Proceedings of the 2007 American Control Conference
Marriott Marquis Hotel at Times Square
New York City, USA, July 11-13, 2007

Sampling Densities of Particle Filter: A Survey and Comparison

Miroslav Šimandl
Department of Cybernetics
University of West Bohemia in Pilsen
Pilsen, Czech Republic
Email: simandl@kky.zcu.cz

Ondřej Straka
Department of Cybernetics
University of West Bohemia in Pilsen
Pilsen, Czech Republic
Email: straka30@kky.zcu.cz

Abstract— **The paper deals with the particle filter in discrete-time nonlinear non-Gaussian system state estimation. One of the key parameters affecting estimation quality of the particle filter is the sampling density (also called importance function or proposal density). In the literature, there are many sampling density proposals based on various ideas. The goal of the paper is to provide a survey of sampling densities, to classify them and to compare estimation quality of the particle filter with various sampling densities in an illustration example.**

Index Terms— **Nonlinear filtering, state estimation, particle filter, sampling density**

I. INTRODUCTION

Recursive state estimation of discrete-time nonlinear stochastic dynamic systems from noisy measurement data has been a subject of considerable research interest in the last three decades. General solution of the state estimation problem is given by the Bayesian recursive relations (BRR). The closed form solution of the BRR is available for a few special cases only so usually an approximative solution has to be employed.

Since the nineties, the particle filter (PF) has dominated in recursive nonlinear state estimation due to its easy implementation in very general settings and cheap and formidable computational power. The PF solves the BRR using Monte Carlo (MC) methods, particularly using the importance sampling method, and approximates the continuous state space by a cloud of samples (particles).

The fundamental paper dealing with the MC solution of the BRR was published in [1] where the bootstrap filter was proposed. It was the first effective PF which started a sequential MC method revival because the MC method was used in estimation already in the seventies [2], [3], [4]. Many improvements of the bootstrap filter have been proposed since, see for example [5]. The crucial parameters affecting estimation quality are the sample size dealt with in e.g. [6], [7] and the sampling density (SD) (also known as the proposal density or the importance function) used for drawing samples. This paper focuses on the SD only.

Since the introduction of the bootstrap filter which used one of the simplest SD's, i.e. the prior SD, many other SD's have been proposed to increase estimation quality. Design of the SD, was discussed in e.g. [8] where basic types of the SD, the prior SD and the optimal SD, were given. The optimal SD represents the best choice of the SD nonetheless it can be applied in a few special cases only. On the other hand, the prior SD represents the easiest choice however its application can lead to a low quality estimate.

Given the SD proposals variety, the goal of the paper is to provide a survey of SD's to make a classification and to compare many of them in a numerical example with emphasis on both point estimates and filtering probability density function (pdf) estimates.

The paper is organized as follows. Section II contains the state estimation problem specification and a general algorithm of the PF. Sampling densities survey is given in Section III. A numerical example is a subject of Section IV and the paper is concluded in Section V.

II. PROBLEM STATEMENT

This section deals with solution of the state estimation problem using the PF. Consider the discrete time nonlinear stochastic system given by the state equation (1) and the measurement equation (2):

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

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

where the vectors $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and $\mathbf{z}_k \in \mathbb{R}^{n_z}$ represent the state of the system and the measurement at time k, respectively, $\mathbf{e}_k \in \mathbb{R}^{n_x}$ and $\mathbf{v}_k \in \mathbb{R}^{n_z}$ are state and measurement white noises, mutually independent and independent of $\mathbf{x}_0$, with known pdf's $p(\mathbf{e}_k)$ and $p(\mathbf{v}_k)$, respectively, $\mathbf{f}_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$, $\mathbf{h}_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ are known vector functions and the pdf $p(\mathbf{x}_0)$ of the initial state $\mathbf{x}_0$ is known. The system given by (1) and (2) can be alternatively described by the transition pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and the measurement pdf $p(\mathbf{z}_k|\mathbf{x}_k)$. The goal of the state estimation problem is to find the filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$ where $\mathbf{z}^k \triangleq [\mathbf{z}_0^{\mathrm{T}}, \ldots, \mathbf{z}_k^{\mathrm{T}}]^{\mathrm{T}}$. Note that sometimes only a point estimate, often the mean value $\mathsf{E}[\mathbf{x}_k|\mathbf{z}^k]$, is of concern.

