On the use and misuse of particle filtering in digital communications

E. Punskaya.[†*] A. Doucet.[‡] W.J. Fitzgerald[†]
[†]Signal Processing Laboratory, Department of Engineering,
University of Cambridge, Trumpington Street, CB2 1PZ, Cambridge, UK.
[‡]Department of Electrical and Electronic Engineering,
The University of Melbourne, Victoria, 3010, Australia.
Email: op205@eng.cam.ac.uk - a.doucet@ee.mu.oz.au - wjf@eng.cam.ac.uk

ABSTRACT

In this paper, application of particle filtering techniques to different classes of problems arising in digital communications is considered. Several approaches are reviewed, and a brief simulation study for demodulation in fading conditions and joint symbol/channel coefficients/code delay estimation for DS spread-spectrum systems is carried out.

1 Introduction

Particle filtering techniques are a set of powerful and versatile simulation-based methods to perform optimal state estimation in nonlinear non-Gaussian state space models. The approach has recently received a lot of interest since it allows a large number of challenging non-linear estimation tasks to be addressed in an efficient on-line manner; see [5] for a survey. The idea is to approximate the posterior distribution of interest by swarms of weighted points in the sample space, called particles, which evolve randomly in time in correlation with each other, and either give birth to offspring particles or die according to their ability to represent the different zones of interest of the state space.

Since many problems arising in digital communications can be considered as optimal filtering problems, the application of particle filters seems only to be a sensible choice. One must be careful, however. In most of the cases, those problems fall into two general classes. In the *first* one, the unknown state of the model - typically the transmitted symbol(s) - takes its values in a finite set; this includes, for example, demodulation in fading channels [2, 10], OFDM systems and multiuser detection in synchronous CDMA [11]. In the *second* class, one faces more challenging problem where the unknown state of interest consists not only of the symbol(s) but also some continuous-valued parameters such as the code delays as in DS spread-spectrum system analyses.

The main objective of this paper is to show that, for a fixed computational complexity, *particle filtering techniques may actually perform worse than simple deterministic algorithms to solve problems in the first class.* For the second class of problems, however, particle filters prove really useful as demonstrated later.

The rest of the paper is organized as follow. In Section 2, we present a general signal model including discrete parameters (transmitted symbols) and continuous parameters (code delays). Section 3 reviews particle filtering techniques in the case where only the symbols are unknown. It outlines the weakness of these techniques in this context. Section 4 considers the case of a mixed continuous-discrete case and presents a generic particle filtering algorithm to address this problem. Finally, a conclusion is drawn in Section 5.

2 Problem statement and estimation objectives

Transmitted waveform. Let us denote for any generic sequence κ_t, $\kappa_{i:j} \triangleq (\kappa_i, \kappa_{i+1}, \ldots, \kappa_j)^{\mathsf{T}}$, and let d_n be the nth information symbol, and $s(\tau)$ be the corresponding analog bandpass spread-spectrum signal waveform transmitted in the symbol interval of duration T_d:

$$s_{\text{trans}}(\tau) = \text{Re}[r_n(d_n) PN(\tau) \exp(j2\pi f_c \tau)], \quad \text{for } (n-1)T_d < \tau \leq nT_d,$$

where $r_n(.)$ performs the mapping from the digital sequence to waveforms and corresponds to the modulation technique employed, f_c denotes the carrier frequency and $PN(\tau)$ is a wide-band pseudo-noise (PN) waveform defined by $PN(\tau) = \sum_{h=1}^{H} a_h \eta(\tau - hT_c)$. Here, $a_{1:H}$ is a spreading code sequence[1] consisting of H chips (with values $\{\pm 1\}$) per symbol, $\eta(\tau - hT_c)$ is a rectangular pulse of unit height and duration T_c, and T_c is the chip interval satisfying the relation $T_c = T_d/H$.

Channel model. The signal is passed through a noisy multipath fading channel which causes random amplitude and phase variations on the signal. The channel can be represented by a time-varying tapped-delayed line with taps spaced T_s seconds apart, where T_s is the Nyquist sampling rate for the transmitted waveform; $T_s = T_c/2$ due to the PN bandwidth being approximately $1/T_c$. The equivalent discrete-time impulse re-

* E. Punskaya is supported by the EPSRC grant, UK.

[1] The extension to a multiuser DS CDMA transmission is straightforward, see [9, 11], for example.

sponse of the channel is given by

$$h_{c,t} = \sum_{n_f=0}^{N_f-1} f_t^{(n_f)} \delta_{t,n_f},$$

where t is a discrete time index, N_f is the number of paths of the channel, $f_t^{(n_f)}$ are the complex-valued time-varying multipath coefficients arranged into the vector $\mathbf{f}_t$, and δ_{t,n_f} denotes the Kronecker delta.

