SAMPLING DENSITIES OF PARTICLE FILTER: A SURVEY AND COMPARISON

Miroslav Šimandl and Ondřej Straka

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

American Control Conference 2007

Outline

- Introduction
- Particle filter
  - State estimation problem
  - Particle filter algorithm
- Sampling densities - direct approach
- Sampling densities - composite approach
- Numerical comparison
- Conclusion

State estimation

Consider a discrete time stochastic system:

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

- $\mathbf{x}_k$ 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 $p(\mathbf{v}_k)$
- $\mathbf{f}_k(\mathbf{x}_k)$ and $\mathbf{h}_k(\mathbf{x}_k)$ are known vector functions

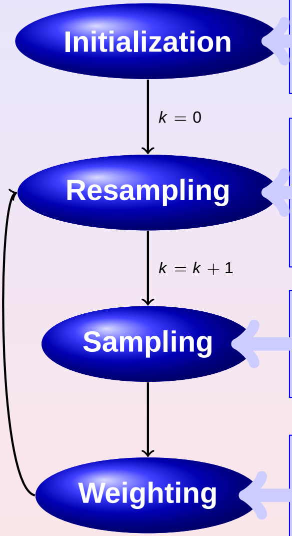
The aim of state estimation here is to find the filtering pdf $p(\mathbf{x}_k|\mathbf{z}^k)$, where $\mathbf{z}^k = [\mathbf{z}_0^{\mathrm{T}}, \ldots \mathbf{z}_k^{\mathrm{T}}]^{\mathrm{T}}$

Particle filter

- General solution of the filtering problem is given by the Bayesian Recursive Relations (BRR).
- Closed form solution of the BRR is available for a few special cases only (e.g. linear Gaussian systems).
- Thus an *approximate* solution of the BRR is usually searched.
- Solution of the BRR by the particle filter is based on approximating the filtering pdf by a set of samples (particles) and corresponding weights as

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

$\mathbf{x}_k^{(i)}$ - samples, $w_k^{(i)}$ - 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 $\{\mathbf{x}_0^{(i)}\}_{i=1}^N$ from $p(\mathbf{x}_0|\mathbf{z}^{-1})$ and set $w_0^{(i)} \propto p(\mathbf{z}_0|\mathbf{x}_0^{(i)})$

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

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

compute the weights

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

Sampling densities - general considerations

- Support of SD must contain support of the filtering pdf (importance sampling)
- Quality perspective: shape of SD must be as close to the filtering pdf as possible
- Implementation perspective: calculation of the weights must be as simple as possible

SD design techniques proceed within two approaches

Direct approach

- develops original concepts of the SD design
- proposes enhancements to the prior SD

Composite approach

- Utilization of another filtering technique - its filtering pdf is used as the SD

Direct approach - optimal and prior SD's

optimal SD (Zaritskii, Svetnik, Shimelevich 1975)

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

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

- ✔ minimizes variance of the weights
- ✔ weights can be computed in advance
- ✘ hard to find its explicit form

prior SD (Handshin 1970, Gordon et al. 1993)

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

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

- used in the famous Bootstrap filter
- ✘ $\mathbf{z}_k$ is ignored

Direct approach - fixed and auxiliary SD's

fixed SD (Tanizaki 1993)

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

$$\tilde{w}_k^{(i)} = \frac{p(\mathbf{z}_k|\mathbf{x}_k^{*(i)})p(\mathbf{x}_k^{*(i)}|\mathbf{x}_{k-1}^{(i)})}{\pi(\mathbf{x}_k^{*(i)})} w_{k-1}^{*(i)}$$

- ✗ sampling without information about the model

auxiliary SD (Pitt and Shephard 1999)

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

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

- primary weight $\lambda_k^{(i)}$ predicts quality of the sample $\mathbf{x}_k^{(i)}$ according to $\mathbf{z}_k$
- various approaches to primary weights calculation
- usually higher quality than prior SD

Direct approach - likelihood SD and others

likelihood SD (Chen 2003)

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

- ✘ a pdf of $\mathbf{x}_k$ must be derived from the likelihood (may not be possible)
- advantageous for measurement pdf tighter than transition pdf

other SD design techniques of the direct approach

- *gradient based prior SD* - sampling from prior SD a moving the sample using the gradient descent technique
- *hybrid SD* - combination of optimal and other SD's for multi-dimensional systems
- *bridging density sampling* - replaces single transition by a sequence of bridging densities placed between the initial density and the final density
- *partitioned sampling* - idea: to partition the state space and to apply the dynamics for each partition sequentially

Composite approach

- combination of the importance sampling technique with another nonlinear filtering method acting as a generator of SD
- the filtering methods include
  - *local methods* providing results valid in a small area in the state space
  - *global methods* providing results valid in almost whole state space
- utilization of another filtering method increases computational demands of particle filters