General solution of the state estimation problem in the form of the filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$ is provided by the BRR

$$p(\mathbf{x}_k|\mathbf{z}^k) = \frac{p(\mathbf{x}_k|\mathbf{z}^{k-1})p(\mathbf{z}_k|\mathbf{x}_k)}{\int p(\mathbf{x}_k|\mathbf{z}^{k-1})p(\mathbf{z}_k|\mathbf{x}_k)\mathrm{d}\mathbf{x}_k} \tag{3}$$

$$p(\mathbf{x}_k|\mathbf{z}^{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}^{k-1})\mathrm{d}\mathbf{x}_{k-1} \tag{4}$$

which produce the filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$ and the predictive pdf $p(\mathbf{x}_k|\mathbf{z}^{k-1})$.

The idea of the PF in nonlinear state estimation is to approximate the filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$, by the empirical filtering pdf $r_{N_k}(\mathbf{x}_k|\mathbf{z}^k)$ which is given by N_k random samples of the state $\{\mathbf{x}_k^{(i)}\}_{i=1}^{N_k}$ and associated weights $\{\mathbf{w}_k^{(i)}\}_{i=1}^{N_k}$. The general algorithm of the PF [9] can be summarized as follows. Note that further fixed sample size, i.e. $N_k = N,\ k = 0, 1\ldots$, will be used for simplicity.

General algorithm of the particle filter

0. **Initialization:** Let $k = 0$. Generate N samples $\{\mathbf{x}_0^{(i)}\}_{i=1}^{N}$ from a SD $\pi(\mathbf{x}_0|\mathbf{z}^0)$, compute the weights $\{\tilde{\mathbf{w}}_0^{(i)}\}_{i=1}^{N}$ as

$$\tilde{\mathbf{w}}_0^{(i)} = \frac{p(\mathbf{z}_0|\mathbf{x}_0^{(i)})p(\mathbf{x}_0^{(i)})}{\pi(\mathbf{x}_0^{(i)}|\mathbf{z}^0)}, \quad \text{i=1,2,\ldots,N} \tag{5}$$

and normalize, i.e. $\mathbf{w}_0^{(i)} = \tilde{\mathbf{w}}_0^{(i)}/\sum_{j=1}^{N}\tilde{\mathbf{w}}_0^{(j)}$. The empirical pdf $r_N(\mathbf{x}_0|\mathbf{z}^0)$ given as

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

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 $\int\delta(\mathbf{x})\mathrm{d}\mathbf{x} = 1$.

1. **Resampling:** Generate a new set $\{\mathbf{x}_k^{*(i)}\}_{i=1}^{N}$ by resampling with replacement N times from $\{\mathbf{x}_k^{(i)}\}_{i=1}^{N}$ with probability $P(\mathbf{x}_k^{*(i)} = \mathbf{x}_k^{(i)}) = \mathbf{w}_k^{(i)}$ and set $\mathbf{w}_k^{*(i)} = \frac{1}{N}$.

2. **Filtering:** Increase k and generate a new set of samples $\{\mathbf{x}_k^{(i)}\}_{i=1}^{N}$ from a global SD $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(1:N)}, \mathbf{z}_k)$ where

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

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

$$\tilde{\mathbf{w}}_k^{(i)} = \frac{p(\mathbf{z}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{*(j_i)})}{\lambda_k^{(j_i)}\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{*(j_i)}, \mathbf{z}_k)}\mathbf{w}_{k-1}^{*(j_i)}. \tag{7}$$

and normalized, i.e. $\mathbf{w}_k^{(i)} = \tilde{\mathbf{w}}_k^{(i)}/\sum_{j=1}^{N}\tilde{\mathbf{w}}_k^{(j)}$. The empirical pdf $r_N(\mathbf{x}_k|\mathbf{z}^k)$ is given by the samples $\{\mathbf{x}_k^{(i)}\}_{i=1}^{N}$ and the weights $\{\mathbf{w}_k^{(i)}\}_{i=1}^{N}$ as

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

and iterate to **Resampling** step.

Note that in step 2 the algorithm uses a general SD

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(1:N)}, \mathbf{z}_k)$$

based on utilization of the current measurement $\mathbf{z}_k$ and so it is not possible to distinguish between the time update step and the measurement update step. This general SD covers most of SD proposals which survey is given in the next section.

Also note that the resampling step need not be executed at each time instant because it introduces further simulation error into estimates [8]. If resampling does not take place, it holds that $\mathbf{x}_k^{*(i)} = \mathbf{x}_k^{(i)}$ and $\mathbf{w}_k^{*(i)} = \mathbf{w}_k^{(i)}$ for each i.

III. SURVEY OF SAMPLING DENSITY DESIGN TECHNIQUES

The choice of the sampling density in step 2 of the algorithm affects quality of the state estimate significantly. So there has been a permanent effort to find a suitable procedure for sampling density design in the last decade.

As the PF exploits the importance sampling technique [10], its SD must fulfill a criterion to ensure convergency of the estimates for $N \to \infty$. The criterion says that the support of the SD must contain the support of the target pdf (i.e. the filtering pdf in this case) [10].

