A Hybrid Importance Function for Particle Filtering

Yufei Huang, *Member, IEEE,* and Petar M. Djurić, *Senior Member, IEEE*

Abstract—Particle filtering has drawn much attention in recent years due to its capacity to handle nonlinear and non-Gaussian dynamic problems. One crucial issue in particle filtering is the selection of the importance function that generates the particles. In this letter, we propose a new type of importance function that possesses the advantages of the posterior and the prior importance functions. We demonstrate its use on the problem of blind detection in flat fading channels and provide simulation results that show its efficiency and performance.

Index Terms—Bind detection, non-Gaussian, , nonlinear, particle filtering, sequential signal processing.

I. INTRODUCTION

IN RECENT years, particle filters [1] have drawn much attention in adaptive processing of nonlinear and non-Gaussian systems. Problems pertaining to particle filtering are concerned with making inferences in a sequential fashion on $\mathbf{x}_t$, the unknowns at time t, based on $\mathbf{y}_{0:t}$, the observations from time 0 to t. For nonlinear and non-Gaussian systems, the optimum solutions require complex high-dimensional integrations, thereby imposing great difficulties for the conventional approaches. Based on the concept of sequential importance sampling [1], particle filters can approximate the posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{0:t})$ regardless of the linearity and Gaussianity of the underlying model. Consequently, they have become a prominent algorithm for adaptive processing of complex systems [1].

A crucial issue in particle filtering is the selection of the importance function. Two standard choices of importance functions are the posterior and the prior. The posterior importance function is defined as $q(\mathbf{x}_t|\mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t}) = p(\mathbf{x}_t|\mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t})$, and the importance weights of the particles are obtained from

$$w_t^{(j)} \propto w_{t-1}^{(j)} p\left(\mathbf{y}_t|\mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1}\right) \tag{1}$$

where the superscript j denotes the j-th trajectory. The posterior importance function is optimal in the sense that it minimizes the variance of the importance weights. As a consequence, better particles can be generated at time instant $t + 1$ and better estimates produced at t. However, a major difficulty in its use is the calculation of the weights since analytical evaluation of $p(\mathbf{y}_t|\mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1})$ is required, which involves complex high-dimensional integrations over $\mathbf{x}_t$. In addition, direct sampling

Y. Huang is with the Department of Electrical Engineering, University of Texas at San Antonio, San Antonio, TX 78249-0669 USA (e-mail: yhuang@utsa.edu).

P. M. Djurić is with the Department of Electrical and Computer Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794 USA (e-mail: djuric@ece.sunysb.edu).

from the posterior importance function may also be prohibited. These difficulties prevent the posterior importance function from being widely used.

Therefore, most frequently it is the prior importance function that is used. The prior importance function is defined as $q(\mathbf{x}_t|\mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t}) = p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(j)})$ and the weights associated are computed according to

$$w_t^{(j)} \propto w_{t-1}^{(j)} p\left(\mathbf{y}_t|\mathbf{x}_t^{(j)}, \mathbf{y}_{0:t-1}\right). \tag{2}$$

Compared with the posterior importance function, it is attractive due to its simplicity in sampling from the prior densities and the calculation of weights. Despite its simplicity, the prior importance function is very inefficient. Since no information from the observations is used, the generated particles often come from the tails of the posterior distributions, and as a result, the weights have large variations and the estimation results are poor. Improving stratagems including the auxiliary particle filter have been proposed [2].

In this letter, we look beyond the above two choices and propose a new type of importance function, which is a hybrid of the aforementioned importance functions. Since the proposed importance function employs most recent observations, it generates better sets of particles than the prior importance function, and that altogether leads to more efficient particle filters. Equally important, the proposed importance function is less restrictive than the posterior function and is thus applicable to a wider range of problems.