We assume here that the channel coefficients $\mathbf{f}_t$ and code delay θ_t propagate according to the first-order autoregressive (AR) model:

$$\mathbf{f}_t = \mathbf{A}_f \mathbf{f}_{t-1} + \mathbf{B}_f \mathbf{v}_t, \mathbf{v}_t \overset{i.i.d.}{\sim} \mathcal{N}_c\left(\mathbf{0}, \mathbf{I}_{N_f}\right), \quad (1)$$

$$\theta_t = \gamma\theta_{t-1} + \sigma_\theta \epsilon_t,\ \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad (2)$$

which corresponds to a Rayleigh uncorrelated scattering channel model; here $\mathbf{A}_f \triangleq diag(\alpha_0, \ldots, \alpha_{N_f-1})$, $\mathbf{B}_f \triangleq diag(\sigma_{f,0}, \ldots, \sigma_{f,N_f-1})$, with σ^2_{f,n_f} being the noise variance, and α_{n_f} accounting for the Doppler spread (see [7] for details and discussion on the use of the higher order AR).

Received signal. The complex output of the channel sampled at the Nyquist rate, (in which case $t = 2H(n-1)+1, \ldots, 2Hn$ samples correspond to the nth symbol transmitted, i.e. $d_n \leftrightarrow y_{2H(n-1)+1:2Hn}$) can, thus, be expressed as

$$y_t = \mathbf{C}(d_{1:n}, \theta_{1:t}) + \sigma\varepsilon_t,\ \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1), \quad (3)$$

where $\mathbf{C}(d_{1:n}, \theta_{1:t}) = \sum_{n_f=0}^{N_f-1} f_t^{(n_f)} s\left((t-n_f)T_s - \theta_t\right)$ and σ^2 being the noise variance[2]. The noise sequences ε_t, ϵ_t and $v_t^{(n_f)}$, $n = 0, \ldots, N_f - 1$ are assumed mutually independent and independent of the initial states $\mathbf{f}_0 \sim \mathcal{N}_c\left(\hat{\mathbf{f}}_0, \mathbf{P}_{\mathbf{f},0}\right)$, $\theta_0 \sim \mathcal{N}\left(\hat{\theta}_0, P_{\theta,0}\right)$.

Estimation objectives. The symbols d_n, which are assumed i.i.d., the channel characteristics $\mathbf{f}_t$ and the code delay θ_t are unknown for $n, t > 0$. Our aim is to obtain sequentially in time an estimate of the joint posterior probability density of these parameters $p\left(d_{1:n}, \mathbf{f}_{0:2Hn}, \theta_{0:2Hn} \middle| y_{1:2Hn}\right)$, and, in particular, some of its characteristics, such as $\mathbb{E}\left(d_{1:n} \middle| y_{1:2Hn}\right)$, $\mathbb{E}\left(\mathbf{f}_{0:2Hn} \middle| y_{1:2Hn}\right)$ and $\mathbb{E}\left(\theta_{0:2Hn} \middle| y_{1:2Hn}\right)$. However, this problem does not admit any analytical solution and, thus, approximate methods must be employed. One of the methods that has proved to be useful in practice is particle filtering techniques, the use of which we investigate in this paper.

3 Particle filtering for demodulation

First, let us develop the particle filtering algorithm for the channel with no delay, $\theta_t = 0$, and no spreading sequence employed, $H = 1$, i.e. let us design a simple particle filtering receiver for demodulation of the transmitted signal in multipath fading conditions.

[2]The case of non-Gaussian noise can be easily treated using the techniques presented in [10].

A similar demodulator has already been considered in [2, 5, 10], although for the flat Rayleigh channels, where the use of *Sequential Importance Sampling and Resampling (SISR)* is proposed. The problem of estimating $p\left(d_{1:n}, \mathbf{f}_{0:n} \middle| y_{1:n}\right)$ is reduced there to one of sampling from a lower-dimensional posterior $p\left(d_{1:n} \middle| y_{1:n}\right)$; based on the fact that, conditional upon the sequence $d_{1:n}$, the density $p\left(\mathbf{f}_{0:n} \middle| y_{1:n}, d_{1:n}\right)$ can be computed using the Kalman filter, see [1, 3, 4], and, thus, $p\left(\mathbf{f}_n \middle| y_{1:n}\right)$ be approximated by a random mixture of Gaussian distributions. According to the algorithm, N particles $\left\{d_{1:n}^{(i)}\right\}_{i=1}^{N}$ are simulated according to an arbitrary convenient importance distribution $\pi\left(d_{1:n} \middle| y_{1:n}\right)$ (such that $p\left(d_{1:n} \middle| y_{1:n}\right) > 0$ implies $\pi\left(d_{1:n} \middle| y_{1:n}\right) > 0$), and the estimate of $p\left(d_{1:n} \middle| y_{1:n}\right)$ is obtained using the importance sampling identity:

$$\hat{p}_N\left(d_{1:n} \middle| y_{1:n}\right) = \sum_{i=1}^{N} \tilde{w}_{1:n}^{(i)} \delta_{\left(d_{1:n}^{(i)}\right)}(d_{1:n}), \quad (4)$$

where $\tilde{w}_{1:n}^{(i)}$ are the so-called *importance weights*

$$\tilde{w}_{1:n}^{(i)} = \frac{w_{1:n}^{(i)}}{\sum_{j=1}^{N} w_{1:n}^{(j)}},\ w_{1:n}^{(i)} \propto \frac{p\left(d_{1:n}^{(i)} \middle| y_{1:n}\right)}{\pi\left(d_{1:n}^{(i)} \middle| y_{1:n}\right)}.$$

An additional condition of $\pi\left(d_{1:n} \middle| y_{1:n}\right) > 0$ having to admit $\pi\left(d_{1:n-1} \middle| y_{1:n-1}\right) > 0$ as a marginal distribution allows to propagate this estimate sequentially in time, and a selection procedure helps to avoid the degeneracy of the algorithm (see [3, 4, 9] for the details of the algorithm).

The computational complexity of this approach largely depends on the importance distribution choice and the selection scheme being employed. The basic idea would be to use the prior distribution as an importance distribution, $\pi\left(d_n \middle| y_{1:n}, d_{n-1}\right) = p\left(d_n \middle| d_{n-1}\right)$, thus calculating just one Kalman filter step for each particle. This can be inefficient, however, as no information carried by y_n is used to explore the state space. The employment of the "*optimal*" importance distribution $\pi\left(d_n \middle| d_{1:n-1}, d_{1:n}\right) = p\left(d_n \middle| d_{1:n-1}, y_{1:n}\right)$, see [2]-[4], which minimizes the conditional variance of $w\left(d_{1:n}\right)$, may, in turn, be quite computationally extensive. Indeed,

$$w_n \propto \sum_{m=1}^{M} p\left(y_n \middle| d_{1:n-1}^{(i)}, d_n = \rho_m, y_{1:n-1}\right),$$

with ρ_m corresponding to the mth $(m = 1, \ldots, M)$ possible realization of d_n (see [9] for details), and M one-step ahead Kalman filters are required.

Moreover, since all the calculations have to be performed anyway, it is better to base our approximation of $p\left(d_{1:n} \middle| y_{1:n}\right)$ (hereafter we refer to this method as *deterministic*) directly on :

$$\hat{p}_{N\times M}\left(d_n \middle| y_{1:n}\right) = \sum_{i=1}^{N} \sum_{m=1}^{M} \tilde{w}_n^{(i,m)} \delta_{\left(\left\{d_{1:n-1}^{(i)}, d_n = \rho_m\right\}\right)}(d_{1:n}),$$

$$w_n^{(i,m)} \propto p\left(y_n \middle| d_{1:n-1}^{(i)}, d_n = \rho_m, y_{1:n-1}\right), \qquad (5)$$

thus, considering all possible "extensions" of the existing state sequences for each particle at step n. In this case, one does not discard unnecessarily any information by selecting randomly one path out of the M available as in SISR. However, a selection procedure still has to be employed, since each particle has M offspring at each step n in this approach, resulting in the exponentially increasing number of them.

The simplest way to perform such selection is to choose the N most likely offspring and discard the others (as, for example, in [12]). The weight of each of $N \times M$ particles in this case depends on the weight of the parent at step $n-1$ as well as the likelihood term (5) computed using the Kalman filter. A more complicated approach involves preserving the particles with high weights and resampling the ones with low weights, thus reducing their total number to N, as, for example, in [6]. An important condition for the design of the selection scheme in this specific context is to resample *without replacement*, as, indeed, *there is no point in carrying along two particles evolving in exactly the same way*, so each of them should appear at most once in the resulting set.

