PARZEN PARTICLE FILTERS

Tue Lehn-Schiøler

ISP, Technical University of Denmark
email: tls@imm.dtu.dk

Deniz Erdogmus & Jose C. Principe

CNEL, University of Florida
email: [deniz,principe]@cnel.ufl.edu

ABSTRACT

Using a Parzen density estimator any distribution can be approximated arbitrarily close by a sum of kernels. In particle filtering this fact is utilized to estimate a probability density function with Dirac delta kernels; when the distribution is discretized it becomes possible to solve an otherwise intractable integral. In this work we propose to extend the idea and use any kernel to approximate the distribution. The extra work involved in propagating small kernels through the nonlinear function can be made up for by decreasing the number of kernels needed, especially for high dimensional problems. A further advantage of using kernels with nonzero width is that the density estimate becomes continuous.

1. INTRODUCTION

The filtering problem can be formulated as

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{v}_{k-1} \qquad (1a)$$
$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{w}_k \qquad (1b)$$

where $\mathbf{v}$ and $\mathbf{w}$ are the process noise and the observation noise. The state transition density is fully specified by $\mathbf{f}$ and the process noise distribution and the observation likelihood is fully specified by $\mathbf{h}$ and the observation noise distribution.

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p_v(\mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1})) \qquad (2a)$$
$$p(\mathbf{z}_k|\mathbf{x}_k) = p_w(\mathbf{z}_k - \mathbf{h}(\mathbf{x}_k)) \qquad (2b)$$

The problem is to find an update formula from $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ to $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, where $\mathbf{z}_{1:k}$ denotes all observation $\{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$ up to time k. The Bayesian approach [1] gives the following update:

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

The problem can be broken down to two subproblems. 1) Find the propagation of a probability density function (pdf)

through a nonlinearity. 2) Modify the pdf according to the recorded measurements $\mathbf{z}_k$. Figure 1(a) illustrates stage one of the problem.

In equation (4) multiplying with $p(\mathbf{z}_k|\mathbf{x}_k)$ can be seen as stage two, and performing the multi-dimensional integration

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

as stage one. In general the integral can not be calculated analytically, hence, we need some way of estimating the distribution $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$.

Algorithms fall into four categories: Extended Kalman Filters, Gaussian Sum Filters, Sigma-Point Kalman Filters and Sequential Monte Carlo Methods (Particle Filters) [2]. Another way to categorize the methods is Gaussian belief (Extended Kalman filters, sigma point filters, moment matching), mixture of Gaussians (Gaussian-sum filter, pseudo-Bayes) and non-parametric methods (Particle filters) [3]. In the Extended Kalman Filter, the distributions are assumed Gaussian, but, the functions are not linear. The functions $\mathbf{f}$ and $\mathbf{h}$ are linearized around the previous state $\mathbf{x}_{k-1}$ using a second order Taylor expansion and then the standard Kalman equations are used. The result is a Gaussian distribution for $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ (see figure 1(b)). For nonlinear systems the solution is better than a normal Kalman filter, and it is accurate to first order. The Unscented Kalman Filter (Sigma Point Filter) [4] propagates points one standard deviation from the previous state $\mathbf{x}_{k-1}$ through the nonlinearity, then uses the points weighted appropriately (Gaussian quadrature like) to estimate mean and co-variance of a Gaussian. Finally this is used in the standard Kalman equations. It is accurate to the second order.

If the process noise distribution is approximated by a mixture of Gaussians the family of Gaussian sum filters arises [5]. In the mixture of Gaussians each mixture component is propagated through an extended Kalman filter. The state update $\mathbf{f}$ is linearized around the means of each mixture component and $\mathbf{h}$ is linearized around the predicted value for the mean of each mixture ($\mathbf{f}(x_{k-1})$). The resulting distribution is again a mixture of Gaussians. If the process noise is also non-Gaussian, this too can be approximated
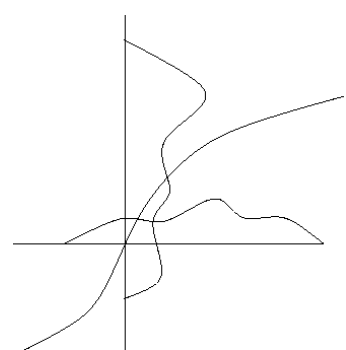
(a) A density propagated through a nonlinear function.
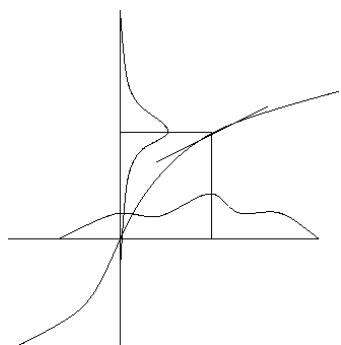
(b) In the extended Kalman filter, the distribution is assumed to be a Gaussian modified by a linearization of the nonlinearity around the previous state.
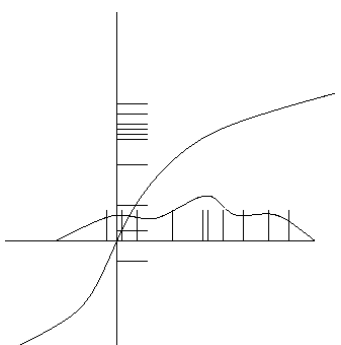
(c) In the particle filter the densities are approximated by discrete samples. These samples can be sent through the nonlinearity to give an estimate of the output density.

