k)$ can be obtained from the measurement equation (2) and has the form $p(\mathbf{y}_k|\mathbf{z}_k) = p_{\mathbf{v}_k}(\mathbf{z}_k - \mathbf{y}_k)$. The pdf $p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)})$ is a pdf of the random variable $\mathbf{y}_k$ obtained by the transformation $\mathbf{h}_k : \mathbb{R}^n \to \mathbb{R}^m$ of the random variable $\mathbf{x}_k$ given by the pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)})$. The pdf $p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)})$ itself can be found using the standard rule of random variable transformation or it is possible to use a completely different approach, e.g. unscented transformation (Julier *et al.*, 2000).

Different measures can be applied for comparison of the pdf's. The Kullback J-divergence (KJD) (Kullback and Leibler, 1951) defined as

$$J(p(\mathbf{y}_k|\mathbf{z}_k)\|p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)})) \triangleq \int [p(\mathbf{y}_k|\mathbf{z}_k) - p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)})] [\log p(\mathbf{y}_k|\mathbf{z}_k) - \log p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)})] \mathrm{d}\mathbf{y}_k, \quad (11)$$

which can be seen as a symmetrized version of the Kullback-Leibler distance, can be applied here as a standard tool.

The primary weight $\mathfrak{v}(\mathbf{z}_k|\mathbf{x}_{k-1}^{(i)})$ should be high for the samples $\mathbf{x}_{k-1}^{(i)}$ with the pdf $p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)})$ close to the pdf $p(\mathbf{y}_k|\mathbf{z}_k)$, i.e. with low value of the considered measure. In this case a simple term

$e^{-J(\cdot\|\cdot)}$ may be used. So the primary weights may have the form

$$\tilde{\mathfrak{v}}(\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k) = e^{-J(p(\mathbf{y}_k|\mathbf{z}_k)\|p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)}))}. \quad (12)$$

Detailed discussion concerning choice of distance for the FSD is subject of the next section.

4. SOME ASPECTS OF FUNCTIONAL SAMPLING DENSITY

Aim of this section is to analyze impact of choice of the reference variable pdf's distance on the FSD. Consider both the pdf's of the reference variable $p(\mathbf{y}_k|\mathbf{z}_k)$ and $p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)})$ are known. The next step of the FSD design consists of their comparison using a distance. Šimandl and Straka (2003) proposed the KJD measure $J(p(\mathbf{y}_k|\mathbf{z}_k)\|p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)}))$ for such comparison and to evaluate the primary weights using (12). Evaluation of the KJD measure has to be calculated numerically except a few special cases. (e.g. both pdf's $p(\mathbf{y}_k|\mathbf{z}_k)$ and $p(\mathbf{y}_k|\mathbf{x}_{k-1}^{(i)})$ are Gaussian). Even for both pdf's given by Gaussian mixtures it is necessary to evaluate the KJD measure numerically. Because numerical integration of the integral in (11) is time consuming especially for high dimensions of the reference variable $\mathbf{y}_k$, it would be suitable to find a distance that can be computed analytically for a larger set of systems.

Also numerical computation of such distance has to be object of attention. To ease notation consider $p_{x^{(i)}}(\mathbf{y}_k) = p(\mathbf{y}_k|\mathbf{x}_k^{(i)})$ and $p_z(\mathbf{y}_k) = p(\mathbf{y}_k|\mathbf{z}_k)$. Consider for example distance called Inaccuracy (Kerridge, 1961) defined as

$$I(p_{x^{(i)}}(\mathbf{y}_k)\|p_z(\mathbf{y}_k) \triangleq \int p_{x^{(i)}}(\mathbf{y}_k) \log \frac{1}{p_z(\mathbf{y}_k)} \mathrm{d}\mathbf{y}_k$$
$$= -\int p_{x^{(i)}}(\mathbf{y}_k) \log p_z(\mathbf{y}_k) \mathrm{d}\mathbf{y}_k. \quad (13)$$

The integral (13) can be computed for $p_{x^{(i)}}(\mathbf{y}_k) = \sum_{i=1}^{L} \alpha_i \mathcal{N}\{\mathbf{y}_k; \mathbf{m}_{k,i}, \mathbf{S}_{k,i}\}$ and for Gaussian $p_z(\mathbf{y}_k)$, $p_z(\mathbf{y}_k) = \mathcal{N}\{\mathbf{y}_k; \mu_k, \mathbf{P}_k\}$. To evaluate the integral (13) numerically it is necessary to resolve problem of limited numerical accuracy of computational environment.

