Sample Size Specification Techniques for Particle Filter

Ondřej Straka

Research Center Data, Algorithms, and Decision
Department of Cybernetics
University of West Bohemia in Pilsen
Czech Republic

March 2008

Outline

1. Particle filtering
2. Sample size specification
3. Computational aspects
4. Numerical examples

System specification

Dynamic stochastic nonlinear non-Gaussian system:

The system is specified by the state and measurement equations

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

and probability density functions $p(\mathbf{e}_k)$, $p(\mathbf{v}_k)$ and $p(\mathbf{x}_0)$,

- The sequences $\{\mathbf{e}_k\}$ and $\{\mathbf{v}_k\}$ are white
- $\mathbf{e}_k$, $\mathbf{v}_k$ and $\mathbf{x}_0$ are mutually independent
- $\mathbf{f}_k(\cdot)$ and $\mathbf{h}_k(\cdot)$ are known

Alternative specification of the system:

by the transition probability density function (pdf) and the measurement pdf

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}),$$
$$p(\mathbf{z}_k|\mathbf{x}_k).$$

State estimation

The aim of state estimation is to find the conditional pdf $p(\mathbf{x}_k|\mathbf{z}^l)$, where $\mathbf{z}^l \triangleq [\mathbf{z}_0^{\mathrm{T}}, \ldots, \mathbf{z}_l^{\mathrm{T}}]^{\mathrm{T}}$.

Particle Filtering - basic idea

- Particle filtering (PF) is currently a rapidly evolving method used for state estimation of discrete-time nonlinear non-Gaussian systems.
- The idea of PF is to approximate the filtering pdf by the empirical filtering pdf r_N, which is given by random samples $\{\mathbf{x}_k^{(i)}\}_{i=1}^N$ of the state and associated weights $\{w_k^{(i)}\}_{i=1}^N$

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

Particle filtering - algorithm

- The general PF algorithm can be decomposed into two principal steps
  1. drawing samples from an importance function specified by the user
  $$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(1:N)},\mathbf{z}_k)$$
  2. computing the corresponding weights utilizing the transition and measurement pdf's as
  $$w_k^{(i)} = \frac{p(\mathbf{z}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}{\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)},\mathbf{z}_k)} w_{k-1}^{(i)}$$
- Obviously, the choice of the sample size N and the importance function π significantly affect quality of the state estimate.

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
- Adaptive techniques for sample size specification
  - always increase computational costs
  - criteria:
    - empirical
    - evaluating *point estimate quality*
    - evaluating *pdf estimate quality*

Non-adaptive techniques

Šimandl M. and Straka O. 2002
Nonlinear estimation by particle filters and Cramér Rao bound.
Proceedings of the 15th Triennial World Congress of IFAC

IDEA: sample size is set according to a distance between the mean square error $\Pi_{k|l}$ and the Cramér Rao bound $\mathbf{C}_{k|l}$

$$V^{CR}_{k|l}(N) = \frac{1}{K}\sum_{k=0}^{K} \mathrm{tr}(|\Pi_{k|l} - \mathbf{C}_{k|l}|)$$

- V^{CR} is calculated for several (N) and the desired N^* is chosen according to the curve "criterion vs sample size"
- the technique requires many MC simulations
- applicable for filtering, prediction and smoothing
- evaluating *point estimate quality*

Likelihood based adaptation

Koller D. and Fratkina R. 1998

Using learning for approximation in stochastic processes.

Proceedings of 15th International Conference on Machine Learning

- Dagum and Luby in “Optimal approximation algorithm for Bayesian inference” argued that actual number of effective samples is their total weight

IDEA: It would be suitable to keep a fixed sum of likelihoods of the whole sample set instead of keeping a fixed sample size

- likelihoods - unnormalized weights
- samples with low weight do not match the target pdf $\Rrightarrow$ more low weighted samples are necessary
- *empirical criterion*

Kullback-Leibler Distance (KLD) sampling

Fox D. 2003

Adapting the Sample Size in Particle Filters through KLD-sampling

International Journal of Robotics Research

IDEA: the error between the true pdf and the empirical pdf is bounded by ε with probability $1-\delta$

- the error is measured by the Kullback-Leibler distance
- the true pdf is supposed to be *discrete, piecewise constant*
- $N_k = \frac{1}{2\varepsilon}\chi^2_{m-1,1-\delta}$
- m is the number of bins of the true pdf with support
- *adaptation with respect to the complexity of the target pdf*

Self adaptive particle filter

Soto A. 2005

Self adaptive particle filter

International Joint Conference on Artificial Intelligence Systems

- elaboration of KLD sampling

IDEA: respecting the fact that the samples are drawn from the importance function (different from the target pdf)

