Quasi-Gaussian Particle Filtering*

Yuanxin Wu, Dewen Hu, Meiping Wu, and Xiaoping Hu

Department of Automatic Control, College of Mechatronics and Automation,
National University of Defense Technology, Changsha, Hunan, P.R. China, 410073
yuanx_wu@hotmail.com

Abstract. The recently-raised Gaussian particle filtering (GPF) introduced the idea of Bayesian sampling into Gaussian filters. This note proposes to generalize the GPF by further relaxing the Gaussian restriction on the prior probability. Allowing the non-Gaussianity of the prior probability, the generalized GPF is provably superior to the original one. Numerical results show that better performance is obtained with considerably reduced computational burden.

1 Introduction

The Bayesian probabilistic inference provides an optimal solution framework for dynamic state estimation problems [1, 2]. The Bayesian solution requires propagating the full probability density function, so in general the optimal nonlinear filtering is analytically intractable. Approximations are therefore necessary, e.g., Gaussian approximation to the probability [3-9]. This class of filters is commonly called as the Gaussian filters, in which the probability of interest, e.g. the prior and posterior probabilities, are approximated by Gaussian distribution. An exception is the so-called augmented unscented Kalman filter [10] where the prior probability is encoded by the nonlinearly transformed deterministic sigma points instead of by the calculated mean and covariance from them. By so doing, the odd-order moment information is captured and propagated throughout the filtering recursion, which helps improve the estimation accuracy. This note will show that similar idea can be applied to the recently-raised Gaussian particle filtering (GPF) [7].

The GPF was developed using the idea of Bayesian sampling under the Gaussian assumption [7]. It actually extends the conventionally analytical Gaussian filters via Monte Carlo integration and the Bayesian update rule [11]. The Gaussian assumption being valid, the GPF is asymptotically optimal in the number of random samples, which means that equipped with the computational ability to handle a large number of samples the GPF is supposed to outperform any analytical Gaussian filter. The GPF also have a lower numerical complexity than particle filters [7].

This work generalizes the GPF by relaxing the assumption of the prior probability being Gaussian. Since the prior probability is allowed to be non-Gaussian, the resulting filter is named as the quasi-Gaussian particle filtering (qGPF) in the sequel. It turns out

* Supported in part by National Natural Science Foundation of China (60374006, 60234030 and 30370416), Distinguished Young Scholars Fund of China (60225015), and Ministry of Education of China (TRAPOYT Project).

V.N. Alexandrov et al. (Eds.): ICCS 2006, Part I, LNCS 3991, pp. 689 – 696, 2006.

that the qGPF outperforms the GPF in both accuracy and computational burden. The contents are organized as follows. Beginning with the general Bayesian inference, Section II derives and outlines the qGPF algorithm. Section III examines two representative examples and the conclusions are drawn in Section IV.

2 Quasi-Gaussian Particle Filtering

Consider a discrete-time nonlinear system written in the form of dynamic state space model as

$$\begin{aligned} x_k &= f_{k-1}\left(x_{k-1}, w_{k-1}\right) \\ y_k &= h_k\left(x_k, v_k\right) \end{aligned} \tag{1}$$

where the process function $f_k : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ and observation function $h_k : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$ are some known functions. The process noise $w_k \in \mathbb{R}^r$ is uncorrelated with the past and current system states; the measurement noise $v_k \in \mathbb{R}^s$ is uncorrelated with the system state and the process noise at all time instants. The probabilities of the process and measurement noises are both assumed to be known.