II. HYBRID IMPORTANCE FUNCTION

Consider a situation where the use of the posterior importance function is extremely difficult. In the cases analyzed here, we assume that the state parameters can be divided into two independent parts, e.g., $\mathbf{x}_t = \{\mathbf{x}_{1t}, \mathbf{x}_{2t}\}$, where sampling from $p(\mathbf{x}_{2t}|\mathbf{x}_{2,t-1}^{(j)})$ and $p(\mathbf{x}_{1t}|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t})$ can be carried out easily. For such cases, we propose to use the following importance function

$$q\left(\mathbf{x}_t|\mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t}\right) = p\left(\mathbf{x}_{1t}|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t}\right) p\left(\mathbf{x}_{2t}|\mathbf{x}_{2,t-1}^{(j)}\right) \tag{3}$$

where $\mathbf{x}_{2t}^{(j)}$ is a sample from $p(\mathbf{x}_{2t}|\mathbf{x}_{2,t-1}^{(j)})$. Apparently, (3) is a hybrid between the posterior and the prior importance functions. The computation of the weights of the particles generated by the hybrid importance function is carried out as in (4), shown at the bottom of the next page. We note from (4) that the distribution $p(\mathbf{y}_t|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1})$ is critical in the computation of the weights, and that therefore, its analytical form should be available.

The advantage of the proposed hybrid importance function over the posterior importance function is in the easy updating of the weights. In addition, since the hybrid importance function

includes information from the observations, it generates samples with smaller variance on the weights than the prior importance function.

The key to the applicability of this importance function is the assumption of possible sampling from $p(\mathbf{x}_{1t}|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t})$ and knowing the analytical form of $p(\mathbf{y}_t|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1})$ (including the normalizing constant). Cases that fall within the assumption, for instance, are the ones where given $\mathbf{x}_{2t}$, $\mathbf{y}_t$ is linear in $\mathbf{x}_{1t}$. It should be noted that, although splitting states is also seen in Rao–Blackwellised filters [3], [4] such as the mixture Kalman filter [5], [6], the purpose of splitting in the hybrid importance function is substantially different. In Rao–Blackwellised filters, $\mathbf{x}_{1t}$ are nuisance states, and Rao–Blackwellised filters provide the possibility to marginalize $\mathbf{x}_{1t}$. However, in our case, $\mathbf{x}_{1t}$ are also the states of interest, and the splitting facilitates the use of the hybrid importance function, which is much more efficient than the prior importance function and more applicable than the posterior importance function. In fact, the hybrid importance function can be used in Rao–Blackwellised filters. In this case, we split the states into three sets, say, $\mathbf{x}_{1t}$, $\mathbf{x}_{2t}$, and $\mathbf{x}_{3t}$ where $\mathbf{x}_{1t}$ are nuisance states. Thus, Rao–Blackwellised filters can be applied to marginalize $\mathbf{x}_{1t}$, and the hybrid importance function can then be applied on $\mathbf{x}_{2t}$ and $\mathbf{x}_{3t}$ to achieve efficient implantation. We provide an example of such usage in the next section. Additional examples can be also found in [7]. Furthermore, the hybrid importance function can be utilized with other techniques for reducing the variance of the weights including the auxiliary particle filter [1], [8].

III. AUXILIARY PARTICLE FILTERING WITH THE HYBRID IMPORTANCE FUNCTION

The hybrid importance function generates the particles of $\mathbf{x}_{2t}$ with the prior distribution. Thus, the aforementioned problems of using the prior importance function are also inherited with the hybrid importance function on the particles of $\mathbf{x}_{2t}$. When the process noise is small, one may get around this problem by combining auxiliary particle filtering with the hybrid importance function. In addition, smoothing kernel can also be incorporated if $\mathbf{x}_{2t}$ are static states. However, it should be noted that, although the use of smoothing kernel proliferates the diversity of particles, such online estimation is suboptimal. We summarize one such procedure as follows.

For $j = 1, \ldots, N$, at time t, compute a point estimate $\boldsymbol{\mu}_{2t}^{(j)} = \alpha \mathbf{x}_{t-1}^{(j)} + (1 - \alpha)\bar{\mathbf{x}}_{t-1}$ where $\bar{\mathbf{x}}_{t-1} = \sum_{j=1}^{N} w_{t-1}^{(j)} \mathbf{x}_{t-1}^{(j)}$.

