Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Adaptive particle filter with fixed empirical density quality

Ondřej Straka [*] **Miroslav Šimandl** [*]

[*] *Department of Cybernetics and Research Centre: Data - Algorithms - Decision, Faculty of Applied Sciences, University of West Bohemia Univerzitní 8, 306 14 Plzeň, Czech Republic; (e-mail: straka30@kky.zcu.cz, simandl@kky.zcu.cz)*

Abstract: The paper deals with the particle filter in state estimation of a discrete-time nonlinear non-Gaussian system. The goal of the paper is to design a sample size adaptation technique to guarantee the quality of an empirical probability density function (pdf) which approximates a target filtering pdf. The quality is measured by inaccuracy (cross-information) between the empirical pdf and the filtering pdf. It is shown that for increasing sample size the inaccuracy converges to the Shannon differential entropy (SDE) of the filtering pdf. The proposed technique adapts the sample size to keep a difference between the inaccuracy and the SDE within pre-specified bounds with a pre-specified probability. The particle filter with the proposed sample size adaptation technique is illustrated in a numerical example.

1. INTRODUCTION

Recursive state estimation of discrete-time nonlinear stochastic dynamic systems from noisy measurement data has been the subject of a considerable research interest over the last three decades. General solution of the state estimation problem is described 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 applied.

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 swarm of samples (particles) with associated relative weights.

The fundamental paper dealing with the MC solution of the BRR was published by Gordon et al. [1993] who proposed the first effective PF called the bootstrap filter. Many improvements of the bootstrap filter have been proposed since that time, see for example Doucet et al. [2001]. Among these improvements, in particular the design of the sampling probability density function (pdf) as one of the key parameters of the PF has to be mentioned. An overview of sampling pdf's can be found in Simandl and Straka [2007].

Another key parameter of the PF significantly affecting estimate quality is sample size (i.e. the number of the particles), nonetheless efficient sample size setting has been disregarded for a long time. The sample size is usually determined empirically. Some advances in a suitable sample size setting were achieved in Simandl and Straka [2002] where the time-invariant sample size was considered and the Cramér Rao bound [Simandl et al., 2001] was used as a gauge for quality evaluation of the PF. Sample size adaptation (SSA) techniques were treated for example in Koller and Fratkina [1998], Fox [2003], Soto [2005], Straka and Simandl [2006]. These papers focused on sample size adaptation with respect to a point estimate quality usually measured by mean square error (MSE).

The goal of this paper is to propose a sample size adaptation technique that focuses on pdf estimation. The intention is to adapt the number of samples while keeping a distance between the empirical pdf produced by the PF and the target filtering pdf fixed. To measure the distance, inaccuracy will be used as a key component of the Kullback-Leibler distance where it serves for actual comparison of the pdf's.

The paper is organized as follows: State estimation by the PF and a short survey of the SSA techniques are given in Section 2. Then the proposed SSA technique with a fixed empirical density quality is presented in Section 3 which consists of the basic idea of the technique, specification of the distance between the empirical pdf and the target filtering pdf, and computational issues. Further, an application of the proposed SSA technique is illustrated in a numerical example in Section 4 and finally, Section 5 concludes the paper.

2. STATE ESTIMATION BY THE PARTICLE FILTER

This section deals with the state estimation using the PF and a brief survey of several SSA techniques proposed so far.

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 $\mathbf{x}_k \in \mathbb{R}^n$ and $\mathbf{z}_k \in \mathbb{R}^m$ represent a state of the system and a measurement at time k, respectively, $\mathbf{e}_k \in \mathbb{R}^n$ and $\mathbf{v}_k \in \mathbb{R}^m$ 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, $\mathbf{f}_k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, $\mathbf{h}_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ are known vector functions and the pdf $p(\mathbf{x}_0)$ of the initial state $\mathbf{x}_0$ is known. The 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 general solution of the state estimation problem in the form of the filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$ with $\mathbf{z}^k \triangleq [\mathbf{z}_0^{\mathrm{T}}, \ldots, \mathbf{z}_k^{\mathrm{T}}]^{\mathrm{T}}$ is provided by the BRR. The PF is based on the importance sampling method [Tanner, 1996] which means that to find a property of an unknown arbitrary target pdf $p(\mathbf{x})$, firstly N samples are drawn from a sampling (proposal) pdf $\pi(\mathbf{x})$, then the weights $w(\mathbf{x}) \propto \frac{p(\mathbf{x})}{\pi(\mathbf{x})}$ are attached to the samples, and finally the property is approximated using the samples and the weights. The idea of the PF in nonlinear state estimation is to approximate the target 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 $\{\mathbf{w}_k(\mathbf{x}_k^{(i)})\}_{i=1}^{N_k}$. The general algorithm of the PF [Liu et al., 2001] can be summarized in Alg. 1 as follows:

Alg. 1: particle filter

Sampling:

- If $k = 0$, draw N_0 samples $\{\mathbf{x}_0^{(i)}\}_{i=1}^{N_0}$ from the prior pdf $p(\mathbf{x}_0|\mathbf{z}^{-1}) = p(\mathbf{x}_0)$.
- If $k > 0$, draw N_k samples $\{\mathbf{x}_k^{(i)}\}_{i=1}^{N_k}$ from the global sampling pdf $\pi(\mathbf{x}_k|\mathbf{x}_k^{*(1:N_{k-1})}, \mathbf{z}_k)$ where

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(1:N_{k-1})},\mathbf{z}_k)=\sum_{i=1}^{N_{k-1}} \mathbf{v}_{k-1}^{(i)}\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)},\mathbf{z}_k). \quad (3)$$

Weighting:

- If $k = 0$, compute the weights $\{\tilde{\mathbf{w}}_0(\mathbf{x}_0^{(i)})\}_{i=1}^{N_0}$

$$\tilde{\mathbf{w}}_0(\mathbf{x}_0^{(i)}) = p(\mathbf{z}_0|\mathbf{x}_0^{(i)}). \quad (4)$$

- If $k > 0$, the weights $\{\mathbf{w}_k(\mathbf{x}_k^{(i)})\}_{i=1}^{N_k}$ are calculated using the following relation

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

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

$$r_{N_k}(\mathbf{x}_k|\mathbf{z}^k) = \sum_{i=1}^{N_k} \mathbf{w}_k(\mathbf{x}_k^{(i)})\delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}),$$

where $\delta(\cdot)$ is the Dirac function defined as $\delta(\mathbf{x}) = 0$ for $\mathbf{x} \neq 0$ and $\int \delta(\mathbf{x})\mathrm{d}\mathbf{x} = 1$.

Resampling: Generate a new set $\{\mathbf{x}_k^{*(i)}\}_{i=1}^{N_k}$ by resampling with replacement N_k times from $\{\mathbf{x}_k^{(i)}\}_{i=1}^{N_k}$ with probability $Prob(\mathbf{x}_k^{*(i)}=\mathbf{x}_k^{(i)}) = \mathbf{w}_k(\mathbf{x}_k^{(i)})$ and set $\mathbf{w}_k(\mathbf{x}_k^{*(i)}) = \frac{1}{N_k}$. Replace the sets $\{\mathbf{x}_k^{(i)}\}_{i=1}^{N_k}$ and $\{w_k(\mathbf{x}_k^{(i)})\}_{i=1}^{N_k}$ by the resampled sets $\{\mathbf{x}_k^{*(i)}\}_{i=1}^{N_k}$ and $\{w_k(\mathbf{x}_k^{*(i)})\}_{i=1}^{N_k}$ respectively.

Increase k and iterate to step **Sampling**.

Note that the algorithm uses a general sampling pdf $\pi(\mathbf{x}_{k+1}|\mathbf{x}_k^{*(1:N_k)}, \mathbf{z}_{k+1})$ based on utilization of the current measurement $\mathbf{z}_{k+1}$. This general sampling pdf covers either the prior sampling pdf with $\mathbf{v}_{k-1}^{(i)} = 1/N$ and $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)}, \mathbf{z}_k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(i)})$, or other sampling pdf's, e.g. the optimal sampling pdf or the auxiliary sampling pdf. A detailed survey of sampling pdf's can be found in Simandl and Straka [2007].

2.1 Sample size specification

The sample size N_k represents a key parameter of the PF significantly affecting estimate quality. It can be changed to a suitable value at each time instant before the sampling step of Alg. 1. There are several papers dealing with a suitable sample size specification and the proposed techniques are briefly described in this subsection.

Setting a suitable sample size according to an analysis that is carried out beforehand was proposed in Simandl and Straka [2002]. The procedure considers a time-invariant sample size and assesses quality of the PF estimates for various sample sizes according to the distance between the MSE matrix of the state estimate conditional mean and the Cramér-Rao bound. No sample size adaptation is conducted during the estimation process but rather a suitable sample size is specified in advance.

In Koller and Fratkina [1998] an SSA technique was published based on the idea that it would be suitable to keep a fixed sum of likelihoods (i.e. unnormalized weights) of the whole sample set instead of keeping a fixed sample size. The intention is that the samples with low weights do not match the target pdf and therefore more samples are necessary for a quality estimate and vice versa.

In Fox [2003] another sample size adaptation technique called Kullbak-Leibler divergence (KLD) sampling was proposed. It adapts the sample size to bound the error between the true pdf and the empirical pdf by ε with probability $1-\delta$. The error is measured by the KLD. The technique assumes that the true pdf can be represented by a discrete piecewise pdf. The drawback of the proposed technique is that the samples of the empirical pdf are assumed to be drawn directly from the target pdf and the relation for sample size calculation utilizes the number of bins with support as the only information concerning the true pdf. Therefore, the KLD sampling technique can be referred to as an adaptation with respect to complexity of the target pdf, without taking into account the sampling pdf.