Denote by $y_{1:k} \triangleq \{y_1, \ldots, y_k\}$ the observations up to time instant k. The purpose of filtering is to recursively estimate the posterior probability $p\left(x_k \mid y_{1:k}\right)$ conditioned on all currently available but noisy observations. The initial probability of the state is assumed to be $p\left(x_0 \mid y_{1:0}\right) \equiv p\left(x_0\right)$. The prior probability is obtained via the Chapman-Kolmogorov equation

$$p\left(x_k \mid y_{1:k-1}\right) = \int_{\mathbb{R}^n} p\left(x_k \mid x_{k-1}\right) p\left(x_{k-1} \mid y_{1:k-1}\right) dx_{k-1}. \tag{2}$$

The transition probability density $p\left(x_k \mid x_{k-1}\right)$ is uniquely determined by the known process function and the process noise probability. Using the Bayesian rule, the posterior probability is given by

$$p\left(x_k \mid y_{1:k}\right) = \frac{p\left(y_k \mid x_k\right) p\left(x_k \mid y_{1:k-1}\right)}{p\left(y_k \mid y_{1:k-1}\right)} \tag{3}$$

where the normalizing constant

$$p\left(y_k \mid y_{1:k-1}\right) = \int_{\mathbb{R}^n} p\left(y_k \mid x_k\right) p\left(x_k \mid y_{1:k-1}\right) dx_k. \tag{4}$$

The likelihood probability density $p\left(y_k \mid x_k\right)$ is uniquely determined by the known observation function and the measurement noise probability. Equations (2)-(4) constitute the foundation of the optimal Bayesian probabilistic inference. Unfortunately, the exact analytic form only exists for a couple of special cases, e.g., when the system (1) is linear and Gaussian. In order to make the filtering problem tractable, approximation must be made.

Next, we start to derive the qGPF by assuming the posterior probability at time instant $k-1$ to be well approximated by a Gaussian distribution, i.e.,

$$p\left(x_{k-1} \mid y_{1:k-1}\right) \approx \mathcal{N}\left(x_{k-1}; m_{k-1}, P_{k-1}\right) \tag{5}$$

Substituting (5) and using the Monte-Carlo integration [12], the prior probability in (2) is

$$\begin{aligned} p\left(x_k \mid y_{1:k-1}\right) &\approx \int_{\mathbb{R}^n} p\left(x_k \mid x_{k-1}\right) \mathcal{N}\left(x_{k-1}; m_{k-1}, P_{k-1}\right) dx_{k-1} \\ &= \int_{\mathbb{R}^n} p\left(x_k \mid x_{k-1}\right) \sum_{i=1}^{M_1} \frac{1}{M_1} \delta\left(x_{k-1} - x_{k-1}^i\right) dx_{k-1} = \frac{1}{M_1} \sum_{i=1}^{M_1} p\left(x_k \mid x_{k-1}^i\right) \end{aligned} \tag{6}$$

where x_{k-1}^i are random samples from the assumed posterior probability at time instant $k-1$, i.e., $\mathcal{N}\left(x_{k-1}^i; m_{k-1}, P_{k-1}\right)$, $i = 1, \ldots, M_1$. The idea of the importance sampling [13, 14] is crucial to numerically implement the Bayesian rule, through which the prior probability is updated to yield the posterior probability using the information provided by the newcome observation. In view of the difficulty of drawing samples directly from the posterior probability $p\left(x_k \mid y_{1:k}\right)$, the importance sampling proposes to sample from a choice importance density $q\left(x_k \mid y_{1:k}\right)$ instead, from which random samples can be readily generated. With (6), the posterior probability in (3) is rewritten as

$$\begin{aligned} p\left(x_k \mid y_{1:k}\right) &\propto p\left(y_k \mid x_k\right) p\left(x_k \mid y_{1:k-1}\right) \propto \frac{p\left(y_k \mid x_k\right) \sum_{i=1}^{M_1} p\left(x_k \mid x_{k-1}^i\right)}{q\left(x_k \mid y_{1:k}\right)} q\left(x_k \mid y_{1:k}\right) \\ &\propto \sum_{j=1}^{M_2} \frac{p\left(y_k \mid x_k^j\right) \sum_{i=1}^{M_1} p\left(x_k^j \mid x_{k-1}^i\right)}{q\left(x_k^j \mid y_{1:k}\right)} \delta\left(x_k - x_k^j\right) \triangleq \sum_{j=1}^{M_2} \hat{w}_k^j \delta\left(x_k - x_k^j\right) \end{aligned} \tag{7}$$

where x_k^j are random samples from the importance density $q\left(x_k \mid y_{1:k}\right)$ and

$$\hat{w}_k^j = \frac{p\left(y_k \mid x_k^j\right) \sum_{i=1}^{M_1} p\left(x_k^j \mid x_{k-1}^i\right)}{q\left(x_k^j \mid y_{1:k}\right)}, \quad j = 1, \ldots, M_2. \tag{8}$$