Further, to achieve the best possible results , the shape of the SD must be as close to the true filtering pdf as possible. Quality of the estimates is related to variance of the weights $\mathbf{w}_k^{(i)}$ conditioned by the last samples and the measurement and thus it is desirable to achieve a minimum variance.

From implementation point of view the SD should be also as simple with respect to the weights evaluation as possible.

Development of SD design techniques proceeds using two main approaches. The former develops the original concepts of the SD design and proposes the prior SD (PSD) enhancements; e.g. in [11] the PF with the SD based on a combination of the PSD and the optimal SD was proposed, or in [12] the auxiliary particle filter (APF) was developed. Thus the approach will be called the *direct* approach. The latter comes out of utilization of another filtering technique producing a filtering pdf which is used as the SD, e.g. [13]. Therefore it will be called the *composite* approach.

A. Direct Approach to Sampling Density Design

The direct approach to the SD design covers the PSD and the optimal SD as historically the first SD's together with various perfections of the PSD proposed later on.

1) Optimal Sampling Density: The global sampling density in the form

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(1:N)}, \mathbf{z}_k) = \sum_{i=1}^{N}\frac{1}{N}p(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)}, \mathbf{z}_k) \tag{8}$$

minimizes variance of the weights [14], [4] and thus it is called the optimal SD. The weights are calculated as

$$\tilde{\mathbf{w}}_k^{(i)} = p(\mathbf{z}_k|\mathbf{x}_{k-1}^{*(i)})\mathbf{w}_{k-1}^{*(i)}. \tag{9}$$

Although the PF with the optimal SD provides high-quality results, utilization of the SD is limited as the pdf

$p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)},\mathbf{z}_k)$ can be found explicitly in a few special cases only and also the pdf $p(\mathbf{z}_k|\mathbf{x}_{k-1}^{(i)})$ is given by the integral $p(\mathbf{z}_k|\mathbf{x}_{k-1}^{(i)}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)})\mathrm{d}\mathbf{x}_k$ which must be solved to compute the weights. Note that a Gaussian system with a linear function $\mathbf{h}_k(\cdot)$ in (2) represents such a case. In other cases a suboptimal SD must be searched.

2) Prior Sampling Density: The PSD is frequently used due to its simplicity and easy weights computation. The global SD is of the following form

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(1:N)},\mathbf{z}_k) = \sum_{i=1}^{N}\frac{1}{N}p(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)}). \tag{10}$$

The local SD is given by the transition pdf which is easily accessible as a part of the system description. The weights computation in (7) is simplified to

$$\tilde{\mathbf{w}}_k^{(i)} = p(\mathbf{z}_k|\mathbf{x}_k^{*(i)})\mathbf{w}_{k-1}^{*(i)}. \tag{11}$$

As it can be seen, the current measurement $\mathbf{z}_k$ is ignored during samples drawing and that is the reason for the SD name. This fact is also a reason for usually low quality estimates. This disregard of the measurement is critical for systems with a peaked measurement pdf when the samples are generated near a tail of the measurement pdf. Such a case may lead to severe samples impoverishment. The PSD was firstly used in [2] and also in the famous bootstrap filter [1].

Note that the algorithm of the PF with the PSD can be divided into two parts, the former providing predictive estimates by sampling from the transient pdf and the latter providing the filtering estimates by evaluating new weights using (11).

3) Fixed Sampling Density: In [15] even a simpler choice of the SD than the PSD was proposed

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(1:N)},\mathbf{z}_k) = \pi(\mathbf{x}_k). \tag{12}$$

It is independent of previous samples and the current measurement. As no information concerning the model is utilized during sampling, the results can be very poor.

4) Auxiliary Sampling Density: To improve estimate quality and preserve simplicity of the PSD, the auxiliary PF based on the auxiliary variable framework [16] considers the following global SD

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)},\mathbf{z}_k) = \sum_{i=1}^{N}\lambda_k^{(i)}p(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)}). \tag{13}$$

The choice will be called the auxiliary SD (ASD). The weights are then given as

$$\tilde{\mathbf{w}}_k^{(i)} = \frac{p(\mathbf{z}_k|\mathbf{x}_k^{(i)})}{\lambda_k^{(j_i)}}\mathbf{w}_{k-1}^{*(j_i)}. \tag{14}$$

The primary weight $\lambda_k^{(i)}$ in (13) should reflect quality of the not yet drawn particle $\mathbf{x}_k^{(i)}$ with respect to the measurement $\mathbf{z}_k$. The unnormalized primary weight $\lambda_k^{(i)}$ is given as

$$\lambda_k^{(i)} = p(\mathbf{z}_k|\mathbf{x}_{k-1}^{*(i)}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)})\mathrm{d}\mathbf{x}_k. \tag{15}$$