In Soto [2005] the KLD sampling technique was elaborated further to grasp the fact that the samples of the empirical pdf are drawn from a sampling pdf which is different from the target pdf. To take into account this fact, the relative accuracy of the estimator between sampling from the target pdf and the sampling pdf was utilized. The accuracy was considered in terms of the point estimate quality.

In Straka and Simandl [2004] the localization-based sample size adaptation (LB-SSA) was proposed based on monitoring quality of the samples generated from the sampling pdf. The quality of the sample set is assessed according to the samples position. Roughly speaking, firstly a criterion is set up with respect to the measurement pdf. Consequently, the samples are being drawn from the sampling density until the criterion respecting their position is met. The LB-SSA allows the estimate quality to be independent of the PF sampling pdf. This means that the PF with a sampling pdf close to the target pdf uses fewer samples than the PF with a sampling pdf which is far from the target pdf.

In Straka and Simandl [2006] another SSA technique was proposed to attain independence of estimate quality from the sampling pdf. The technique is based on a fixed efficient sample size and is much simpler than the LB-SSA technique.

So it can be seen that there are several techniques dealing with sample size adaptation from the viewpoint of a point estimate and its error but none taking the PF as a technique for estimation of the whole filtering pdf and respecting this fact during sample size adaptation. Note that the pdf estimate provides besides the mean of the state many additional pieces of information like for example modes, high-order moments, etc.

3. SAMPLE SIZE ADAPTATION FOR FIXED EMPIRICAL DENSITY QUALITY

As it was mentioned earlier, SSA techniques usually focus on estimation quality measured by MSE of a point estimate, usually mean. The aim of the SSA technique proposed in this paper is to focus on estimation of a complete pdf. As the PF approximates the target filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$ by the empirical filtering pdf $r_{N_k}(\mathbf{x}_k|\mathbf{z}^k)$, the idea is to measure quality of this approximation and keep the quality fixed by adapting sample size. To measure the quality, a suitable criterion has to be chosen. The Kullback-Leibler (KL) distance given by

$$D(p_1, p2) \triangleq \int p_1(\mathbf{x}) \log \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} \mathrm{d}\mathbf{x} \tag{6}$$

is a generic choice of information measure to quantify a discrepancy between two pdf's $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$. The KL distance can be also written as a difference of two components

$$D(p_1, p_2) = \underbrace{\int p_1(\mathbf{x}) \log \frac{1}{p_2(\mathbf{x})} \mathrm{d}\mathbf{x}}_{\mathrm{K}(p_1, p_2)} - \underbrace{\int p_1(\mathbf{x}) \log \frac{1}{p_1(\mathbf{x})} \mathrm{d}\mathbf{x}}_{\mathrm{H}(p_1)} \tag{7}$$

where the former $\mathrm{K}(p_1, p_2)$ is inaccuracy [Kerridge, 1961] and the latter $\mathrm{H}(p_1)$ is the Shannon differential entropy (SDE). The inaccuracy measures actual discrepancy between the pdf's $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$, while the SDE measures entropy of $p_1(\mathbf{x})$. The inaccuracy has opportune form for the comparison of $r_N(\mathbf{x}_k|\mathbf{z}^k)$ and $p(\mathbf{x}_k|\mathbf{z}^k)$ as the empirical pdf is a mixture of Dirac functions and the relation $\int \delta(x-a) f(x) \mathrm{d}x = f(a)$ holds. The SDE component of the KL distance will be dropped as $\mathrm{H}(r_N) = -\infty$. It must be noted that the inaccuracy alone may be negative nevertheless it still provides a measure of disagreement between the pdf's.

3.1 SDE as a limiting value of inaccuracy

For the sake of clarity, consider only the importance sampling method for the derivation of the adaptation technique and subsequently, the proposed technique will be transfered into the PF framework. The sampling density will be denoted as $\pi(\mathbf{x})$, the target (filtering) density $p(\mathbf{x})$ and the importance weight $w(\mathbf{x})$. Note that the weight is unnormalized, i.e. $w(\mathbf{x}) = c\frac{p(\mathbf{x})}{\pi(\mathbf{x})} \propto \frac{p(\mathbf{x})}{\pi(\mathbf{x})}$ where $c = \mathsf{E}_\pi\{w(\mathbf{x})\}$. Suppose N samples $\{\mathbf{x}^{(i)}\}_{i=1}^N$ are drawn from $\pi(\mathbf{x})$; then the empirical pdf $r_N(\mathbf{x})$ approximating $p(\mathbf{x})$ is given as

$$r_N(\mathbf{x}) = \frac{\frac{1}{N}\sum_{i=1}^N w(\mathbf{x}^{(i)})\delta(\mathbf{x} - \mathbf{x}^{(i)})}{\frac{1}{N}\sum_{j=1}^N w(\mathbf{x}^{(j)})}. \tag{8}$$