Considering the normalization condition, the posterior probability at time instant k is approximated by

$$p\left(x_k \mid y_{1:k}\right) = \sum_{j=1}^{M_2} w_k^j \delta\left(x_k - x_k^j\right) \tag{9}$$

where

$$w_k^j = \hat{w}_k^j \Big/ \sum_{j=1}^{M_2} \hat{w}_k^j . \tag{10}$$

Table 1. Quasi-Gaussian Particle Filtering

1. Draw samples from the posterior probability at time instant $k-1$, i.e., $x_{k-1}^i \sim p\left(x_{k-1} \mid y_{1:k-1}\right) \approx \mathcal{N}\left(x_{k-1}; m_{k-1}, P_{k-1}\right)$, $i = 1,\ldots,M_1$;
2. Draw samples from the important density, that is, $x_k^j \sim q\left(x_k \mid y_{1:k}\right)$, $j = 1,\ldots,M_2$;
3. Assign each sample x_k^j a weight w_k^j according to (8) and (10);
4. Calculate the mean and covariance according to (11), then $p\left(x_k \mid y_{1:k}\right) \approx \mathcal{N}\left(x_k; m_k, P_k\right)$.

Then approximate the posterior probability at time instant k by $\mathcal{N}\left(x_k; m_k, P_k\right)$ in which

$$m_k = \sum_{j=1}^{M_2} w_k^j x_k^j, \quad P_k = \sum_{j=1}^{M_2} w_k^j \left(x_k^j - m_k\right)\left(x_k^j - m_k\right)^T. \tag{11}$$

This ends the derivation and the resulting algorithm is summarized and outlined in Table I. It is clear from above that we only assume the posterior probability to be Gaussian while do not impose any restriction on the prior probability, which is the major difference from the GPF. Recall that the GPF approximates both the prior and posterior probabilities by Gaussian densities. To be more specific, the GPF approximates the discrete representation of the prior probability $p\left(x_k \mid y_{1:k-1}\right)$ by a Gaussian density, from which random samples are regenerated to be weighted by the likelihood $p\left(y_k \mid x_k\right)$, while the qGPF directly employs the discrete representation of the prior probability. This resembles the difference between the non-augmented UKF and augmented UKF [10]. By a peer-to-peer comparison between the qGPF and the GPF ([7], Table I), we see that the qGPF needs not to calculate the sample mean and covariance for the assumed Gaussian prior probability and thus has lower numerical complexity.

The following theorem says that the mean and covariance in (11) converge almost surely to the true values under the condition that the posterior probability at time instant $k-1$ is well approximated by a Gaussian distribution.

Theorem: If the posterior probability $p\left(x_{k-1} \mid y_{1:k-1}\right)$ is a Gaussian distribution, then the posterior probability expressed in (9) converges almost surely to the true posterior probability $p\left(x_k \mid y_{1:k}\right)$.

Proof: it is a very straightforward extension of Theorem 1 in [7] and thus omitted here.

It follows as a natural corollary that the mean and covariance in (11) converge almost surely to the true value. Therefore the qGPF is provably better than the GPF in accuracy because the former takes the non-Gaussianity of the prior probability into consideration. Note that the non-Gaussianity of the prior probability is not uncommon for nonlinear/non-Gaussian systems.

In theory, we could assume the posterior probability to be any other distribution, as long as the samples from the distribution were easily obtained, e.g. mixed Gaussian [15-17]. The derivation procedure and theoretical proof would be analogical to the above.