As an analytic computation of (15) can not be usually performed, in [12] the following approximation of (15) was proposed

$$\lambda_k^{(i)} = p(\mathbf{z}_k|\mu_k^{(i)}), \tag{16}$$

where $\mu_k^{(i)}$ is the mean, the mode or another likely value associated with $p(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)})$. The auxiliary sampling density (13) with the primary weights (16) will be called the point auxiliary sampling density (PASD). The primary weights (16) can be seen as a comparison of predictive and measurement information about the state $\mathbf{x}_k$. The measurement information is given by the measurement pdf and the predictive information is given by the predictive point estimate $\mu_k^{(i)}$. Although the PASD uses the measurement $\mathbf{z}_k$ and contains more information than the PSD, the PF with the PASD need not necessarily yield better results due to employing a point estimate $\mu_k^{(i)}$ only [12].

In [17] it was suggested to solve the integral in (15) using the unscented transformation [18] to achieve better results than the PASD.

Finally, in [19] the primary weight $\lambda_k^{(i)}$ was considered as a product of a comparison of prior information given by $p(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)})$ and measurement information given by $p(\mathbf{z}_k|\mathbf{x}_k)$. The comparison is accomplished using a metric. Also this primary weights computation should provide results of higher quality as full prior information and measurement information are compared.

5) Likelihood Sampling Density: In [20] the likelihood SD, which is given by the measurement pdf $p(\mathbf{z}_k|\mathbf{x}_k)$ (or likelihood), was mentioned . To employ the measurement pdf as a SD, it must be firstly inverted to become a pdf of $\mathbf{x}_k$. This may not be always possible, especially when $\mathbf{h}_k(\mathbf{x}_k)$ is a many-to-one function (usually for $n_x > n_z$).

The weights are then proportional to the transition pdf. Note that the likelihood SD can be advantageous for the measurement pdf tighter than the transition pdf and for the measurement pdf close to the filtering pdf.

6) Gradient-Based Prior Sampling Density: In [21] the gradient-based PSD was discussed. The procedure is based on sampling from the PSD $\mathbf{x}_k^{(i)} \sim p(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)})$ followed by moving the samples into a high-likelihood region using the well-known gradient descent technique as

$$\bar{\mathbf{x}}_k^{(i)} = \mathbf{x}_k^{(i)} - \eta\frac{\partial(\mathbf{z}_k - \mathbf{h}_k(\mathbf{x}))^2}{\partial\mathbf{x}}|_{\mathbf{x}=\mathbf{x}_k^{(i)}},$$

where the parameter $0 < \eta < 1$ is a learning rate parameter. This shifting of the samples should increase estimate quality with respect to the PSD. After resampling the samples $\{\mathbf{x}_{k+1}^{(i)}\}_{i=1}^N$ at the next time instant are generated from $p(\mathbf{x}_{k+1}|\bar{\mathbf{x}}_k^{*(i)})$.

A similar technique was proposed in [22] where the PF was used for training neural networks under the name hybrid gradient descent/sampling importance resampling algorithm (HySIR).

7) Hybrid Sampling Density: For state spaces with $n_x > 1$ a combination of the optimal SD and the PSD was adopted in [11] under the name hybrid importance function. To apply the hybrid SD, the state is partitioned into two parts in a suitable way (e.g. $\mathbf{x}_k = (\mathbf{x}_{1,k}, \mathbf{x}_{2,k})$ and then the first part of each sample is drawn from the PSD or possibly the ASD and the second part is drawn from the optimal SD. The local SD has then the following form

$$\pi(\mathbf{x}_k|\mathbf{x}^{*(i)}_{k-1}, \mathbf{z}_k) = p(\mathbf{x}_{1,k}|\mathbf{x}^{*(i)}_{1,k-1}).p(\mathbf{x}_{2,k}|\mathbf{x}^{(i)}_{1,k}, \mathbf{x}^{*(i)}_{k-1}, \mathbf{z}_k) \quad (17)$$

This hybrid SD is advantageous for systems where the optimal SD can not be found for whole state but for a certain part only.

Note that this idea can be used combining any other SD's provided that the state is partitioned suitably and choosing a proper SD for each part.