For $j = 1, \ldots, N$:

- Sample an auxiliary variable from $\{1, \cdots, N\}$ with probability $\lambda_t^{(i)} \propto w_{t-1}^{(i)} p(\mathbf{y}_t|\boldsymbol{\mu}_{2t}^{(i)}, \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1})$, and call the sample index k.
- Sample $\mathbf{x}_{2t}^{(j)} \sim \mathcal{N}(\boldsymbol{\mu}_{2t}^{(k)}, h^2 \mathbf{V}_t)$ where $\mathbf{V}_t$ is the weighted sample covariance matrix.
- Sample $\mathbf{x}_{1t}^{(j)} \sim p(\mathbf{x}_{1t}|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(k)}, \mathbf{y}_{0:t})$ and define $\mathbf{x}_{0:t}^{(j)} = \{\mathbf{x}_{1t}^{(j)}, \mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(k)}\}$.
- Evaluate the corresponding weight by

$$w_t^{(j)} \propto \frac{p\left(\mathbf{y}_t|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(k)}, \mathbf{y}_{0:t-1}\right)}{p\left(\mathbf{y}_t|\boldsymbol{\mu}_{2t}^{(k)}, \mathbf{x}_{0:t-1}^{(k)}, \mathbf{y}_{0:t-1}\right)}. \tag{5}$$

It is suggested in [8] that $\alpha = \sqrt{1 - h^2}$, $h^2 = 1 - ((3\epsilon - 1)/2\epsilon)^2$, and ϵ is a discount factor typically from the set [0.95, 0.99].

IV. ADAPTIVE BLIND DETECTION IN FLAT FADING CHANNELS WITH UNKNOWN NOISE VARIANCE

We demonstrate the use and the advantage of the proposed hybrid importance function on the problem of adaptive blind detection in flat fading channels. A similar problem has been addressed in [6]. Here, however, we consider a more realistic situation, where we assume that the observation noise variance is unknown. Consequently, the scheme proposed in [6] is no longer feasible.

A. Problem Formulation

The state-space model for detection of digital signals in flat fading channels can be represented as follows:

$$\begin{cases} \mathbf{h}_t = \mathbf{D}\mathbf{h}_{t-1} + \mathbf{g}v_t, & \sigma_t^2 = \sigma_{t-1}^2 \\ y_t = \mathbf{g}^{\mathrm{T}}\mathbf{h}_t s_t + e_t \end{cases} \tag{6}$$

where s_t is an M-ary transmitted signal, e_t is a complex Gaussian noise with zero mean and unknown variance σ^2, h_t is a complex Rayleigh fading coefficients whose temporal correlation is modeled by an AR(2) process with parameters a_1 and a_2, $\mathbf{h}_t = [h_t \; h_{t-1}]^{\mathrm{T}}$, $v_t \sim \mathcal{CN}(0,1)$, $\mathbf{g} = [1 \; 0]^{\mathrm{T}}$, and

$$\mathbf{D} = \begin{bmatrix} -a_1 & -a_2 \\ 1 & 0 \end{bmatrix}.$$

$$\begin{aligned} w_t^{(j)} &\propto w_{t-1}^{(j)} \frac{p\left(\mathbf{x}_{0:t}^{(j)}|\mathbf{y}_{0:t}\right)}{p\left(\mathbf{x}_{0:t-1}^{(j)}|\mathbf{y}_{0:t-1}\right) p\left(\mathbf{x}_{1t}^{(j)}|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t}\right) p\left(\mathbf{x}_{2t}^{(j)}|\mathbf{x}_{2,t-1}^{(j)}\right)} \\ &\propto w_{t-1}^{(j)} \frac{p\left(\mathbf{y}_t|\mathbf{x}_{1t}^{(j)}, \mathbf{x}_{2t}^{(j)}\right) p\left(\mathbf{x}_{1t}^{(j)}, \mathbf{x}_{2t}^{(j)}|\mathbf{x}_{1,t-1}^{(j)}, \mathbf{x}_{2,t-1}^{(j)}\right) p\left(\mathbf{x}_{0:t-1}^{(j)}|\mathbf{y}_{0:t-1}\right)}{p\left(\mathbf{x}_{0:t-1}^{(j)}|\mathbf{y}_{0:t-1}\right) p\left(\mathbf{x}_{1t}^{(j)}|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t}\right) p\left(\mathbf{x}_{2t}^{(j)}|\mathbf{x}_{2,t-1}^{(j)}\right)} \\ &= w_{t-1}^{(j)} \frac{p\left(\mathbf{y}_t|\mathbf{x}_{1t}^{(j)}, \mathbf{x}_{2t}^{(j)}\right) p\left(\mathbf{x}_{1t}^{(j)}|\mathbf{x}_{1,t-1}^{(j)}\right) p\left(\mathbf{y}_t|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1}\right)}{p\left(\mathbf{y}_t|\mathbf{x}_{1t}^{(j)}, \mathbf{x}_{2t}^{(j)}\right) p\left(\mathbf{x}_{1t}^{(j)}|\mathbf{x}_{1,t-1}^{(j)}\right)} \\ &= w_{t-1}^{(j)} p\left(\mathbf{y}_t|\mathbf{x}_{2t}^{(j)}, \mathbf{x}_{0:t-1}^{(j)}, \mathbf{y}_{0:t-1}\right) \end{aligned} \tag{4}$$