3 Numerical Results

This section examines the qGPF via the univariate nonstationary growth model and bearing only tracking, which have been extensively investigated in the literature [2, 7, 18, 19]. We also carried out the GPF for comparison. The prior probability was selected as the importance density for both filters, i.e., $q\left(x_k \mid y_{1:k}\right) = p\left(x_k \mid y_{1:k-1}\right)$.

Univariate Nonstationary Growth Model
The model is formulated as

$$\begin{aligned} x_k &= f_{k-1}\left(x_{k-1}, k\right) + w_{k-1} \\ y_k &= h_k\left(x_k\right) + v_k, \quad k = 1, \ldots, N \end{aligned} \tag{12}$$

where $f_{k-1}\left(x_{k-1}, k\right) = 0.5x_{k-1} + 25\frac{x_{k-1}}{1+x_{k-1}^2} + 8\cos\left(1.2\left(k-1\right)\right)$, $h_k\left(x_k\right) = x_k^2/20$. The process noise w_{k-1} and measurement noise v_k are zero-mean Gaussian with variances Q_{k-1} and R_k, respectively. In our simulation, $Q_{k-1} = 10$ and $R_k = 1$. This model has significant nonlinearity and is bimodal in nature depending on the sign of observations. The reference data were generated using $x_0 = 0.1$ and $N = 100$. The initial probability $p\left(x_0\right) \sim \mathcal{N}\left(0,1\right)$.

The mean square error (MSE) averaged across all time instants defined as $\text{MSE} = \sum_{k=1}^{N}\left(x_k - x_{k|k}\right)^2 \Big/ N$ is used to quantitatively evaluate each filter. We carried out 50 Monte Carlo runs for $M_1 = M_2 = 20, 50, 100, 200, 400$, respectively. Figure 1 shows MSEs as a function of the number of samples. The qGPF remarkably outperforms the GPF. With the same number of samples, MSE of the qGPF is less than half of that of the GPF; on the other hand, to achieve comparable performance the GPF needs at least as twice samples as the qGPF does. The average running time of the qGPF is about 20 percent less than that of the GPF.

Bearing Only Tracking
The target moves within the $s-t$ plane according to the standard second-order model

$$x_k = \Phi x_{k-1} + \Gamma w_{k-1}, \qquad k = 1, \ldots, N \tag{13}$$

where $x_k = \left[s, \dot{s}, t, \dot{t}\right]_k^T$, $w_k = \left[w_s, w_t\right]_k^T$,

$$\Phi = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \Gamma = \begin{bmatrix} 0.5 & 0 \\ 1 & 0 \\ 0 & 0.5 \\ 0 & 1 \end{bmatrix}.$$

Here s and t denote Cartesian coordinates of the moving target. The system noise $w_k \sim \mathcal{N}(0, QI_2)$. A fixed observer at the origin of the plane takes noisy measurements of the target bearing

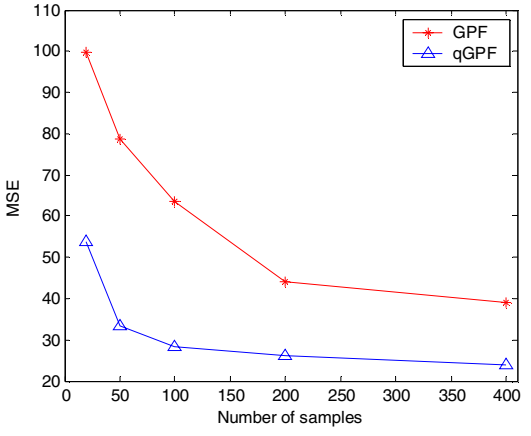

Fig. 1. MSE as a function of the number of samples for both filters

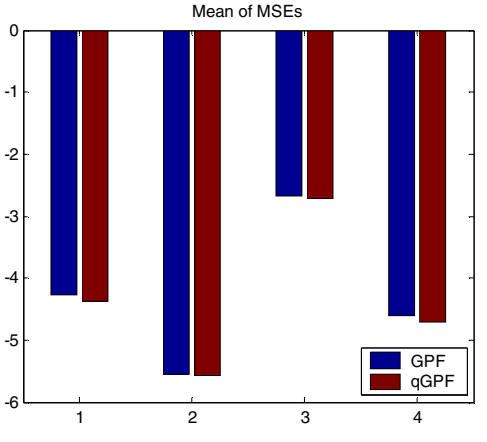

Fig. 2. Averaged MSE of all four coordinates (logarithmic in y axis)