8) Bridging Density Sampling: The bridging density is a technique for drawing samples proposed in [23]. The idea is to replace a single transition of samples using the SD by a sequence of bridging densities placed between the initial density $\Pi_{k,0} = \pi(\mathbf{x}_k|\mathbf{x}^{*(i)}_{k-1}, \mathbf{z}_k)$ and the final density $\Pi_{k,M+1} = p(\mathbf{x}_k|\mathbf{z}^k)$. The intervening densities $\Pi_{k,1} \dots \Pi_{k,M}$ are given as

$$\Pi_{k,m} \propto \pi(\mathbf{x}_k|\mathbf{x}^{*(i)}_{k-1}, \mathbf{z}_k)^{\alpha_m} p(\mathbf{x}_k|\mathbf{z}^k)^{1-\alpha_m},$$

with $1 > \alpha_1 > \alpha_2 > \cdots \alpha_M > 0$. As it holds that $p(\mathbf{x}_k|\mathbf{z}^k) \propto p(\mathbf{x}_k|\mathbf{z}^{k-1})p(\mathbf{z}_k|\mathbf{x}_k)$, it can be seen that the bridging densities gradually introduce more of the likelihood starting with flat likelihood for $\alpha_0 = 1$ and finishing with a peaked one for $\alpha_{M+1} = 1$. Further details concerning the bridging densities can be found in [24]. A very similar technique to bridging densities called annealed importance sampling was proposed in [25]. They differ in the way the samples are drawn.

9) Partitioned Sampling: Partitioned sampling is a strategy introduced in [26] and motivated by objects tracking. The key idea is to divide the state space into two or more "partitions" and to apply the dynamics for each partition sequentially. Sampling of each partition is followed by a weighted resampling that acts as an operation "populating" area of interest with samples.

The point of partitioning the state space here is not to use different SD for each partition as in the hybrid SD but to decrease computational demands as the number of samples devoted to each partition may vary and thus many computations of the measurement pdf may be saved.

The weighted resampling operation corresponds in fact to the ASD as in the weighted sampling the samples are resampled according to new suitably chosen importance weights that are in accordance with the primary weights.

Of course, partitioned sampling can be realized under several conditions, e.g. it must be possible to partition the state space, to decompose the dynamics and also the weighting function used in weighted resampling must be defined for the partitions. For further details refer to [27].

B. Composite Approach to Sampling Density Design

The composite approach to SD design represents a combination of the importance sampling technique with another nonlinear filtering method acting as a generator of SD for the PF. These methods include both local estimation methods providing results valid in a small area in the state space only and global estimation methods providing results valid in almost whole state space. It should be mentioned that the utilization of another estimation method as a generator of the SD increases computational demands of the PF.

To clarify this approach, suppose samples $\{\mathbf{x}^{(i)}_{k-1}\}_{i=1}^N$ and corresponding weights $\{\mathbf{w}^{(i)}_{k-1}\}_{i=1}^N$ are given at the time instant $k-1$. The approximate predictive pdf $\hat{p}(\mathbf{x}_k|\mathbf{z}^{k-1})$ is given by

$$\hat{p}(\mathbf{x}_k|\mathbf{z}^{k-1}) = \sum_{i=1}^{N} \mathbf{w}^{(i)}_k p(\mathbf{x}_k|\mathbf{x}^{*(i)}_{k-1}),$$

where $p(\mathbf{x}_k|\mathbf{x}^{*(i)}_{k-1}) = p_{\mathbf{v}_k}(\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}^{*(i)}_{k-1}))$. The closed form of the optimal local SD $p(\mathbf{x}_k|\mathbf{x}^{*(i)}_{k-1}, \mathbf{z}_k)$ given by the BRR (3) as $p(\mathbf{x}_k|\mathbf{x}^{*(i)}_{k-1}, \mathbf{z}_k) \propto p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}^{*(i)}_{k-1})$ can not be usually found so other estimation methods (e.g. the extended Kalman filter (EKF), the Gaussian sum filter (GSF), the sigma point filter etc.) are utilized to approximate it.

1) Extended Kalman Filter Sampling Density: The EKF approximates the optimal SD by a Gaussian pdf. It uses the Taylor series expansion first order of the function $\mathbf{h}_k(\cdot)$ and also assumes Gaussian random variables so only the first two moments of the state noise $\mathbf{e}_k$ and the measurement noise $\mathbf{v}_k$ are taken into account.

As the EKF supposes a Gaussian pdf of the noises and the linearization of $\mathbf{h}_k(\cdot)$ is done at a point only, it may not provide quality approximation of the optimal SD for non-Gaussian pdf's and especially for multimodal pdf's. The EKF SD was adopted in e.g. [28], [22]. Note that a combination of the PF and the Kalman filter also appeared in [29] for conditional linear Gaussian systems.

2) Gaussian Sum Sampling Density: A straightforward extension of the EKF SD is utilization of the GSF [30] as a SD generator. The GSF can be seen as a mixture of EKF's where all the random variables are described by weighted sums of Gaussian pdf's. As $\mathbf{h}_k(\cdot)$ is linearized at multiple points and any pdf can be approximated by a weighted sum of Gaussian pdf with arbitrary accuracy [31], this should be a preferable choice for non-Gaussian pdf's of the noises.