- the relative accuracy measured in terms of point estimates (MSE)

$$\frac{\mathrm{var}_p(\mathbf{x})}{N_k} = \frac{\sigma_{IS}^2}{N_{IS,k}} \quad \Rightarrow \quad N_{IS,k} = \frac{\sigma_{IS}^2}{\mathrm{var}_p(\mathbf{x})} N_k$$

- $N_{IS,k} = \frac{\sigma_{IS}^2}{\mathrm{var}_p(\mathbf{x})} \frac{1}{2\varepsilon} \chi^2_{m-1,1-\delta}$

Self adaptive particle filter - asymptotic normal approximation

IDEA: instead of checking the accuracy of the *true pdf estimate*, it is possible to check the accuracy of the particle filter in *estimation of a moment of the true pdf*

- using strong law of large numbers – estimate of the mean is asymptotically unbiased
- if the variance of estimator is finite – the central limit theorem justifies asymptotic normal approximation for it
- $N_{IS,k} = \frac{\sigma^2_{IS}}{\mathrm{E}_p(x_k)^2}\frac{1}{\varepsilon^2}Z^2_{1-\alpha/2}$

Information theoretic rule adaptation

Lanz O. 2007

An information theoretic rule for sample size adaptation in particle filtering

14th International Conference on Image Analysis and Processing (ICIAP 2007)

- Consequence of the **Asymptotic equipartition property theorem :** *For a given density p and n large, the volume of the smallest n-sized sample set that contains most of the probability is approximately* $e^{nH(p)}$

IDEA: The number of i.i.d. samples needed to properly represent a density p with resolution ρ is

$$N_\rho(p) \sim \rho\, e^{H(p)}$$

- Shannon differential entropy is approximated using kernel density estimation

Localization-basic adaptation

Straka O. and Šimandl M. 2004

Sample size adaptation for particle filters

Proceedings of the 16th IFAC symposium on Automatic Control in Aerospace

IDEA: checking position of the generated samples according to the measurement pdf

- at least ν_k of N_k generated samples are located in the significant part $\mathscr{A}_k$ of the measurement pdf support.
- ν_k can be specified using the measurement pdf (idea of relative accuracy)
- $$N_k = \bar{N} \frac{\int_{\mathscr{A}_k} p(y_k|z_k)\mathrm{d}y_k}{\int_{\mathscr{A}_k} \pi(y_k|\mathbf{x}_{k-1}^{(1:\nu_{k-1})}, z_k)\mathrm{d}y_k}$$

$y_k = h_k(x_k)$

Fixed efficient sample size adaptation

Straka O. and Šimandl M. 2006

Adaptive particle filter based on fixed efficient sample size

Proceedings of the 14th IFAC symposium on System Identification

IDEA: to preserve *Efficient sample size (ESS)* and to adapt sample size accordingly

- *Efficient sample size (ESS):* The ESS describes the number of samples drawn from the filtering pdf necessary to attain the same estimate quality as N_k samples drawn from the importance function.
- $$N_k = N_k^* \int \frac{[p(\mathbf{x}_k)]^2}{\pi(\mathbf{x}_k)} \mathrm{d}\mathbf{x}_k$$
- evaluating point estimate quality

Adaptive particle filter with fixed empirical density quality

Straka O. and Šimandl M. 2008

Adaptive particle filter with fixed empirical density quality

17th IFAC World Congress

IDEA: to guarantee the quality of an empirical pdf

- The quality is measured by inaccuracy (cross-information) between the empirical pdf and the filtering pdf
- with $N \to \infty$ inaccuracy converges to the Shannon differential entropy (SDE) of the filtering pdf
- the difference between the inaccuracy and the SDE

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

Adaptive particle filter with fixed empirical density quality

- $$N_k = z^2_{1-\delta/2} \frac{\sigma^2_W r^2_{1-\delta/2} - 2\text{cov}(Y,W) r_{1-\delta/2} + \sigma^2_Y}{(\mu_W r_{1-\delta/2} - \mu_Y)^2}$$
  with $z_{1-\delta/2}$ as $1-\delta/2$ quantile of the standard normal distribution
- N_k 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 $1-\delta$.