At any time instant t, the unknowns of the problem are s_t, h_t and σ_t^2, and our main objective is to detect the transmitted signal s_t sequentially without sending pilot signals. For convenience, we denote σ^2 by θ.

B. Particle Filtering Solution

First we observe that given θ_t and s_t, (6) is linear in h_t. Therefore, the mixture Kalman filter can be used to marginalize out h_t and particle filtering is then only applied on s_t and θ_t. Due to the presence of θ_t, the posterior importance function is intractable. Alternatively, since s_t is a discrete variable, we can adopt the hybrid importance function as

$$q\left(s_t, \theta_t | s_{t-1}^{(j)}, \theta_{t-1}^{(j)}, \mathbf{y}_{0:t}\right) = p\left(s_t | \theta_t, \mathbf{y}_{0:t}\right) \delta\left(\theta_t - \theta_{t-1}^{(j)}\right) \quad (7)$$

where $\delta(\cdot)$ is the Dirac delta function. The corresponding weight is obtained from (4) as

$$w_t^{(j)} \propto w_{t-1}^{(j)} \sum_{s_t \in \mathcal{A}} p\left(y_t | s_t, \theta_t^{(j)}, \mathbf{y}_{0:t-1}\right) \quad (8)$$

where $\mathcal{A} = \{a_1, \cdots, a_M\}$ is the alphabet space of s_t. Since $p(s_t | \theta_t^{(j)}, \mathbf{y}_{0:t}) \propto p(y_t | s_t, \theta_{t-1}^{(j)}, \mathbf{y}_{0:t-1})$, we find that the sampling of s_t and the calculation of the weight are achieved by computing $p(y_t | s_t, \theta_t^{(j)}, \mathbf{y}_{0:t-1})$. This distribution is the likelihood function after marginalizing out h_t and is obtained from the predictive step of the Kalman filter. It is given by

$$p\left(y_t | s_t, \theta_t^{(j)}, \mathbf{y}_{0:t-1}\right) = \mathcal{CN}\left(m_t^{(j)}, c_t^{(j)}\right) \quad (9)$$

where $m_t^{(j)} = \mu_t^{(j)} s_t$, $\mu_t^{(j)} = \mathbf{g}^T \mathbf{D} \boldsymbol{\gamma}_{t-1}^{(j)}$ and $c_t^{(j)} = \mathbf{g}^T \boldsymbol{\Sigma}_t^{(j)} \mathbf{g} + \theta_t^{(j)}$ with $\boldsymbol{\Sigma}_t = \mathbf{D} \mathbf{P}_{t-1}^{(j)} \mathbf{D}^H + \mathbf{g}\mathbf{g}^\top$, and where H denotes the Hermitian transpose. Moreover, $\boldsymbol{\gamma}_{t-1}^{(j)}$ and $\mathbf{P}_{t-1}^{(j)}$ are computed from the update steps of the Kalman filter and can be expressed according to $\boldsymbol{\gamma}_t^{(j)} = \mathbf{D} \boldsymbol{\gamma}_{t-1}^{(j)} + \mathbf{K}_t^{(j)}(y_t - m_t^{(j)})$ and $\mathbf{P}_t^{(j)} = (\mathbf{I} - \mathbf{K}_t^{(j)} \mathbf{g} s_t^{(j)}) \boldsymbol{\Sigma}_t^{(j)}$, where $\mathbf{K}_t^{(j)} = \boldsymbol{\Sigma}_t^{(j)} \mathbf{g} c_t^{(j)-1} s_t^{(j)}$. As for θ_t, we notice that no sampling is required. Although this simplifies the sampling process, the absence of sampling introduces lack of diversity on θ_t. To address this problem, smoothing kernel techniques can be used during the resampling procedure.