Fig. 1. Propagation of a pdf through a nonlinearity 1(a) and different approximations to the propagated distribution 1(b) and 1(c). This is the prediction step corresponding to equation (1a), the resulting pdf is then modified to match the measurements equation (1b).

with a mixture of Gaussians. However, in this case the number of mixing components increases quickly.

Nonparametric methods are an entirely different approach to nonlinear filtering. In the Particle Filter it is assumed that the distributions $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ and $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ from equation (4) can be estimated by discrete distributions (figure 1(c)). Samples are drawn from the posterior distribution using importance sampling and a proposal distribution. In the generic particle filter the transition probability $(p(\mathbf{x}_k|\mathbf{x}_{k-1}))$ is used as proposal, but other proposals has been proposed in e.g. the extended Kalman particle filter and the unscented particle filter [6], in these methods a filter (ekf or ukf) is calculated for each particle and the resulting mixture of Gaussians is used as proposal distribution for the particle filter. In an attempt to combine the particle filter and the Gaussian sum filter the Gaussian Sum Particle Filtering was proposed [7]. In this approach both the density and the process noise is considered mixture of Gaussians, in each time step samples are drawn from the mixture approximating $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$. These samples are propagated through the nonlinearity and used to offset the means in a mixture describing $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, then samples are drawn from this distribution too. In this way a discrete approximation of $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is obtained and the sample mean and covariance of the new mixtures can be estimated. Unfortunately the number of mixtures explode, to avoid this mixtures with small weight can be thrown away. In a similar manner, the Gaussian mixture Sigma Point Particle Filter [2] uses a bank of sigma point filters to update $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ then samples are drawn from the mixture and the importance weights are calculated before a Gaussian mixture is fitted to produce the posterior estimate.

In this paper an algorithm based on the Parzen density estimator is presented. The algorithm is best categorized as non-parametric, since it can be seen as a direct extension to the particle filter. The basic concept is to improve the performance of the particle filter by using a better density estimate.

The algorithm is similar to the Gaussian Sum Particle Filter and the Kernel Filter [8], however, it is derived in a different manner that allows use of any kernel type. The derivation of the algorithm uses a sample mean estimate of the integral $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and a particle filter like update of the weights. In section 2 the algorithm is derived and in section 3 experimental results are provided.

2. KERNEL METHOD

With a Parzen density estimator [9, 10] a distribution can be approximated arbitrarily close by a number of identical kernels centered on points chosen from the distribution. In the particle filter the kernels are delta functions, but information can be gained by using a broader kernel.

The distribution at time $k-1$ can be approximated by:
$p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) \approx \sum_i^N w_{k-1}^i K(\mathbf{A}_{k-1}^i(\mathbf{x}_{k-1}-\mathbf{x}_{k-1}^i))$, where $\mathbf{A}^i$ is a transformation matrix used to keep track of distortions of the kernel. Each kernel can be propagated through the mapping $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ by using a local linearization, yielding a continuous output distribution $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, this is again a sum of kernels but the kernels are no longer identical (in the sense that they are from the same family of functions, yet they have different parameters).

Using the kernel representation equation (4) can be written as:

$$\sum_i^N w_{k-1}^i \int p_v(\mathbf{x}_k - f(\mathbf{x}_{k-1})) K(\mathbf{A}_{k-1}^i(\mathbf{x}_{k-1}-\mathbf{x}_{k-1}^i))\mathrm{d}\mathbf{x}_{k-1}$$
(5)