The inaccuracy $\mathrm{K}(r_N, p)$ given by

$$\mathrm{K}(r_N, p) = \int \frac{\frac{1}{N}\sum_{i=1}^N w(\mathbf{x}^{(i)})\delta(\mathbf{x} - \mathbf{x}^{(i)})}{\frac{1}{N}\sum_{j=1}^N w(\mathbf{x}^{(j)})} \log \frac{1}{p(\mathbf{x})} \mathrm{d}\mathbf{x} \tag{9}$$

can be written as

$$\mathrm{K}(r_N, p) = \frac{\frac{1}{N}\sum_{i=1}^N w(\mathbf{x}^{(i)}) \log \frac{1}{p(\mathbf{x}^{(i)})}}{\frac{1}{N}\sum_{j=1}^N w(\mathbf{x}^{(j)})}. \tag{10}$$

Since the set of samples $\{\mathbf{x}^{(i)}\}_{i=1}^N$ together with the set of weights $\{w(\mathbf{x}^{(i)})\}_{i=1}^N$ approximate the target pdf $p(\mathbf{x})$, the inaccuracy in (10) can be seen as an MC approximation of the integral $\int p(\mathbf{x}) \log \frac{1}{p(\mathbf{x})} \mathrm{d}\mathbf{x}$, which corresponds to $\mathrm{K}(p, p)$, based on the importance sampling method. With $N \to \infty$ the inaccuracy $\mathrm{K}(r_N, p)$ approaches $\mathrm{K}(p, p) = \mathrm{H}(p)$.

$$\lim_{N\to\infty} \mathrm{K}(r_N, p) = \mathrm{K}(p, p) = \mathrm{H}(p).$$

Due to the fact that inaccuracy may be negative, it is hard to specify its desired value. Nonetheless, it is possible to take advantage of the fact that the inaccuracy approaches the SDE and the idea of measuring a distance between the empirical pdf $r_N(\mathbf{x})$ and the target pdf $p(\mathbf{x})$ through the inaccuracy can be converted to measuring a distance between the inaccuracy $\mathrm{K}(r_N, p)$ and the SDE $\mathrm{H}(p) = \mathrm{K}(p, p)$. As the inaccuracy $\mathrm{K}(r_N, p)$ is a random variable, the distance must be understood in a probabilistic sense. Setting up a constraint of this distance while adapting sample size represents the main idea of the proposed SSA which will be denoted information measure SSA (IM-SSA).

The distance between $\mathrm{K}(r_N, p)$ and $\mathrm{H}(p)$ will be measured through their difference for which

$$\lim_{N\to\infty} (\mathrm{K}(r_N, p) - \mathrm{H}(p)) = 0$$

holds. The aim of the adaptation procedure is to allow the user to specify a probability that the difference will be within a user specified bounds. This corresponds to adapting sample size with respect to a user specified quantile of the difference.

3.2 Quantile of a difference between inaccuracy and SDE

The difference between the inaccuracy $\mathrm{K}(r_N, p)$ and the SDE $\mathrm{H}(p)$ is given as

$$\mathrm{K}(r_N, p) - \mathrm{H}(p) = \frac{\frac{1}{N}\sum_{i=1}^{N} w(\mathbf{x}^{(i)})\left(\log\frac{1}{p(\mathbf{x}^{(i)})} - \mathrm{H}(p)\right)}{\frac{1}{N}\sum_{j=1}^{N} w(\mathbf{x}^{(j)})}. \tag{11}$$

Let us denote the ratio in (11) by R. Both nominator and denominator in (11) are given by sample means and according to the central limit theorem (CLT), for $N \to \infty$ they converge to a Gaussian distribution. Note that the CLT can only be applied if means and variances of the terms in the sums are finite. Denote $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N}\left[w(\mathbf{x}^{(i)})\left(\log(\frac{1}{p(\mathbf{x}^{(i)})}) - \mathrm{H}(p)\right)\right]$ and $\overline{W} = \frac{1}{N}\sum_{i=1}^{N} w(\mathbf{x}^{(i)})$, i.e. $Y = W(L - \mathrm{H}(p))$ and $W = w(\mathbf{x})$ with $L = \log(\frac{1}{p(\mathbf{x})})$, then according to the CLT

$$p(\overline{Y}) \xrightarrow[N\to\infty]{} \mathcal{N}\{\overline{Y} : \mu_{\overline{Y}}, \sigma^2_{\overline{Y}}\}$$
$$p(\overline{W}) \xrightarrow[N\to\infty]{} \mathcal{N}\{\overline{W} : \mu_{\overline{W}}, \sigma^2_{\overline{W}}\},$$

where $\mu_{\overline{Y}} = \mu_Y$, $\mu_{\overline{W}} = \mu_W$ $\sigma^2_{\overline{Y}} = \frac{\sigma^2_Y}{N}$ and $\sigma^2_{\overline{W}} = \frac{\sigma^2_W}{N}$ with

$$\mu_Y = 0 \tag{12}$$
$$\mu_W = \mathsf{E}_\pi(W) \tag{13}$$
$$\sigma^2_Y = \mathsf{E}_\pi(W^2L^2) - 2\mathsf{E}_\pi(W^2L)\frac{\mathsf{E}_\pi(WL)}{\mathsf{E}_\pi(W)} + \mathsf{E}_\pi(W^2)\frac{\mathsf{E}^2_\pi(WL)}{\mathsf{E}^2_\pi(W)} \tag{14}$$
$$\sigma^2_W = \mathsf{E}_\pi(W^2) - \mathsf{E}^2_\pi(W). \tag{15}$$