V. SIMULATION

The performance of the proposed particle filtering scheme using the hybrid importance function was studied for a system with $a_1 = -1.99348$ and $a_2 = 0.996$. They reflect a physical scenario of a Doppler spread of 113 Hz and data rate of 10 Kbps. This AR process was normalized to have a unit power. The transmitted signal was BPSK modulated with differential coding. In the following example, four different particle filtering schemes were examined. Two of them used the hybrid importance function: one with a resampling procedure at every ten transmissions (HIF-R) and the other, with a smoothing kernel for σ^2, applied resampling at every transmission (HIF-RS). There were also two schemes that used the prior importance function. One of them employed resampling at every ten transmissions (PIF-R) and the other, used a smoothing kernel (PIF-RS) and resampling at every step. Also, the HIFs used 150 particles, whereas the
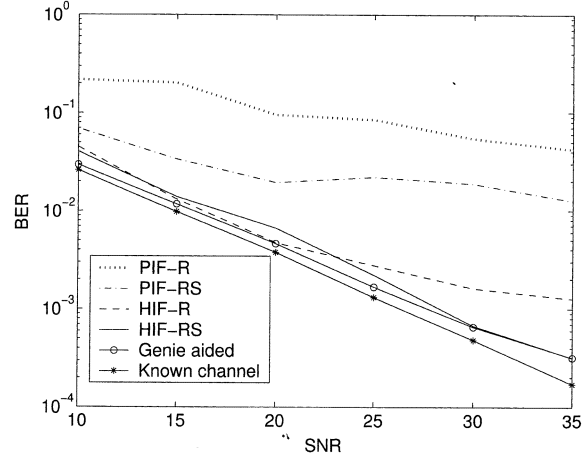

Fig. 1. Comparison plot of BERs versus SNRs of various detectors.

PIFs employed 2000 particles. In all the cases we calculated the MMSE estimator for s_t.

In Fig. 1, we plotted the bit error rates (BERs) of the four particle filters for various SNRs. The results show that the PIF-RS solution provided improvement over the results of PIF-R and that HIF-R clearly achieved the best performance. It should be noted that it only used about 1/13 of the number of particles used by the two PIF filters. Also, we observed that by applying the smoothing kernel to σ^2, significant improvements were achieved by HIF-RS. Especially, at high SNRs, the performance of the HIF-RS filter approached that of the genie aided detector with known noise variance.

VI. CONCLUSION

In this letter, we proposed a hybrid importance function which encompasses the advantages of both the posterior and the prior importance functions. We have shown its application to blind detection in flat fading channels. Simulation results showed much improved performance over particle filters that use the prior importance function only.

REFERENCES

[1] A. Doucet, J. de Freitas, and N. Gordon, Eds., *Sequential Monte-Carlo Methods in Practice*. Berlin, Germany: Springer-Verlag, 2000.
[2] M. Pitt and N. Shephard, "Filtering via simulation: auxiliary particle filter," *J. Amer. Statist. Assoc.*, vol. 94, pp. 590–599, 1999.
[3] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statist. Comput.*, vol. 10, no. 3, pp. 197–208, 2000.
[4] C. Ch Andrieu, A. Doucet, and E. Punskaya, "Sequential Monte Carlo methods for optimal filtering," in *Sequential Monte Carlo Methods in Practice*, J. F. G. De Freitas, A. Doucet, and N. J. Gordon, Eds. New York: Springer-Verlag, 2001.
[5] R. Chen and J. Liu, "Mixture Kalman filters," *J. R. Statist. Soc. B*, vol. 62, pp. 493–508, 2000.
[6] R. Chen, X. Wang, and J. S. Liu, "Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2079–2094, Sept. 2000.
[7] Y. Huang and P. M. Djurić, "A blind particle filtering detector for joint channel estimation and symbol detection over flat fading channels," *IEEE Trans. Signal Processing*, 2004, to be published.
[8] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, J. F. G. De Freitas, and N. J. Gordon, Eds. New York: Springer-Verlag, 2000.