Application of some correction step for resolving limited numerical accuracy together with normalization of the primary weights may result in quite different values of the primary weights given by numerically evaluated Inaccuracy in comparison to the primary weights given by analytically evaluated Inaccuracy.

To illustrate incorrect primary weights evaluation given by numerical computation of the integral (13) caused by limited accuracy of computational environment, consider for example
$p_{x^{(i)}}(\mathbf{y}_k) = 0.5\mathcal{N}\{\mathbf{y}_k; -3, 0.01\} + 0.5\mathcal{N}\{\mathbf{y}_k; 3, 0.01\}$ and $p_z(\mathbf{y}_k) = \mathcal{N}\{\mathbf{y}_k; \mu, 0.01\}$. Figure 1 contains the values of integral (13) computed both analytically and numerically for different values of μ. It can be seen that shape of the curves are quite different and thus the numerical solution of (13) can not be used here.

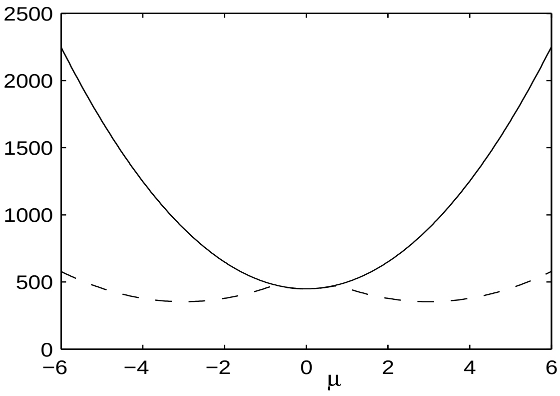

Fig. 1. Values of Inaccuracy for its analytical computation (solid) and numerical computation (dashed)

Moreover, it is important to note that application of Inaccuracy is inappropriate for the considered pdf's because it does not respect multimodality of the pdf's (see solid curve in Figure 1). Thus a deeper analysis of the distance for given reference variable pdf's is necessary.

5. USING BHATTACHARYYA DISTANCE

This section contains proposal of a reference variable pdf's distance for the FSD that can be computed for a large set of systems analytically and that is also suitable for numerical computation in other cases. A very simple relation measuring distance between two pdf's is defined as follows

$$\varepsilon_{\text{Bayes}} = \int \min(p_{x^{(i)}}(\mathbf{y}_k), p_z(\mathbf{y}_k)) \mathrm{d}\mathbf{y}_k, \quad (14)$$

where the distance between two pdf's $p_{x^{(i)}}(\mathbf{y}_k)$ and $p_z(\mathbf{y}_k)$ is defined as $D(p_{x^{(i)}}(\mathbf{y}_k)\|p(\mathbf{y}_k|\mathbf{z}_k)) = 1 - \varepsilon_{\text{Bayes}}$. The relation (14) is frequently used as the Bayes error (BE). The value of the BE is close to one for similar pdf's and close to zero for dissimilar pdf's and thus it can be used for primary weights evaluation directly. Numerical computation of the integral in (14) does not require any modification to avoid problems caused by limited numerical accuracy of computational environment. Unfortunately, the integral in (14) can not be evaluated analytically even for two Gaussian pdf's. Nevertheless due to the inequality

$$\min[a, b] \leq a^s b^{1-s}, \ 0 \leq s \leq 1$$

with $s = 0.5$ it is possible to propose an upper bound for (14) in the form

$$\varepsilon_{\mathrm{Bayes}} \leq \int \sqrt{p_{x^{(i)}}(\mathbf{y}_k)\ p_z(\mathbf{y}_k)}\mathrm{d}\mathbf{y}_k. \qquad (15)$$