A quantile of the ratio R in (11) as a function of N can not be computed directly due to intricate distribution of R, nevertheless the Geary-Hinkley transformation to normality [Hayya et al., 1975] can be applied. It says that, under a certain condition, the random variable R given by a ratio of two possibly correlated, normal random variables $\overline{Y}$ and $\overline{W}$ may be transformed to a standard normal variable T using the transformation

$$T = \frac{\mu_{\overline{W}}R - \mu_{\overline{Y}}}{\sqrt{\sigma^2_{\overline{W}}R^2 - 2\mathsf{cov}(\overline{Y},\overline{W})R + \sigma^2_{\overline{Y}}}} = \frac{\mu_W R - \mu_Y}{\sqrt{\frac{\sigma^2_W}{N}R^2 - 2\frac{\mathsf{cov}(Y,W)}{N}R + \frac{\sigma^2_Y}{N}}}. \tag{16}$$

The covariance $\mathsf{cov}(\overline{Y},\overline{W}) = \frac{\mathsf{cov}(Y,W)}{N}$ can be expressed as

$$\mathsf{cov}(Y,W) = \mathsf{E}_\pi(W^2L) - \mathsf{E}_\pi(W^2)\frac{\mathsf{E}_\pi(WL)}{\mathsf{E}_\pi(W)}. \tag{17}$$

The transformation (16) holds even for quantiles, i.e.

$$t_{1-\delta/2} = \frac{\mu_W r_{1-\delta/2} - \mu_Y}{\sqrt{\frac{\sigma^2_W}{N}r^2_{1-\delta/2} - 2\frac{\mathsf{cov}(Y,W)}{N}r_{1-\delta/2} + \frac{\sigma^2_Y}{N}}}, \tag{18}$$

where $t_{1-\delta/2}$ is $1-\delta/2$ quantile of the standard normal distribution and $r_{1-\delta/2}$ is $1-\delta/2$ quantile of the distribution of R. The equation (18) represents a relation between a quantile of the standard normal distribution, a quantile of $R = \mathrm{K}(r_N, p) - \mathrm{H}(p)$ and sample size N. Therefore, it is possible to introduce the following relation for sample size N:

$$N = t^2_{1-\delta/2}\frac{\sigma^2_W r^2_{1-\delta/2} - 2\mathsf{cov}(Y,W)r_{1-\delta/2} + \sigma^2_Y}{(\mu_W r_{1-\delta/2} - \mu_Y)^2}. \tag{19}$$

The sample size N given by (19) is necessary for the difference $\mathrm{K}(r_N, p) - \mathrm{H}(p)$ to be within the interval $(-r_{1-\delta/2}, +r_{1-\delta/2})$ with probability $1-\delta$.

The constraint for the sample size N is given by the confidence coefficient $1-\delta$ and the value of $1-\delta/2$ quantile $r_{1-\delta/2}$ which are both chosen by the user.

Note that according to Hayya et al. [1975] the transformation (16) holds as long as the coefficient of variation of the denominator $\overline{W}$, denoted $C_{\overline{W}} = \sigma_{\overline{W}}/\mu_{\overline{W}}$ is less than 0.39. If this condition does not apply, the quantile can be computed numerically but not in terms of a function of the sample size N. In such a case, it is possible to determine an upper bound for the sample size N using Chebychev's inequality

$$Prob(|\mathrm{K}(r_N, p) - \mathrm{H}(p)| \geq \varepsilon) \leq \sqrt{\mathsf{var}\left(\mathrm{K}(r_N, p) - \mathrm{H}(p)\right)}/\varepsilon,$$

in the following form

$$N = \frac{1}{\varepsilon^2\delta}\mathsf{var}(\mathrm{K}(r_N, p) - \mathrm{H}(p)). \tag{20}$$