A combination of the GSF and the PF was proposed in [32] where a Gaussian sum approximation of the filtering and the predictive pdf's was retrieved from the samples and associated weights after each step of the PF algorithm.

3) Sigma Point Kalman Filter Sampling Density: A very similar idea as for the EKF SD was adopted in [13] where the unscented Kalman filter (UKF) [33], belonging to sigma point Kalman filters, was used to provide the SD for the PF. As the UKF provides higher estimation quality than the EKF, the UKF based SD should provide higher quality samples than the EKF SD.

4) Gaussian Mixture Sigma Point Kalman Filter Sampling Density: In [13] also a straightforward extension to the UKF SD was proposed. The Gaussian mixture sigma point Kalman filter is in this case a mixture of UKF's and thus this SD is an analogy to the GSF SD. It can be said that this choice of SD should provide a higher quality PF than the UKF SD especially for non-Gaussian pdf's of the noises.

5) H_∞ Filter Sampling Density: Another filter for generating SD was adopted in [34] where the extended H_∞ filter was applied.

IV. ILLUSTRATION EXAMPLE

In this section the PF with some interesting SD's from both direct and composite approaches were applied to estimate the state of the following one-dimensional linear non-Gaussian system with nonlinear measurement [13]

$$x_{k+1} = \varphi_1 x_k + 1 + \sin(\omega\pi k) + e_k$$
$$z_k = \varphi_2 x_k^2 + \upsilon_k$$

with $p(x_0) = \mathcal{N}\{x_0; 0, 12\}$, $p(e_k) = G\{e_k, 3, 2\}$, $p(\upsilon_k) = \mathcal{N}\{\upsilon_k; 0, 10^{-5}\}$,$k = 0, 1 \dots K$, $K = 19$, $\varphi_1 = 0.5$, $\varphi_2 = 0.2$, $\omega = 0.04$. The following SD's were considered: the prior (PSD), the point auxiliary (PASD), the unscented transformation based auxiliary (UTASD), the likelihood (LSD), the EKF generated (EKFSD), the GSF generated (GSFSD), the UKF generated (UKFSD), and the Gaussian mixture UKF generated (GMUKFSD). The results were compared considering both point estimates and filtering pdf estimates.

A. Point Estimates Comparison

To compare point estimates of the PF, firstly the mean square error (MSE) Ξ_k is estimated using S MC simulations with $S = 1000$ as $\Xi_k = \frac{1}{S}\sum_{s=1}^{S}\left(x_k(s) - \hat{x}_k(s)\right)^2$, where $x_k(s)$ is the state in the sth simulation and $\hat{x}_k(s) = \sum_{i=1}^{N}\mathbf{w}_k^{(i)} x_k^{(i)}(s)$ is the filtering mean in the sth simulation. Using the MSE the following criterion is evaluated

$$V_{MSE} = \frac{1}{K}\sum_{k=0}^{K}\Xi_k. \tag{18}$$

The values of the criterion (18) for $N = 100$ are given in Table I. It can be seen that value of the criterion (18) is

TABLE I

ESTIMATION QUALITY AND COMPUT. TIME WITH DIFFERENT SD'S

	PSD	PASD	UTASD	LSD
V_{MSE}	13.53	13.72	7.27	0.86
T	0.0020	0.0020	0.0030	0.0080

	EKFSD	GSFSD	UKFSD	GMUKFSD
V_{MSE}	2.06	1.28	1.49	1.28
T	0.0075	0.0170	0.0060	0.0185

low for the PF with the SD of the composite approach and with the LSD. As far as the PF with the PSD, PASD and UTASD are concerned, their quality is bad in terms of point estimates. The highest quality is provided by the LSD. This is caused by variance of the measurement noise ($\Sigma_{\upsilon,k} = 10^{-5}$) which is substantially lower than the variance of the state noise ($\Sigma e, k = 12$). For higher number of samples ($N = 500$ and $N = 1000$) the value of the criterion (18) for the PSD, the PASD and the UTASD decreases rapidly while it almost does not change for the LSD and the composite approach SD's.