Each part of the sum can be handled individually, and under the assumption that the kernels are small compared to the dynamic in the nonlinearity, $\mathbf{f}$ can be locally linearized. By linearizing $\mathbf{f}$ around $\mathbf{x}_{k-1}^i$ the jacobian $\mathbf{J}|_{\mathbf{x}_{k-1}^i} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}|_{\mathbf{x}_{k-1}^i}$ is introduced and the following change of variables can be employed: $\hat{\mathbf{x}}_{k-1} = \mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1}^i) - \mathbf{J}|_{\mathbf{x}_{k-1}^i}(\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^i)$. Inserting this in the integral from equation (5) yields:

$$\left|\mathbf{J}|_{\mathbf{x}_{k-1}^i}\right|^{-1} \int [p_v(\hat{\mathbf{x}}_{k-1}) \qquad (6)$$
$$K \quad \left(\mathbf{A}_{k-1}^i\mathbf{J}|_{\mathbf{x}_{k-1}^i}^{-1}(\mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1}^i) - \hat{\mathbf{x}}_{k-1})\right) \quad]\,\mathrm{d}\hat{\mathbf{x}}_{k-1}$$

This integral is an expectation over the process noise $E_{p_v}\left[K\left(\mathbf{A}_{k-1}^i\mathbf{J}|_{\mathbf{x}_{k-1}^i}^{-1}(\mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1}^i) - \hat{\mathbf{x}}_{k-1})\right)\right]$ and can be approximated by a sample mean. In the extreme case a single sample drawn from p_v can be used, and the result is a translation of the kernel by the noise sample:

$$E_{p_v}\left[K\left(\mathbf{A}_{k-1}^i\mathbf{J}|_{\mathbf{x}_{k-1}^i}^{-1}(\mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1}^i) - \hat{\mathbf{x}}_{k-1})\right)\right]$$
$$\approx \quad K\left(\mathbf{A}_{k-1}^i\mathbf{J}|_{\mathbf{x}_{k-1}^i}^{-1}(\mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1}^i) - \mathbf{v}_{k-1})\right), \mathbf{v}_{k-1} \sim p_v$$

Writing $p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \sum_i^N w_k^i K(\mathbf{A}_k^i(\mathbf{x}_k - \mathbf{x}_k^i))$ It is possible to identify the centers of the kernels $\mathbf{x}_k^i = \mathbf{f}(\mathbf{x}_{k-1}^i) + \mathbf{v}_{k-1}$ and the transformation matrix $\mathbf{A}_k^i = \mathbf{A}_{k-1}^i\mathbf{J}|_{\mathbf{x}_{k-1}^i}^{-1}$. By considering equation (4), (5) and (6) the weight update can be found to be $w_k^i = w_{k-1}^i p(\mathbf{z}_k|\mathbf{x}_k^i)\left|\mathbf{J}|_{\mathbf{x}_{k-1}^i}\right|^{-1}$. This derivation holds for any kernel, however, for simplicity, in this paper the kernels are considered Gaussian.

For a Gaussian a change of variables can be employed such that the update of $\mathbf{A}_k^i$ can be replaced with an update of the covariance matrix as follows:

$$\mathbf{A}_k^i = \mathbf{A}_{k-1}^i \qquad (7a)$$
$$\boldsymbol{\Sigma}_k^i = \mathbf{J}|_{\mathbf{x}_{k-1}^i}\boldsymbol{\Sigma}_{k-1}^i\mathbf{J}|_{\mathbf{x}_{k-1}^i}^T \qquad (7b)$$

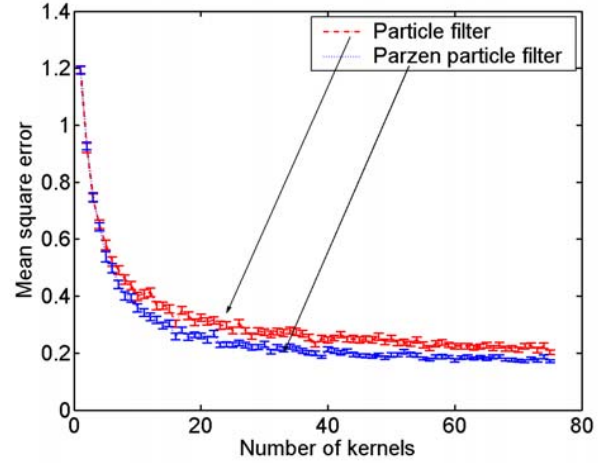The transformation matrix is absorbed in the covariance matrix.

Fig. 2. Problem with nonlinear state transition, nonlinear observation process and Gaussian noise. Note that the performance of a Parzen particle filter with ≈ 10 kernels equals that of a normal particle filter with ≈ 20 kernels.

The transformation matrix $\mathbf{A}$ (or $\boldsymbol{\Sigma}$ in the Gaussian case) is distorted in each iteration, to avoid to much distortion a resampling schema can be applied. With a suitable frequency the distribution can be re-approximated by a Parzen estimator by drawing samples from $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, choosing $\mathbf{A}$ or $\boldsymbol{\Sigma}$ to take their initial values and setting the weights equal.