$$y_k = \arctan\left(t_k / s_k\right) + v_k \tag{14}$$

where the measurement noise $v_k \sim \mathcal{N}(0, R)$. The reference data were generated using $Q = 0.001^2$, $R = 0.005^2$ and $N = 24$. The initial true state of the system was $x_0 = [-0.05, 0.001, 0.7, -0.055]^T$ and the initial estimate was $x_{0|0} = x_0$ with covariance $P_{0|0} = \mathrm{diag}\left(\left[0.1^2, 0.005^2, 0.1^2, 0.01^2\right]^T\right)$.

We carried out 100 random runs for $M_1 = M_2 = 1000$ and the averaged MSEs of all four coordinates are given in Fig. 2. We see that the qGPF is smaller in MSE, though marginally, than the GPF. Similar observations were obtained for various number of samples and is omitted here for brevity. In the simulation, eighteen percent of computational time was spared by using the qGPF.

4 Conclusions

This note proposes the qGPF filter that generalizes the GPF by allowing the prior probability to be non-Gaussian. It has provable superiority over the GPF. The numerical results show that the qGPF achieves (sometimes remarkably) better improvement in estimation accuracy with lower numerical complexity than the GPF. Theoretically, the posterior probability could be assumed to be any other distribution as long as it was readily sampled, e.g., mixed Gaussian. In such a case, it is promising for the qGPF to be used to construct more superior filter than the GPF-based Gaussian sum particle filter in [17].

References

[1] Y. C. Ho and R. C. K. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, vol. AC-9, pp. 333-339, 1964.

[2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, 2002.

[3] K. Ito and K. Q. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910-927, 2000.

[4] A. H. Jazwinski, *Stochastic Processing and Filtering Theory*. New York and London: Academic Press, 1970.

[5] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in *Signal Processing, Sensor Fusion, and Target Recognition VI*, vol. 3068, *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, 1997, pp. 182-193.

[6] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 34-45, 1960.

[7] J. H. Kotecha and P. A. Djuric, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592-2601, 2003.

[8] M. Norgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, no. 11, pp. 1627-1638, 2000.

[9] Y. Wu, D. Hu, M. Wu, and X. Hu, "A Numerical-Integration Perspective on Gaussian Filters," *IEEE Transactions on Signal Processing, to appear*, 2005.
[10] Y. Wu, D. Hu, M. Wu, and X. Hu, "Unscented Kalman Filtering for Additive Noise Case: Augmented versus Non-augmented," *IEEE Signal Processing Letters*, vol. 12, no. 5, pp. 357-360, 2005.
[11] Y. Wu, X. Hu, D. Hu, and M. Wu, "Comments on "Gaussian particle filtering"," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3350-3351, 2005.
[12] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*: New York, Academic Press, 1975.
[13] A. F. M. Smith and A. E. Gelfand, "Bayesian statistics without tears: a sampling-resamping perspective," *The American Statistician*, vol. 46, no. 2, pp. 84-88, 1992.
[14] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*. New York: Wiley, 1981.
[15] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465-479, 1971.
[16] D. L. Alspach and H. W. Sorenso, "Nonlinear Bayesian estimation using Gaussian sum approximation," *IEEE Transactions on Automatic Control*, vol. 17, pp. 439-448, 1972.
[17] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2602-2612, 2003.
[18] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time-series," *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1032-1063, 1987.
[19] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings-F*, vol. 140, no. 2, pp. 107-113, 1993.