B. Filtering PDF Estimates Comparison

The filtering pdf's of the PF $r_N(x_k|z^k)$ were compared using the following criterion

$$V_{PDF} = 1 - n\sum_{k=1,s=1}^{K,S}\int \min\left(p(x_k|z^k(s)), r_N(x_k|z^k(s))\right)\mathrm{d}x_k \tag{19}$$

with $n = \frac{1}{KS}$. The criterion measures a distance between the exact filtering pdf $p(x_k|z^k)$ and the approximate filtering pdf given by the PF. The criterion is attractive as it assumes values between zero and one (i.e. $V_{PDF} \in [0,1]$) and the lower the value is the closer the pdf's are. Note that the term $\int \min\left(p(x_k|z^k), r_N(x_k|z^k)\right)\mathrm{d}x_k$ is called the Bayesian error. The filtering pdf $p(x_k|z^k)$ is computed using the point mass method [35] which performs numerical calculations of the BRR (3) and (4). The experiments were completed using $N = 50, 100, 500, 1000$ samples. Values of the criterion were computed using $S = 1000$ MC simulations and are given in Fig. 1. The black bar is for $N = 50$ samples, the dark gray one for $N = 100$ samples, the light gray one for $N = 500$ samples and the white one for $N = 1000$ samples. As far

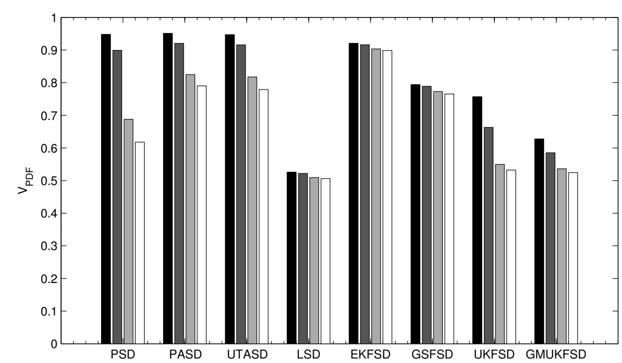

Fig. 1. Filtering pdf estimates comparison

as the filtering pdf estimates are considered, it can not be generally said which SD provides quality estimates as the results of the criterion (19) markedly depends on the sample size. Rate of quality increase with increasing sample size depends on the SD in this case significantly which is a very interesting fact. For example quality increases for the PSD much faster than for the LSD.

For the composite approach SD's it holds that the better estimator provides the SD the higher PF estimate quality is. Also the PASD and UTASD can provide worse results in terms of the criterion (19) than the PSD as their estimates of the integral in (15) might be bad. The best estimates are provided by the LSD as in the point estimate case.

C. Computational time

As far as computational time is considered, the length of a time step in the MATLAB environment on a 3.2 GHz

PC are given in Table I. Low computational time is typical for the SD's of the direct approach while the SD's of the composite approach are almost by an order more computationally demanding. This fact results from application of other estimation method for each sample in the PF algorithm.

V. CONCLUSIONS

The paper provided a survey of sampling densities for the PF. The SD's were classified and both, the direct and the composite approach to the SD design were discussed. Estimation quality and computational demands of the PF with different SD's were analyzed in an illustration example.

The illustration example indicated that utilization of the composite approach in the SD design brings quality increase in terms of both the point estimates and the filtering pdf estimates. The direct approach may also provide high quality results but it is necessary to respect the system characteristics and to use a suitable SD. In this case the variance of the measurement noise was substantially lower than the variance of the state noise and thus the likelihood SD based on the measurement pdf rendered the highest quality results here. Higher computational demands are drawback of the composite approach SD's as they may be almost an order higher than for the direct approach SD's. Also utilization of another estimator as the SD generator poses higher theoretical requirement to the composite approach SD.

For high-dimensional cases the hybrid SD or the partitioned sampling from the direct approach should be employed as they could increase estimation quality.

VI. ACKNOWLEDGMENTS

The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic, project No. 1M0572 and by the Grant Agency of the Czech Republic, project GACR 102/05/2075.

REFERENCES

[1] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE proceedings-F*, vol. 140, pp. 107–113, 1993.

[2] J. Handshin, "Monte Carlo techniques for prediction and filtering of nonlinear stochastic processes," *Automatica*, vol. 6, pp. 553–563, 1970.

[3] J. Handshin and D. Mayne, "Monte Carlo techniques to estimate the condidional expectation in multi-stage non-linear filtering," *International Journal of Control*, vol. 9, pp. 547–559, 1969.

[4] V. Zaritskii, V. Svetnik, and L. Shimelevich, "Monte Carlo technique in problems of optimal data processing," *Automatic Remote Control*, pp. 95–103, 1975.

[5] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001, (Ed. Doucet A., de Freitas N., and Gordon N.).

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

[7] O. Straka and M. Šimandl, "Adaptive particle filter based on fixed efficient sample size," in *Preprints of the 14th IFAC Symposium on System Identification*, Newcastle, 2006.

[8] J. Liu and R. Chen, "Sequential Monte Carlo Methods for Dynamic Systems," *J. Amer. Statist. Assoc.*, vol. 93, no. 443, pp. 1032–1044, 1998.

