ADAPTIVE PARTICLE FILTER WITH FIXED EMPIRICAL DENSITY QUALITY

Ondřej Straka and Miroslav Šimandl

Research Centre Data, Algorithms, and Decision Making
Department of Cybernetics
University of West Bohemia
Czech Republic

17th IFAC World Congress, 2008

Outline

1. Motivation
2. Particle Filter
3. Sample size adaptation
4. Numerical Illustration
5. Conclusion

Motivation

- the problem of suitable sample size specification usually overlooked in particle filtering
- usually constant sample size is considered while estimate quality varies
- there are few sample size specification techniques adapting with respect to point estimate quality but none respecting pdf estimate quality

- the aim of the proposed sample size specification technique is to adapt the sample size such that the Kullback-Leibler distance between the empirical filtering pdf and the true filtering pdf is preserved

State estimation

Consider a discrete time stochastic system:

$$
\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{f}_k(\mathbf{x}_k, \mathbf{e}_k), \quad k = 0, 1, 2, \ldots \qquad [p(\mathbf{x}_{k+1}|\mathbf{x}_k)] \\
\mathbf{z}_k &= \mathbf{h}_k(\mathbf{x}_k, \mathbf{v}_k), \quad k = 0, 1, 2, \ldots \qquad [p(\mathbf{z}_k|\mathbf{x}_k)]
\end{aligned}
$$

- $\mathbf{x}_k$ is nx dimensional state vector with $p(\mathbf{x}_0)$
- $\mathbf{z}_k$ is nz dimensional measurement vector
- $\mathbf{e}_k$ is white noise with known $p(\mathbf{e}_k)$
- $\mathbf{v}_k$ is white noise with known $p(\mathbf{v}_k)$
- $\mathbf{f}_k(\cdot,\cdot)$ and $\mathbf{h}_k(\cdot,\cdot)$ are known vector functions

The aim of state estimation

$p(\mathbf{x}_k|\mathbf{z}^k) = ?$, with $\mathbf{z}^k = [\mathbf{z}_0^{\mathrm{T}}, \ldots \mathbf{z}_k^{\mathrm{T}}]^{\mathrm{T}}$

Particle filter

General solution of the filtering problem

- given by the Bayesian Recursive Relations (BRR).
- closed form solution available for a few special cases only (e.g. linear Gaussian systems).
- usually *approximate* solution

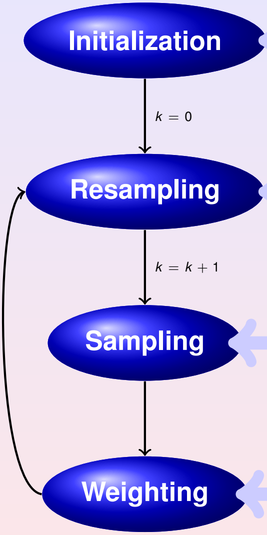
Solution of the BRR by the particle filter

- based on approximating the filtering pdf by a set of N_k samples (particles) and corresponding weights as

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

$\mathbf{x}_k^{(i)}$ - samples, $\omega_k^{(i)}$ - normalized weights,

δ - the Dirac function ($\delta(\mathbf{x}) = 0$ for $\mathbf{x} \neq 0$, $\int \delta(\mathbf{x})\mathrm{d}\mathbf{x} = 1$).



draw $\{\mathbf{x}_0^{(i)}\}_{i=1}^{N_0}$ from $p(\mathbf{x}_0|\mathbf{z}^{-1})$ and set $\omega_0^{(i)} \propto p(\mathbf{z}_0|\mathbf{x}_0^{(i)})$

generate $\{\mathbf{x}_k^{*(i)}\}_{i=1}^{N_k}$ by resampling with replacement from $\{\mathbf{x}_k^{(i)}\}_{i=1}^{N_k}$ according to $P(\mathbf{x}_k^{*(i)} = \mathbf{x}_k^{(i)}) = \omega_k^{(i)}$

draw new samples $\{\mathbf{x}_k^{(i)}\}_{i=1}^{N_k}$ from $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{*(1:N_{k-1})}, \mathbf{z}_k)$

compute the weights

$$\{\omega_k^{(i)}\}_{i=1}^{N_k} \propto \frac{p(\mathbf{z}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{*(i)})}{\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{*(i)},\mathbf{z}_k)}$$

Sample size specification

Only a few papers address sample size

- Non-adaptive sample size specification
  - constant sample size, i.e. $N_k = N$
  - calculating N in advance according to a criterion evaluating estimate quality
  - no increase of computational costs of the actual algorithm

Boers Y.
On the number of samples to be drawn in particle filtering
IEE Colloquium on Target Tracking: Algorithms and Applications, 1999.