Whether we choose to preserve the most likely particles or employ the selection scheme proposed in [6], the computational load of the resulting algorithms at each step n is that of $N \times M$ Kalman filters, and the selection step in both cases is implemented in $O(N \times M \log N \times M)$ operations compared to $O(N)$ when, for example, the stratified sampling [8] in SISR is employed. Of course, if M is large, which is the case in many applications (see Section 4, for example), both these methods are too computationally extensive to be used.

To conclude, one could hope that randomization "helps" by allowing particles with a small weight to survive, but simulations presented in Section 5 and in [11] show that it is not necessarily the case. In this very specific but important context, particle filtering algorithms do not perform better than the simplest deterministic method which consists of keeping at each time step the best N hypothesis!

In the case where it is to costly to explore M hypothesis for each particle, though, that is one uses an important distribution different from the optimal one, particle filtering could prove useful if one could develop suboptimal importance distributions with "good" properties. This problem has to be addressed on a case by case basis. An interesting way to explore consists of randomizing standard deterministic algorithms such as successive interference cancellation or iterative least squares. However, advanced deterministic pruning strategies can also be developed using, for example, a coordinate ascent version of the algorithm proposed in [4]. The comparison of deterministic and randomized algorithms deserves further study.

4 Particle filtering for joint demodulation and code delay estimation

Let us now consider the problem of joint estimation of the symbols, channel coefficients and code delay for DS spread-spectrum systems, i.e. let us focus on the estimation of the joint posterior distribution $p\left(d_{1:n}, d\mathbf{f}_{0:2Hn}, d\theta_{0:2Hn} \middle| y_{1:2Hn}\right) = p\left(d_{1:n}, \mathbf{f}_{0:2Hn}, \theta_{0:2Hn} \middle| y_{1:2Hn}\right) d\mathbf{f}_{0:2Hn} d\theta_{0:2Hn}$.

Again, we can restrict ourselves to approximating the lower-dimensional distribution $p\left(d_{1:n}, d\theta_{0:2Hn} \middle| y_{1:2Hn}\right)$ through particle filtering:

$$\begin{aligned}\hat{p}\left(d_{1:n}, d\theta_{1:2Hn} \middle| y_{1:2Hn}\right) \\ = \textstyle\sum_{i=1}^{N} \tilde{w}_n^{(i)} \delta_{\left(d_{1:n}^{(i)}, \theta_{1:2nH}^{(i)}\right)}\left(d_{1:n}, d\theta_{0:2Hn}\right),\end{aligned}$$

and, then, if necessarily, calculating $p\left(\mathbf{f}_{1:2Hn} \middle| y_{1:2Hn}\right)$ as a mixture of Gaussians computed through the Kalman filter associated with the Eq.(1) and (3):

$$p\left(\mathbf{f}_{1:2Hn} \middle| y_{1:2Hn}\right) = \textstyle\sum_{i=1}^{N} p(\mathbf{f}_{1:2Hn} | y_{1:2Hn}, d_{1:n}^{(i)}, \theta_{1:2Hn}^{(i)}),$$

Then, given for symbol $n-1$, N particles $(d_{1:(n-1)}^{(i)}, \theta_{1:2H(n-1)}^{(i)})$, $i = 1, \ldots, N$ distributed approximately according to $p\left(d_{1:n-1}, d\theta_{1:2H(n-1)} \middle| y_{1:2H(n-1)}\right)$, the basic particle filtering receiver proceeds as follows:

Particle Filtering Algorithm

Sequential Importance Sampling Step

- For $i = 1, \ldots, N$, sample $(\widetilde{d}_n^{(i)}, \widetilde{\theta}_{2H(n-1)+1:2Hn}^{(i)}) \sim \pi(d_n, \theta_{2H(n-1)+1:2Hn} | d_{1:n-1}^{(i)}, \theta_{1:2H(n-1)}^{(i)}, y_{1:2Hn})$.
- For $i = 1, \ldots, N$, evaluate the importance weights $w_n^{(i)}$ up to a normalizing constant.
- For $i = 1, \ldots, N$, normalize $w_n^{(i)}$ to obtain $\tilde{w}_n^{(i)}$.