Earlier attempts use the kernels in the resampling phase where the shape and kernel size are selected based on the particle statistics (e.g. covariance) [11] . However, the proposed method iterates these properties of the kernel through the system equations, thus there is no need for optimization of kernel parameters at every step. In addition, the approximation of the integral stochastically using a sample drawn from p_v includes an inherent resampling step at every iteration, which allows the particle filter accuracy to survive longer than the standard version.

3. EXPERIMENTS

In this section the performance of the Parzen particle filter will be compared to the performance of the standard particle filter[1]. The method is tested on a one dimensional problem:

$$x_k = \frac{x_{k-1}}{2} + 25\frac{x_{k-1}}{(1+x_{k-1}^2)} + 8\cos(1.2k) + v_k \qquad (8a)$$
$$z_k = 10\arctan(\frac{x_k}{10}) + w_k \qquad (8b)$$

Where v_k and w_k are drawn from Gaussian distributions $G(0,1)$ (figure 2) and from gamma distributions $\Gamma(3,2)$ (fig-

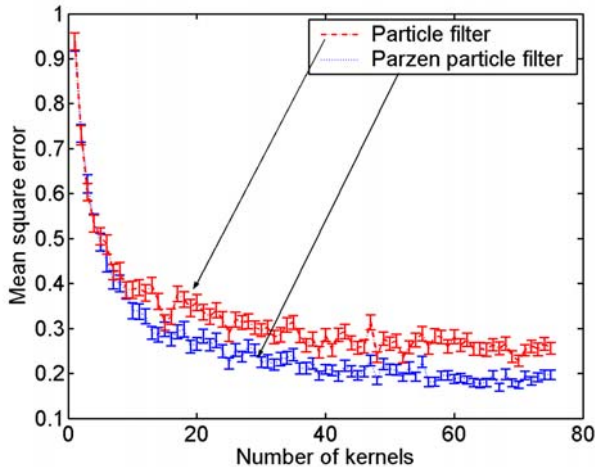[1]Code to reproduce the results can be found at www.imm.dtu.dk/~tls

Fig. 3. Problem with nonlinear state transition, nonlinear observation process and gamma distributed noise.

ure 3).

The Parzen particle filter and the generic particle filter has been used on 100 time series generated using equation (8).

In figure 2 the mean square error is plotted as a function of the number of kernels. It can be seen that with few kernels the methods perform equally good (or bad), but as the number of kernels increases the kernel method becomes better. It can be seen that for this one dimensional example the kernel methods perform equally well, but the number of particles can be reduced drastically by improving the density estimate. It is expected that this effect will be even more impressive in higher dimensional problems.

4. CONCLUSION

A novel algorithm for nonlinear filtering is presented, the algorithm is based on Parzen density estimates and particle filter like propagation of the kernel through local linearizations of the nonlinear function.

It is shown that the improved density estimate help performance both with Gaussian and non-Gaussian noise. In this work only the special case with a Gaussian kernel is examined, however it is expected that a broader kernel would be well suited for long tailed noise, since it will be more likely to get the particles spread out.

The basic formulation for the arbitrary kernel case has been derived and performances of various kernel choices will be compared in a future publication.

5. REFERENCES

[1] N. Gordon, D. Salmond, and A. F. M. Smith, "Novel approach to non-linear and non-gaussian bayesian state estimation," *IEE Proceedings-F*, vol. 140, pp. 107–113, 1993.

[2] R. van der Merwe and E. Wan, "Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models," *Speech and Signal Processing (ICASSP)*, April 2003.

[3] T. Minka, "Bayesian inference in dynamic models – an overview, www.stat.cmu.edu/∼ minka/dynamic.html," 2003.

[4] S. J. Julier and J. Uhlmann, "A new extension of the kalman filter to nonlinear systems," 1997.

[5] D. L. Alspach and H. W. Sorensen, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, August 1972.

[6] R. van der Merwe, N. de Freitas, A. Doucet, and E. Wan, "The unscented particel filter," *Advances in Neural Information Processing Systems 13*, Nov 2001.

[7] J.H. Kotecha and P.M. Djuric, "Gaussian sum particle filtering for dynamic state space models," *Acoustics, Speech, and Signal Processing, 2001. Proceedings. 2001 IEEE International Conference on*, vol. 6, pp. 3465 –3468 vol.6, 2001.

[8] M. Hurzeler and H.R. Kunsch, "Monte carlo approximations for general state space models," *Journal of Computational and Graphical Statistics*, vol. 7, no. 2, pp. 175–193, 1998.

[9] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat*, vol. 27, pp. 1065–1076, 1962.

[10] L. Devroye and G. Lugosi, *Combinatorial Methods in Density Estimation*, Springer, New York, 2001.

[11] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.