Further consider both pdf's $p_{x^{(i)}}(\mathbf{y}_k)$ and $p_z(\mathbf{y}_k)$ to be mixtures of some arbitrary pdf's

$$p_{x^{(i)}}(\mathbf{y}_k) = \sum_{i=l}^{A} \alpha_l p_{l,x^{(i)}}(\mathbf{y}_k) \qquad (16)$$

and

$$p_z(\mathbf{y}_k) = \sum_{j=1}^{B} \beta_j p_{j,z}(\mathbf{y}_k) \qquad (17)$$

It holds that

$$\varepsilon_{\mathrm{Bayes}} \leq \int \sqrt{(\sum_{l=1}^{A} \alpha_l p_{l,x^{(i)}}(\mathbf{y}_k))(\sum_{j=1}^{B} \beta_j p_{j,z}(\mathbf{y}_k))\mathrm{d}\mathbf{y}_k}$$

$$= \int \sqrt{\sum_{l=1}^{A}\sum_{j=1}^{B} \alpha_l \beta_j p_{l,x^{(i)}}(\mathbf{y}_k) p_{j,z}(\mathbf{y}_k)\mathrm{d}\mathbf{y}_k}.$$

Applying the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ as proposed in Mak (1995) it holds that

$$\varepsilon_{\mathrm{Bayes}} \leq \sum_{l=1}^{A}\sum_{j=1}^{B} \sqrt{\alpha_l \beta_j} \int \sqrt{p_{l,x^{(i)}}(\mathbf{y}_k) p_{j,z}(\mathbf{y}_k)}\mathrm{d}\mathbf{y}_k. \qquad (18)$$

Thus if it is possible to evaluate the integral in (18) analytically, then it is possible to evaluate an upper bound for the BE. For $p_{l,x^{(i)}}(\mathbf{y}_k) = \mathcal{N}\{\mathbf{y}_k; \mathbf{m}_{k,l}^{(i)}, \mathbf{S}_{k,l}^{(i)}\}$ and $p_{j,z}(\mathbf{y}_k) = \mathcal{N}\{\mathbf{y}_k; \mu_{k,j}, \mathbf{P}_{k,j}\}$ the value of the integral in (18) is

$$\int \sqrt{\mathcal{N}\{\mathbf{y}_k; \mathbf{m}_{k,l}^{(i)}, \mathbf{S}_{k,l}^{(i)}\}\mathcal{N}\{\mathbf{y}_k; \mu_{k,j}, \mathbf{P}_{k,j}\}}\mathrm{d}\mathbf{y}_k =$$

$$= e^{-\mathrm{D}_{\mathrm{Bhat}}(\mathcal{N}\{\mathbf{y}_k; \mathbf{m}_{k,l}^{(i)}, \mathbf{S}_{k,l}^{(i)}\}||\mathcal{N}\{\mathbf{y}_k; \mu_{k,j}, \mathbf{P}_{k,j}\})}, \qquad (19)$$

where $\mathrm{D}_{\mathrm{Bhat}}(p_1(x)||p_2(x))$ is the Bhattacharyya distance between two densities $p_1(x)$ and $p_2(x)$ defined as (Basseville, 1989)

$$\mathrm{D}_{\mathrm{Bhat}}(p_1(x)||p_2(x)) \triangleq -\log \int \sqrt{p_1(x) p_2(x)}\mathrm{d}x. \qquad (20)$$

The distance for the two Gaussian pdf's considered in (19) has the following form

$$\mathrm{D}_{\mathrm{Bhat}}(\mathcal{N}\{\mathbf{y}_k; \mathbf{m}_{k,l}^{(i)}, \mathbf{S}_{k,l}^{(i)}\}||\mathcal{N}\{\mathbf{y}_k; \mu_{k,j}, \mathbf{P}_{k,j}\}) =$$

$$= \frac{1}{8}(\mathbf{m}_{k,l}^{(i)} - \mu_{k,j})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{m}_{k,l}^{(i)} - \mu_{k,j}) +$$