The relation states that if sample size N given by (20) is used, then

$$Prob(|\mathrm{K}(r_N, p) - \mathrm{H}(p)| \geq \varepsilon) \leq \delta.$$

The term $\mathsf{var}(\mathrm{K}(r_N, p) - \mathrm{H}(p))$ in (20) can be computed using the standard delta method for ratio statistics as

$$\mathsf{var}_\pi(\mathrm{K}(r_N, p) - \mathrm{H}(p)) \approx \Big(\frac{\mathsf{var}_\pi(WL)}{\mathsf{E}^2_\pi(W)} - 2\frac{\mathsf{E}_\pi(WL)\mathsf{cov}_\pi(W, WL)}{\mathsf{E}^3_\pi(W)} + \frac{\mathsf{E}^2_\pi(WL)\mathsf{var}_\pi(W)}{\mathsf{E}^4_\pi(W)}\Big) \tag{21}$$

which can be further explored using second order raw moments as

$$\mathsf{var}_\pi(\mathrm{K}(R_N, p) - \mathrm{H}(p)) \approx \Big(\frac{\mathsf{E}_\pi(W^2L^2)}{\mathsf{E}^2_\pi(W)} - 2\frac{\mathsf{E}_\pi(WL)\mathsf{E}_\pi(W^2L)}{\mathsf{E}^3_\pi(W)} + \frac{\mathsf{E}^2_\pi(WL)\mathsf{E}_\pi(W^2)}{\mathsf{E}^4_\pi(W)}\Big) = \frac{\sigma^2_Y}{\mathsf{E}^2_\pi(W)}. \tag{22}$$

Note that for the sample size given by (20) no information concerning distribution of $\mathrm{K}(r_N, p) - \mathrm{H}(p)$ is used; therefore this condition for the sample size is loose and the relation (19) should be used if possible.

As the terms σ^2_Y and $\mathsf{cov}(Y,W)$ in (19) and (22) depend on the target pdf $p(\mathbf{x})$ which is unknown, the following substitution $p(\mathbf{x}) = \frac{w(\mathbf{x})\pi(\mathbf{x})}{c}$ will be used and after a few arrangements the terms can be expressed as

$$\sigma^2_Y = \mathsf{E}_\pi(W^2\tilde{L}^2) - 2\mathsf{E}_\pi(W^2\tilde{L})\frac{\mathsf{E}_\pi(W\tilde{L})}{\mathsf{E}_\pi(W)} + \mathsf{E}_\pi(W^2)\frac{\mathsf{E}^2_\pi(W\tilde{L})}{\mathsf{E}^2_\pi(W)}$$
$$\mathsf{cov}(Y,W) = l\mathsf{E}_\pi(W^2) + \mathsf{E}_\pi(W^2\tilde{L}) - \mathsf{E}_\pi(W^2)\frac{l\mathsf{E}_\pi(W) + \mathsf{E}_\pi(W\tilde{L})}{\mathsf{E}_\pi(W)}$$

with $\tilde{L} = \log\frac{1}{w(\mathbf{x})\pi(\mathbf{x})}$ and $l = \log(\mathsf{E}_\pi(W))$.

3.3 Computational issues

To implement the proposed IM-SSA in the PF framework, several issues must be discussed. First of all, let us restate individual terms used in the sample size adaptation (19):

$$\mathsf{E}_\pi(W) = \int w(\mathbf{x}_k)\pi(\mathbf{x}_k|\mathbf{z}^k)\mathrm{d}\mathbf{x}_k \tag{23}$$

$$\mathsf{E}_\pi(W^2) = \int w(\mathbf{x}_k)^2\pi(\mathbf{x}_k|\mathbf{z}^k)\mathrm{d}\mathbf{x}_k \tag{24}$$

$$\mathsf{E}_\pi(W\tilde{L}) = \int w(\mathbf{x}_k)\log(\tfrac{1}{w(\mathbf{x}_k)\pi(\mathbf{x}_k|\mathbf{z}^k)})\pi(\mathbf{x}_k|\mathbf{z}^k)\mathrm{d}\mathbf{x}_k \tag{25}$$

$$\mathsf{E}_\pi(W^2\tilde{L}) = \int w(\mathbf{x}_k)^2\log(\tfrac{1}{w(\mathbf{x}_k)\pi(\mathbf{x}_k|\mathbf{z}^k)})\pi(\mathbf{x}_k|\mathbf{z}^k)\mathrm{d}\mathbf{x}_k \tag{26}$$

$$\mathsf{E}_\pi(W^2\tilde{L}^2) = \int w(\mathbf{x}_k)^2\log^2(\tfrac{1}{w(\mathbf{x}_k)\pi(\mathbf{x}_k|\mathbf{z}^k)})\pi(\mathbf{x}_k|\mathbf{z}^k)\mathrm{d}\mathbf{x}_k. \tag{27}$$

Value of the integrals (23-27) cannot be usually computed analytically, therefore they must be calculated either numerically or using MC integration. In the case of the numerical integration, calculation of $\pi(\mathbf{x}) = \pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(1:N_{k-1})},\mathbf{z}_k)$ at an arbitrary point for large N_{k-1} can be computationally demanding as N_{k-1} evaluations of local sampling densities is required for each point. A simple solution to this problem is approximation of the global sampling density by a piecewise linear function which evaluation at an arbitrary point is modest from the computational point of view.