[9] J. Liu, R. Cheng, and T. Logvinenko, *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. Springer, 2001, ch. A Theoretical Framework for Sequential Importance Sampling with Resampling, pp. 225–246, (Ed. Doucet A., de Freitas N., and Gordon N.).

[10] J. Geweke, "Bayesian inference in econometric models using Monte Carlo Integration," *Econometrika*, vol. 24, pp. 1317–1399, 1989.

[11] Y. Huang and P. Djuric, "A hybrid importance function for particle filtering," *IEEE Signal Processing Letters*, vol. 11, no. 3, pp. 404–406, 2004.

[12] M. Pitt and N. Shephard, *Sequential Monte Carlo Methods in Practice*. Springer, 2001, ch. Auxiliary Variable Based Particle Filter, pp. 273–294, (Ed. Doucet A., de Freitas N., and Gordon N.).

[13] R. van der Merwe and E. Wan, "Sigma-point particle filters for sequential probabilistic inference in dynamic state-space models," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Hong Kong: IEEE, 4 2003.

[14] H. Akashi and H. Kumamoto, "Random sampling approach to state estimation in switching environments," *Automatica*, pp. 429–434, 1977.

[15] H. Tanizaki, *Nonlinear Filters: Estimation and Applications*, ser. Lecture Notes in Economics and Mathematical Systems. Berlin: Springer, 1993.

[16] M. Pitt and N. Shephard, "Filtering via simulation: auxiliary particle filter," *J. Amer. Statist. Assoc.*, vol. 94, pp. 590–599, 1999.

[17] C. Andrieu, M. Davy, and A. Doucet, "Improved auxiliary particle filtering: Applications to time-varying spectral analysis," *IEEE Workshop on Statistical Signal Processing*, 2001.

[18] S. Julier and J. Uhlmann, "The scaled unscented transformation," in *Proceedings of the IEEE American Control Conference*. Anchorage AK, USA: IEEE, 2002, pp. 4555–4559.

[19] O. Straka and M. Šimandl, "Using the bhattacharyya distance in functional sampling density of particle filter," in *Preprints of the 16th IFAC World Congress*, P. Horacek, M. Simandl, and P. Zitek, Eds. Prague: IFAC, 2005.

[20] Z. Chen, "Bayesian filtering: From kalman filters to particle filters, and beyond." 2003, [Online], http://soma.crl.mcmaster.ca/zhechen/homepage.htm.

[21] ——, "Robust particle filters with application in tracking and communications," Adaptive Syst. Lab., McMaster Univ., Hamilton, ON, Canada, Tech. Rep., 2003.

[22] J. De Freitas, M. Niranjan, A. Gee, and A. Doucet, "Sequential Monte Carlo Methods to Train Neural Network Models," *Neural Computation*, no. 12, pp. 955–993, 2000.

[23] S. Godsill, A. Doucet, and M. West, "Maximum a posteriori sequence estimation using Monte Carlo particle filters," *Ann. Inst. Statist. Math*, no. 1, pp. 82–96, 2001.

[24] S. Godsill and T. Clapp, *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. Springer, 2001, ch. Improvement Strategies for monte Carlo Particle Filters, pp. 139–158, (Ed. Doucet A., de Freitas N., and Gordon N.).

[25] R. Neal, "Annealed importance sampling," University of Toronto, Toronto, Ontario, Canada, Tech. Rep. 9805, 1998.

[26] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," in *Proc. 7th International Conf. Computer Vision*, 1999, pp. 572–578.

[27] J. MacCormick and M. Isard, "Partitioned sampling, articulated objects, and interface-quality hand tracking," Tech. Rep., 2000.

[28] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[29] R. Chen and J. Liu, "Mixture Kalman filters," *Journal of the Royal Statistics Society*, vol. 62, pp. 493–508, 2000.

[30] H. Sorenson and D. Alspach, "Approximation of Density Function by a Sum of Gaussians for Nonlinear Bayesian Estimation," in *Proceedings of the 1st Symposium for Nonlinear Estimation Theory and Its Applications*, San Diego, 1970, pp. 88–90.

[31] ——, "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, vol. 7, pp. 465–479, 1971.

[32] J. Kotecha and P. Djuric, "Gaussian sum particle filtering," in *IEEE Transactions on Signal Processing*, ser. 51, no. 10, 2003.

[33] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proceedings of the American Control Conference*, 1995, pp. 1628–1632.

[34] K. Nishiyama, "Fast and effective generation of the proposal distribution for particle filters," *Signal Processing*, no. 85, pp. 2412–2417, 2005.

[35] M. Šimandl, J. Královec, and T. Söderström, "Anticipative Grid Design in Point-Mass Approach to Nonlinear State Estimation," *IEEE Transactions on Automatic Control*, vol. 47, no. 4, 2002.