$$+ \frac{1}{2}\log\frac{|\mathbf{R}|}{\sqrt{|\mathbf{S}_{k,l}^{(i)}||\mathbf{P}_{k,j}|}}, \qquad (21)$$

where $\mathbf{R} = \frac{\mathbf{S}_{k,l}^{(i)} + \mathbf{P}_{k,j}}{2}$. Thus the upper bound for the BE can be evaluated analytically even for pdf's given by a mixture of Gaussians. This is important because it greatly extends set of systems for which the primary weights for the FSD can be computed analytically and thus it allows to accelerate computation of the FSD. Further, the distance respects multimodality of the densities which makes it fitting for the FSD in case of any reference variable pdf being multimodal.

6. NUMERICAL EXAMPLES

To show different performance of the BF, the APF and the FPF, the system with multi-modal pdf of state noise is considered. The aim is to compare quality of SD of particular PF with exact filtering pdf produced by the Gaussian sum filter (GSF). To calculate exact filtering pdf using the GSF, a linear non-gaussian system has been chosen in the following form

$$x_k = 0.9\ x_{k-1} + w_k, \qquad (22)$$

$$z_k = x_k + v_k, \qquad (23)$$

with $p(v_k) = \mathcal{N}(v_k; 0, 0.01)$, $p(x_0) = \mathcal{N}(v_k; 0, 0.001)$, $p(w_k) = 0.1\,\mathcal{N}(w_k; -1, 0.001) + 0.9\,\mathcal{N}(w_k; 1, 0.001)$. The system was simulated for $k = 1, 2, \ldots 8$. Four particle filters were chosen for quality comparison: the BF, the APF with μ_k chosen as mean, labeled as the APFM, the APF with μ_k chosen as a sample, labeled as the APFS and the FPF with primary weights evaluated using the upper bound (19). The sample size $N = 100$ was common for all the particle filters.

The exact filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$, in the form $p(\mathbf{x}_k|\mathbf{z}^k) = \sum_{j=1}^{M} \gamma_j p_j(\mathbf{x}_k|\mathbf{z}^k)$ was calculated using the GSF. The comparison was accomplished using the Bhattacharyya distance between the SD of particular PF $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)}, \mathbf{z}_k)$ and the exact filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$ as

$$\mathrm{D}_{\mathrm{Bhat}}(p(\mathbf{x}_k|\mathbf{z}^k)||\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)}, \mathbf{z}_k)) = -\log(\sum_{j=1}^{M}\sum_{i=1}^{N}$$

$$\sqrt{\gamma_j \mathfrak{v}(\mathbf{x}_k^{(i)}, \mathbf{z}_k)} e^{-\mathrm{D}_{\mathrm{Bhat}}(p_j(\mathbf{x}_k|\mathbf{z}^k)||p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)})})$$

Exact value of the Bhattacharyya distance was approximated using the MC method with $S = 10000$ realizations $\{\mathbf{z}_k(j)\}_{k=1}^{8}$, $j = 1, 2, \ldots S$ and thus a criterion J_k was considered in the following form:

$$J_k = \sum_{s=1}^{S} \frac{\mathrm{D}_{\mathrm{Bhat}}(p(\mathbf{x}_k|\mathbf{z}^k(s))||\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)}, \mathbf{z}_k(s)))}{S},$$

where $\mathbf{z}_k(s)$ represents measurement for the s^{th} realization.

Figure 2 contains comparison of time evolution of J_k for the BF, the APFM, the APFS and the FPF. Due to multimodality of the state noise $\mathbf{w}_k$ it can be seen that estimate quality of the APFM is worse than that of the BF which does not employ primary weights. The FPF provides highest estimate quality among all the particle filters.

Comparison of computational demands of a time step between the FPF with the Bhattacharyya distance (FPFB) and the FPF with numerically calculated J-divergence (FPFJ) can be seen in Table 1.