MC integration of the expectations (23-27) is preferable especially for high dimension of the state $\mathbf{x}_k$. To calculate the expectations using MC integration, generate N_{MC} samples from $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(1:N_{k-1})},\mathbf{z}_k)$, utilize them to compute N_k according to (19) and finally, the remaining $N_k - N_{MC}$ samples are drawn.

The PF with IM-SSA will be denoted as the information measure adaptive PF (IM-APF) and its algorithm can be summarized as

Alg. 2: information measure adaptive particle filter

Sample size adaptation: Compute value of the integrals (23-27). Choose the confidence coefficient $1-\delta$ and length of the interval $r_{1-\delta/2}$ and calculate the sample size N_k according to (19).
Sampling: Corresponds to Alg. 1
Weighting: Corresponds to Alg. 1
Resampling: Corresponds to Alg. 1

Increase k and iterate to step **Sample size adaptation**.

Note that for $k = 0$ the sampling density is $\pi(\mathbf{x}) = p(\mathbf{x}_0|\mathbf{z}^{-1})$ and the weight $w(\mathbf{x})$ is given by (4). For $k > 0$ the sampling density $\pi(\mathbf{x})$ is given by (3) and the weight $w(\mathbf{x})$ by (5).

4. NUMERICAL EXAMPLE

To illustrate the proposed IM-SSA, a scalar nonlinear Gaussian system is considered:

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

with $p(x_0) = \mathcal{N}\{x_0; 0, 12\}$, $p(e_k) = \mathcal{G}\{e_k, 3, 2\}$, $p(v_k) = \mathcal{N}\{v_k; 0, 1\}$, $\varphi_1 = 0.5$, $\varphi_2 = 0.2$, $\omega = 0.04$. The state is estimated by the PF for $k = 0, \ldots 29$. The PF considers $p(x_0|z^{-1}) = p(x_0)$ and prior sampling density. The system was simulated 1000 times. Due to lack of space the IM-APF is compared with the unadapted PF only. The IM-APF considered the confidence coefficient $1-\delta/2 = 0.99$ and $r_{1-\delta/2} = 1$. To compare IM-APF and the unadapted PF meaningfully, the sample size for the unadapted PF was calculated in the following way. Firstly, an average N_{AV} of all the sample sizes of the IM-APF was calculated and rounded, i.e.

$$N_{AV} = \lceil \tfrac{1}{30}\tfrac{1}{1000}\sum_{k=0}^{29}\sum_{s=1}^{1000} N_k(s)\rceil = 410,$$

where $N_k(s)$ is sample size of the IM-APF in s-th simulation at the time instant k. Consequently, the unadapted PF was applied twice with $N = N_{AV}$, and $N = 2\cdot N_{AV}$. It should be remembered that the unadapted PF uses some information obtained from the IM-APF (N_{AV}) but does not adapt the sample size N in time.

To compute the true filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$ used for comparison of the results, the point-mass method [Simandl et al., 2006] was used with a large number of grid points. Therefore, its filtering pdf estimate can be treated as the true filtering pdf.

Fig. 1 contains 0.99 quantiles of the difference $\mathrm{K}(r_{N_k}) - \mathrm{H}(p)$ calculated from the simulations for the IM-APF (solid) and unadapted PF with $N = N_{AV}$ (dot-dashed), and $N = 2\cdot N_{AV}$ (dashed). An example of IM-APF sample size evolution for several simulations is given in Fig. 2 (dashed), together with an shaded area formed by 95% sample sizes. From Fig. 1 it is clear that the IM-APF

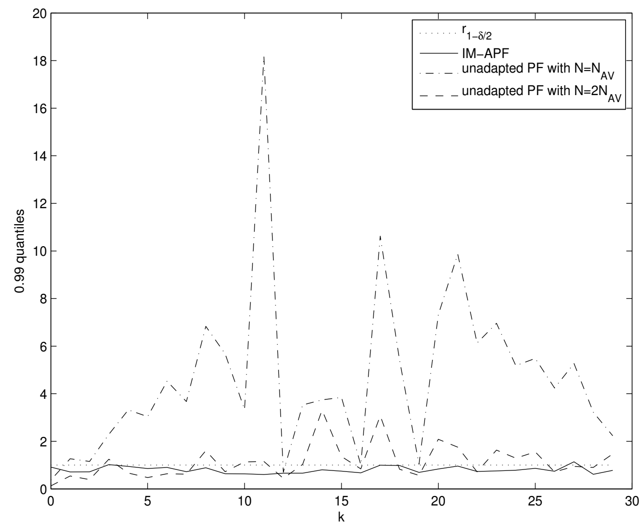

Fig. 1. 0.99 quantiles of the difference between inaccuracy and SDE

adapts the sample size to keep the difference $\mathrm{K}(r_{N_k}) - \mathrm{H}(p)$ within the interval $(-r_{0.99}, r_{0.99})$, $r_{0.99} = 1$ with probability 0.99. The unadapted PF with the same sample size on average as the IM-APF provides much worse results