Šimandl M. and Straka O.
Nonlinear estimation by particle filters and Cramér Rao bound.
Proceedings of the 15th triennial world congress of IFAC, 2002.

Adaptive sample size specification

- Adaptive techniques for sample size specification
  - always increase computational costs
  - empirical or criteria respecting point estimate quality:

Koller, Fratkina – Using learning for approximation in stochastic processes.
Proc. of 15th Int. Conf. on Machine Learning, 1998.

Fox – KLD sampling: Adaptive particle filter for mobile robot localization.
Advances in Neural Information Processing Systems, 2001.

Soto – Self adaptive particle filter
International Joint Conference on Artificial Intelligence Systems,2005

Straka, Šimandl – Adaptive particle filter based on fixed efficient sample size
Proceedings of the 14th IFAC symposium on System Identification, 2006

Lanz O. – An information theoretic rule for sample size adaptation in particle filtering
14th International Conference on Image Analysis and Processing,2007

Adaptation with fixed empirical pdf quality

Idea:

to keep **Kullback-Leibler (KL) distance** between empirical pdf r_N and true pdf p fixed and to adapt sample size accordingly

$$D(r_N, p) \triangleq \int r_N \log \frac{r_N}{p} \mathrm{d}\mathbf{x} = \underbrace{\int r_N \log \frac{1}{p} \mathrm{d}\mathbf{x}}_{\mathrm{K}(r_N, p)} - \underbrace{\int r_N \log \frac{1}{r_N} \mathrm{d}\mathbf{x}}_{\mathrm{H}(r_N)}$$

- $\mathrm{K}(r_N, p)$ – inaccuracy measuring actual discrepancy between r_N and p
- $\mathrm{H}(r_N)$ – Shannon differential entropy (SDE), further dropped as $\mathrm{H}(r_N) = -\infty$

Adaptation with fixed empirical pdf quality (cont.)

From KL distance to difference between inaccuracy and SDE $\mathrm{K}(p)$

- Instead of KL distance, inaccuracy will be further considered
- the limiting value of inaccuracy $\mathrm{K}(r_N, p)$ is not zero
- it can be shown that

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

- therefore the idea of monitoring the KL distance between r_N and p can be converted to monitoring the distance between inaccuracy $\mathrm{K}(r_N, p)$ and SDE $\mathrm{H}(p)$ as

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

Adaptation with fixed empirical pdf quality (cont.)

The difference between inaccuracy and SDE

$$\mathrm{K}(r_N, p) - \mathrm{H}(p) = \frac{\frac{1}{N}\sum_{i=1}^{N} w(\mathbf{x}^{(i)}) \left(\log \frac{1}{p(\mathbf{x}^{(i)})} - \mathrm{H}(p)\right)}{\frac{1}{N}\sum_{j=1}^{N} w(\mathbf{x}^{(j)})} = \frac{\overline{Y}}{\overline{W}} = R$$

- According to Central Limit Theorem

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

- a quantile of R as a function of N can not be found directly
- nevertheless the Geary-Hinkley transformation to normality can be applied

Adaptation with fixed empirical pdf quality (cont.)

Geary-Hinkley transformation to normality

$$T = \frac{\mu_{\overline{W}} R - \mu_{\overline{Y}}}{\sqrt{\sigma^2_{\overline{W}} R^2 - 2\text{cov}(\overline{Y}, \overline{W}) R + \sigma^2_{\overline{Y}}}}$$

- T has approximately standard normal distribution (under a certain condition)

$$\mu_{\overline{Y}} = 0, \qquad \mu_{\overline{W}} = \mathrm{E}_\pi(W) \qquad \sigma^2_{\overline{W}} = \frac{1}{N}\left[\mathrm{E}_\pi(W^2) - \mathrm{E}^2_\pi(W)\right]$$

$$\sigma^2_{\overline{Y}} = \frac{1}{N}\left[\mathrm{E}_\pi(W^2L^2) - 2\mathrm{E}_\pi(W^2L)\frac{\mathrm{E}_\pi(WL)}{\mathrm{E}_\pi(W)} + \mathrm{E}_\pi(W^2)\frac{\mathrm{E}^2_\pi(WL)}{\mathrm{E}^2_\pi(W)}\right]$$

$$\text{cov}(\overline{Y}, \overline{W}) = \frac{1}{N}\left[\mathrm{E}_\pi(W^2L) - \mathrm{E}_\pi(W^2)\frac{\mathrm{E}_\pi(WL)}{\mathrm{E}_\pi(W)}\right],$$

with $W = w(\mathbf{x})$, $L = \log(\frac{1}{p(\mathbf{x})})$ and $Y = W(L - \mathrm{H}(p))$

Adaptation with fixed empirical pdf quality (cont.)