Computational aspects of the adaptation techniques

- Most of the relations for sample size adaptation depends on the true filtering pdf
- Naturally, the importance sampling technique is utilized for approximation of the filtering pdf

Convenient procedure:

1. set admissible sample sizes $[N_{min}, N_{max}]$
2. **Burn-in:** Draw N_{min} samples from the importance function and compute the corresponding weights, set N_{curr} to N_{min} and N_k to N_{max}
3. **Iterate:**
   - *while* $Ncurr < \min(N_k, N_{max})$
   - draw new N_Δ samples, compute the weights
   - set N_{curr} to $N_{curr} + N_\Delta$
   - update N_k using the new samples

Example I: Fixed efficient sample size adaptation

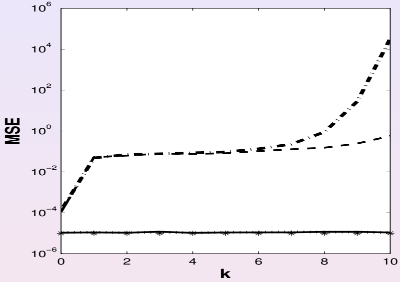
System

$$x_{k+1} = x_k - 0.2 \cdot x_k^2 + e_k \qquad p(e_k) = \mathcal{N}\{e_k : 0, 0.1\}$$
$$z_k = x_k + v_k \qquad p(v_k) = \mathcal{N}\{v_k : 0, 0.0001\}$$
$$p(x_0) = \mathcal{N}\{x_0 : 0, 0.001\}$$

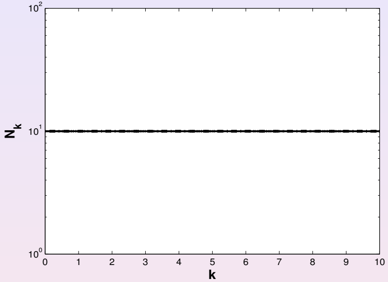
Particle filter

- importance function
  - optimal importance function
  - prior importance function
  - auxiliary importance function
  - filtering pdf
- $k = 0, 1, \ldots 10$, $N_k^* = 10$
- Criterion - MSE $\Pi_k = \mathrm{E}[x_k - \hat{x}_k]^2$, 1000 simulations

Example I: Results without sample size adaptation



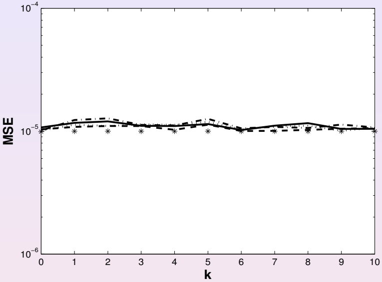
Mean square error



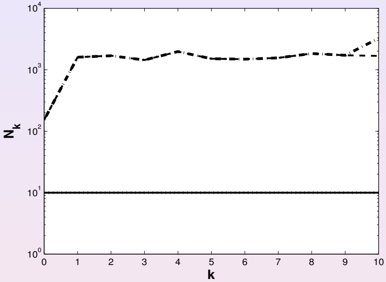
Sample size

solid -	**optimal** importance function
dashed -	**prior** importance function
dot-dashed -	**auxiliary** importance function
dotted -	**filtering** pdf
stars -	**Cramér Rao bound**

Example I: Results with sample size adaptation



Mean square error



Sample size

solid - **optimal** importance function
dashed - **prior** importance function
dot-dashed - **auxiliary** importance function
dotted - **filtering** pdf
stars - **Cramér Rao bound**

Example II: Adaptive particle filter with fixed empirical density quality

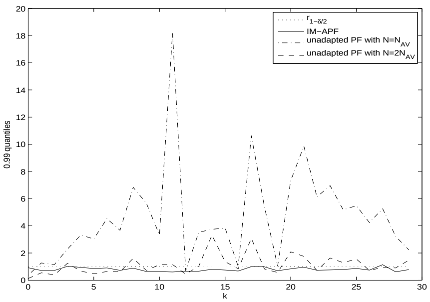
System

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

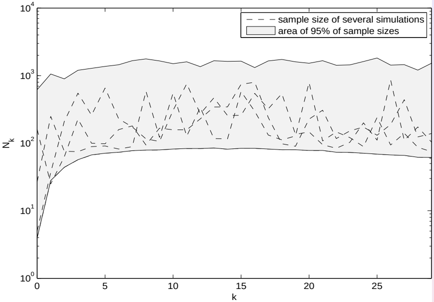
Particle filter

- 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 II: Results



0.99 quantiles of the difference between inaccuracy and SDE



Sample sizes of the IM-APF

solid - IM-APF

dot-dashed - unadapted PF with $N = N_{AV}$

dashed - unadapted PF with $N = 2N_{AV}$

dotted - $r_{1-\delta/2}$

Example II: 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
$\overline{\mathrm{var}(SE)}$	31.868	131.795	86.854

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

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

Concluding notes and remarks

- The problem of sample size setting requires close attention
- The techniques should take into account
  - System specification (determines posterior pdf of the state)
  - Importance function (determines samples location)
- The presented design principles – incomparable w.r.t estimate quality (different criteria)
- Operating conditions of the PF
  - Fixed duration of an estimation step
  - Possibility of incessant estimate quality improvement (sample as much as you can)

QUESTIONS ?