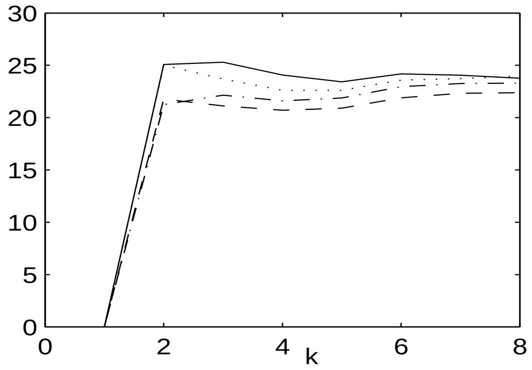

Fig. 2. Time evolution of J_k for the BF (dash-dot), the APFM (solid), the APFS (dotted) and the FPF (dashed)

Table 1. Comp. demands of FPFB and FPFJ

PF algorithm	time of a step [s]
FPFB	0.2
FPFJ	34.9

Note that $N = 100$ samples and one-dimensional reference variable were considered for the comparison and that the difference would be even higher for n-dimensional reference variable. Also note that the FPFB approximately as computationally demanding as the APF.

7. CONCLUSION

The paper dealt with functional sampling density design for the PF. Distance measures between pdf's of the reference variable were analyzed. The measure based on the Bhattacharyya distance for comparison of the reference variable pdf's was proposed. It is more suitable for numerical computation than the Kullback J-divergence proposed earlier and also it can be computed analytically even for reference variable given by mixture of densities. Moreover the proposed measure can respect multimodality of the reference variable pdf's. Estimate quality of the FPF with proposed measure was illustrated in a numerical example.

8. ACKNOWLEDGMENT

The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic, projects No. MSM 235200004 and 1M6798555601 and by the Grant Agency of the Czech Republic, project GACR 102/05/2075.

REFERENCES

Basseville, M. (1989). Distance measures for signal processing and pattern recognition. *European Journal Signal Processing* **18**(4), 349–369.

Fox, D. (2001). KLD-sampling: Adaptive particle filters and mobile robot localization. In: *Advances in Neural Information Processing Systems (NIPS)*.

Gordon, N., D. Salmond and A.F.M. Smith (1993). Novel approach to nonlinear/ non-Gaussian Bayesian state estimation. *IEE proceedings-F* **140**, 107–113.

Julier, S.J., J.K. Uhlmann and H.F. Durrant-white (2000). A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control* **45**(3), 477–482.

Kerridge, D.F. (1961). Inaccuracy and inference. *J. Royal Statist. Society* (23), 184–194.

Koller, D. and R. Fratkina (1998). Using learning for approximation in stochastic processes. In: *Proc. 15th International Conf. on Machine Learning*. Morgan Kaufmann, San Francisco, CA. pp. 287–295.

Kullback, S. and R.A. Leibler (1951). On Information and Sufficiency. *Annals of Math. Statist.* (22), 79–86.

Liu, J.S., R. Cheng and T. Logvinenko (2001). *Sequential Monte Carlo Methods in Practise*. Chap. A Theoretical Framework for Sequential Importance Sampling with Resampling. Statistics for Engineering and Information Science. Springer.

Mak, B. (1995). A distance measure of speech phones and its application to phonetic context clustering. Technical report. OGI School of science and engineering. Research Proficiency Examination paper.

Pitt, M.K. and Shephard, N., Eds.) (2001). *Monte Carlo Methods in Practise*. Chap. Auxiliary Variable Based Particle Filters. Springer. (Ed. A. Doucet, N. de Freitas and N. Gordon).

Šimandl, M. and O. Straka (2002). Nonlinear estimation by particle filters and Cramér-Rao bound. In: *Proceedings of the 15th Triennial World Congress of the IFAC*. Barcelona.

Šimandl, M. and O. Straka (2003). Sampling density design for particle filters. In: *Proceedings of the 13th IFAC Symposium on System Identification*. Rotterdam.

Straka, O. and M. Šimandl (2004). Sample size adaptation for particle filters. In: *preprints of the 16th IFAC Symposium on Automatic Control in Aerospace* (Alexander Nebylov, Ed.). Vol. 1. Saint Petersburg, Russia. pp. 444–449.

van der Merwe, R. and E. A. Wan (2003). Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. Hong Kong.