The transformation holds for quantiles $\Longrightarrow$

$$N = t^2_{1-\delta/2} \frac{\sigma_W^2 r^2_{1-\delta/2} - 2\mathrm{cov}(Y, W) r_{1-\delta/2} + \sigma_Y^2}{(\mu_W r_{1-\delta/2} - \mu_Y)^2}$$

- with user specified parameters
  - confidence coefficient $1-\delta$
  - value of $1-\delta/2$ quantile $r_{1-\delta/2}$
- and $t_{1-\delta/2}$ being $1-\delta/2$ quantile of the standard normal distribution

---

The relation means that N given by it 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 $1-\delta$.

Adaptation with fixed empirical pdf quality (cont.)

Computational aspects

- The second moments $\mathrm{E}_\pi(W)$, $\mathrm{E}_\pi(W^2)$, $\mathrm{E}_\pi(WL)$, $\mathrm{E}_\pi(W^2L)$, $\mathrm{E}_\pi(W^2L^2)$ are computed using Monte Carlo method
  1. N_{MC} samples are firstly generated from π
  2. the second moments are enumerated
  3. the sample size N_k is calculated
  4. $N_k - N_{MC}$ remaining samples are drawn from π
- information measure adaptive PF (IM-APF)
- if the condition for Geary-Hinkley transformation (coefficient of variation of the denominator $\overline{W}$ is less than 0.39) is not fulfilled, Chebychev inequality must be used (providing loose bound for sample size)

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

Example: Adaptive particle filter with fixed empirical density quality

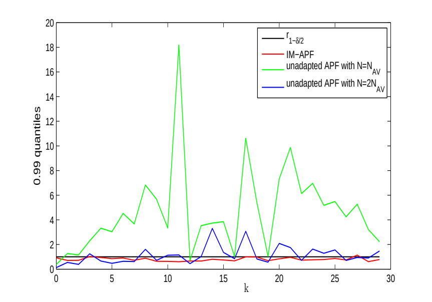
System

$$x_{k+1} = \varphi_1 x_k + 1 + \sin(\omega \pi k) + e_k \qquad p(e_k) = \mathcal{G}\{e_k, 3, 2\}$$
$$z_k = \varphi_2 x_k^2 + v_k \qquad p(v_k) = \mathcal{N}\{v_k : 0, 1\}$$
$$p(x_0) = \mathcal{N}\{x_0 : 0, 12\}$$
$$\varphi_1 = 0.5,\ \varphi_2 = 0.2,\ \omega = 0.04.$$

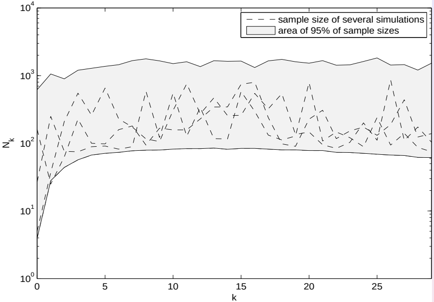
Particle filter

- prior importance function
- $k = 0, 1, \ldots 29$, 1000 MC simulations
- adaptive PF: $1 - \delta/2 = 0.99$ and $r_{1-\delta/2} = 1$
- unadapted PF: $N = N_{AV}$, $N = 2N_{AV}$

Example: Results



0.99 quantiles of the difference between inaccuracy and SDE



Sample sizes of the IM-APF

red -	IM-APF
green -	unadapted PF with $N = N_{AV}$
blue -	unadapted PF with $N = 2N_{AV}$
black -	$r_{1-\delta/2}$

Example: Point estimate quality

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
$\overline{\text{var}(SE)}$	31.868	131.795	86.854

$\overline{MSE}$ - average mean squared error estimate

$\overline{\text{var}(SE)}$ - average variance of squared error

Conclusion

- A sample size adaptation technique was proposed.
- The adaptation is done with respect to empirical pdf quality.
- The difference between inaccuracy $\mathrm{K}(r_N, p)$ and Shannon differential entropy $\mathrm{H}(p) = \mathrm{K}(p, p)$ is kept within a user-specified interval r with user-specified probability $1 - \delta/2$.
- Enumeration of the adapted sample size introduces little extra computational overheads as the samples generated for computing N are reused for computing the empirical pdf.