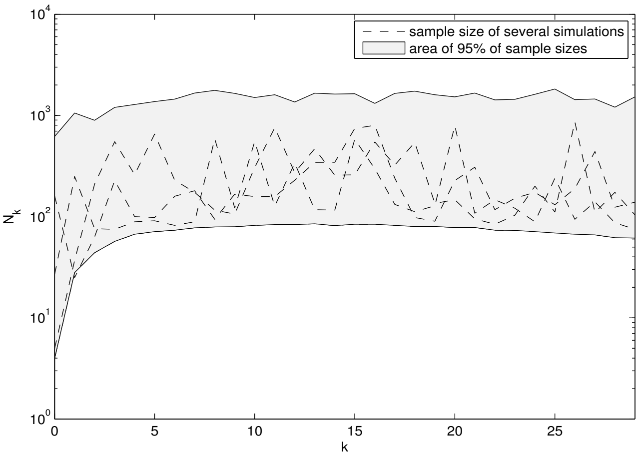

Fig. 2. Sample sizes of the IM-APF

Table 1. Comparison of point estimates quality

	IM-APF	PF, $N = N_{AV}$	PF, $N = 2 \cdot N_{AV}$
$\overline{MSE}$	0.555	0.748	0.588
$\overline{\mathsf{var}(SE)}$	31.868	131.795	86.854

in terms of empirical pdf quality than the IM-APF. If the sample size of the unadapted PF is considered twice as large as the average N_{AV}, then the empirical pdf quality is approximately the same as with the IM-APF. Therefore, in this case the unadapted PF requires more than twice as many samples as the IM-APF to guarantee the same quality of the empirical pdf.

As a matter of interest, the quality of point estimates of the state, i.e. mean, is shown in Table 1. The notation $\overline{MSE}$ represents average mean squared error estimate and the notation $\overline{\mathsf{var}(SE)}$ represents average variance of squared error which exposes variability of the squared error. Both indicators demonstrate superiority of the IM-APF over both unadapted PF's.

5. CONCLUSION

The paper dealt with the particle filter for nonlinear state estimation problem. The information measure sample size adaptation technique was proposed to keep quality of the empirical filtering pdf fixed. The ground of the technique is the Kullback-Leibler distance which can be decomposed into inaccuracy and Shannon differential entropy. The technique is based on the fact that the inaccuracy between the empirical pdf and the target filtering pdf approaches the Shannon differential entropy of the filtering pdf for infinity sample size. The difference between the inaccuracy and the Shannon entropy is utilized for measuring quality of the empirical filtering pdf. The proposed relation for sample size adaptation guarantees that the difference is within a user-specified boundary with a user-specified probability. The paper also discussed implementation issues of the proposed sample size adaptation technique. The particle filter with the information measure sample size adaptation technique was illustrated in a numerical example.

ACKNOWLEDGEMENTS

The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic, project No. 1M0572.

REFERENCES

A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, 2001. (Ed. Doucet A., de Freitas N., and Gordon N.).

D. Fox. Adapting the sample size in particle filters through kld-sampling. *International Journal of Robotics Research*, 22:985–1003, 2003.

N. Gordon, D. Salmond, and A. F. M. Smith. Novel approach to nonlinear/ non-gaussian bayesian state estimation. *IEE Proceedings-F*, 140:107–113, 1993.

J. Hayya, D Armstrong, and N Gressis. A note on the ratio of two normally distributed variables. *Management Science*, 21(11):338–1341, 7 1975.

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

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

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

M. Simandl and O. Straka. Nonlinear estimation by particle filters and cramér-rao bound. In *Proceedings of the 15th Triennial World Congress of the IFAC*, pages 79–84, Barcelona, 2002.

M. Simandl and O. Straka. Sampling densities of particle filter: a survey and comparison. In *Proceedings of the 26th American Control Conference (ACC)*, pages 4437–4442. AACC, 7 2007.

M. Simandl, J. Kralovec, and P. Tichavsky. Filtering, predictive and smoothing cramér-rao bounds for discrete-time nonlinear dynamic filters. *Automatica*, 37(11): 1703–1716, 2001.

M. Simandl, J. Kralovec, and T. Soderstrom. Advanced point – mass method for nonlinear state estimation. *Automatica*, 42(7):1133–1145, 2006.

A. Soto. Self adaptive particle filter. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1398–1406. International Joint Conference on Artificial Intelligence, 2005.

O. Straka and M. Simandl. Sample size adaptation for particle filters. In A. Nebylov, editor, *Preprints of the 16th Symposium on Automatic Control in Aerospace*, volume 1, pages 444–449, Saint Petersburg, Russia, 2004.

O. Straka and M. Simandl. Adaptive particle filter based on fixed efficient sample size. In *Preprints of the 14th IFAC Symposium on System Identification*, Newcastle, 2006.

M. A. Tanner. *Tools for Statistical Inference.* Springer Series in Statistics. Springer Verlag, New York, 3rd edition edition, 1996.