Composite approach - general scheme

at each time instant the approximate filtering pdf is given by

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

according to the Bayesian recursive relations

$$\hat{p}(\mathbf{x}_{k+1}|\mathbf{z}^k) = \int p(\mathbf{x}_{k+1}|\mathbf{x}_k) r_N(\mathbf{x}_k|\mathbf{z}^k) \mathrm{d}\mathbf{x}_k = \sum_{i=1}^{N} w_k^{(i)} p(\mathbf{x}_{k+1}|\mathbf{x}_k^{*(i)})$$

approximation of the filtering pdf given by

$$\hat{p}(\mathbf{x}_{k+1}|\mathbf{z}^{k+1}) = C^{-1} \hat{p}(\mathbf{x}_{k+1}|\mathbf{z}^k) p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})$$

is computed using the filtering method and used as a SD

Composite approach - filtering methods

Some of the filtering methods used a SD generator:

- Extended Kalman filter (e.g. De Freitas et al. 2000, Chen, Liu 2000)
- Gaussian sum filter (Kotecha, Djuric 2003)
  - preferable for multimodal pdf of the noises
- Sigma point Kalman filter (van der Merwe, Wan 2003)
- Gaussian mixture sigma point Kalman filter (van der Merwe, Wan 2003)
  - preferable for multimodal pdf of the noises
- H_∞ filter (Nishiyama 2005)

Numerical illustration

System - nonlinear, non-Gaussian

$$x_{k+1} = \varphi_1 \cdot x_k + 1 + \sin(\omega \pi k) + e_k \qquad p(e_k) = G\{e_k : 3, 2\}$$

$$z_k = \varphi_k \cdot x_k^2 + v_k \qquad p(v_k) = \mathcal{N}\{v_k : 0, 10^{-5}\}$$

$$p(x_0) = \mathcal{N}\{x_0 : 0, 12\}$$

$$\varphi_1 = 0.5,\ \varphi_2 = 0.2,\ \omega = 0.04$$

$$k = 0, 1, \ldots, 19$$

Note: variance of measurement noise ($\mathrm{var}[v_k] = 10^{-5}$) is smaller by six orders than variance of state noise ($\mathrm{var}[e_k] = 12$)

Sampling densities utilized

Direct approach

- prior (PSD)
- point auxiliary (PASD)
- unscented transformation auxiliary (UTASD)
- likelihood (LSD)

Composite approach

- extended Kalman filter based (EKFSD)
- Gaussian sum filter based (GSFSD)
- unscented Kalman filter based (UKFSD)
- Gaussian mixture unscented Kalman filter based (GMUKFSD)

Point estimates comparison

Criterion - mean square error

$$V_{MSE} = \frac{1}{KS} \sum_{k=1,s=1}^{K,S} (x_k(s) - \hat{x}_k(s))^2 , \quad \hat{x}_k(s) = \sum_{i=1}^{N} w_k^{(i)}(s) x_k^{(i)}(s)$$

Results for N=100 samples, S=1000 simulations

	PSD	PASD	UTASD	LSD
V_{MSE}	13.53	13.72	7.27	0.86

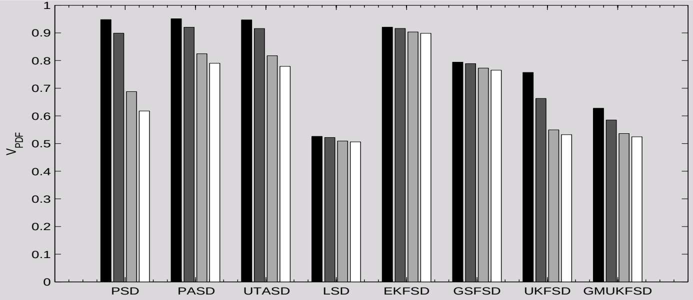
	EKFSD	GSFSD	UKFSD	GMUKFSD
V_{MSE}	2.06	1.28	1.49	1.28

Filtering PDF estimates comparison

Criterion

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

$V_{PDF} \in [0,1]$, $S = 1000$ MC simulations, $N = 50, 100, 500, 1000$ samples



Computational time

Length of a time step in the MATLAB environment, 3.2 GHz PC

	PSD	PASD	UTASD	LSD
T	0.0020	0.0020	0.0030	0.0080

	EKFSD	GSFSD	UKFSD	GMUKFSD
T	0.0075	0.0170	0.0060	0.0185

- SD's of the composite approach almost by an order more computational demanding than the SD's of the direct approach

Conclusion

Direct approach SD's

- generally, it is essential to choose SD according to system description (functions and pdf's of the noises)
- the **optimal** SD is the best choice but its explicit form is hard to be found
- **prior** SD is generic choice but usually provide low quality estimates
- **auxiliary** SD's a slightly better choice than the prior SD
- due to very accurate measurement in the example, the **likelihood** SD provided excellent results

Composite approach SD's

- SD's bring usually estimate quality increase paid by theoretical complexity
- high computational demands with respect to direct approach SD's

Conclusion

- for high-dimensional cases, hybrid SD may represent a profitable choice
- bridging density sampling improves degeneracy problem and accuracy of the samples at the expense of high computational burden

Thank you