Selection Step

- Multiply/discard particles with respect to high/low $\tilde{w}_n^{(i)}$ to obtain N particles $(d_{1:n}^{(i)}, \theta_{1:2Hn}^{(i)})$

If, say, the prior is taken to be the importance distribution i.e.

$$\begin{aligned}\pi(d_n, \theta_{2H(n-1)+1:2Hn} | d_{1:n-1}, \theta_{1:2H(n-1)}, y_{1:2Hn}) = \\ p(d_n) \textstyle\prod_{t=2H(n-1)+1}^{2Hn} p(\theta_t | \theta_{t-1}),\end{aligned}$$

then w_n becomes

$$\begin{aligned}w_n \propto p\left(y_{2H(n-1)+1:2Hn} \middle| y_{1:2H(n-1)}, \tilde{d}_{1:n}^{(i)}, \tilde{\theta}_{1:2Hn}^{(i)}\right) \\ = \textstyle\prod_{t=2H(n-1)+1}^{2Hn} p\left(y_t \middle| d_{1:n}, \theta_{1:t}, y_{1:t-1}\right),\end{aligned}$$

and requires evaluation of $2H$ one-step Kalman filter updates.

In cases where $H \gg 1$, this method is unlikely to perform well as the state space to explore is large. An alternative algorithm consists of, first, only sampling $\left(d_n, \theta_{2H(n-1)+1}\right)$ and updating the distribution with $y_{2H(n-1)+1}$. Then, in the next steps, one samples θ_k (k going from $2H(n-1)+2$ to $2Hn$) according to $p(\theta_k | \theta_{k-1})$, and updates the distribution with y_k.

5 Simulation Results

First, computer simulations were carried out in order to compare the bit-error-rate (BER) performance of the algorithms presented in Section 3 for demodulation of $2PSK$ symbols transmitted over fast Rayleigh fading channels with normalized channel Doppler frequency 0.05 and $N_f = 4$. The results for different average signal to noise ratio (SNR) are presented in Fig.1 for $N = 50$ (pilot symbol rate is $1:20$). They are interesting in the sense that, in this case, the deterministic approach preserving the N most likely particles (MLP) turned out to be the most efficient one. With other simulation parameters, however, one may find that the results between the different algorithms are much less pronounced.

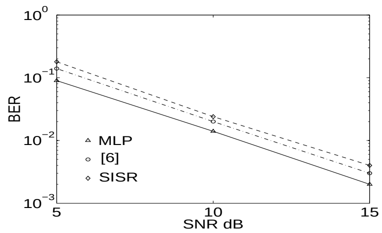

Figure 1: Demodulation. Bit error rate.

In the second experiment, a basic SISR algorithm presented in Section 4 (the deterministic approach is not applicable in this case) was applied to perform joint symbols/channel coefficients/code delay estimation for DS spread-spectrum systems with $H = 15$, $N_f = 4$ and the multipath channel response as in [7, channel B]. The AR parameters corresponded to the case of nearly constant coefficients and constant delay were set, $\alpha_{n_f} = 0.999$, $\gamma = 0.999$, $\sigma^2_{f,n_f} = 0.001$, $\sigma^2_\theta = 0.001$. As it is shown in Fig.2, the algorithm exhibits good performance even for just $N = 50$ particles being employed.

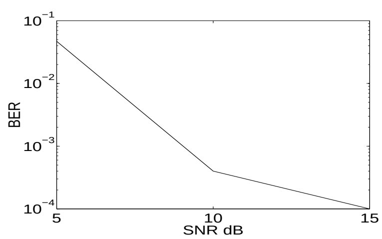

Figure 2: DS spread spectrum system. Bit error rate.

6 Conclusion

In this paper, we study the application of particle filtering techniques to different types of problems arising in digital communications. In the context where only symbols are unknown, as in demodulation, standard particle filtering methods, although quite capable of providing good performance, do not necessarily compare favorably with deterministic approaches: as simulations show, the most basic deterministic algorithm preserving the N most likely particles also turns out to be the most efficient one. However, for more complex problems involving continuous-valued unknown parameters, such as DS spread-spectrum systems analyses, or, indeed, in situations where MLP and similar methods are of no use due to their computational complexity, as with more efficient M-ary modulation (M being relatively large), additive non-Gaussian noise and multiuser detection [9], these deterministic approaches do not apply and particle methods appears to be really useful.

7 Acknowledgment

The authors would like to acknowledge Simon Maskell for constructive and insightful comments which led us to a more critical evaluation of our methodology.

References

[1] R. Chen, and J. Liu, "Mixture Kalman Filters," *J. Roy. Statist. Soc.* B, vol. 62, pp. 493-508, 2000.

[2] R. Chen, X. Wang, and J.S. Liu, "Adaptive Joint Detection and Decoding in Flat-Fading Channels via Mixture Kalman Filtering," *IEEE Trans. Information Theory*, vol. 46, no. 6, pp. 2079-2094, 2000.

[3] A. Doucet, S. Godsill and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing,* vol. 10, no. 3, pp. 197-208, 2000.

[4] A. Doucet, N.J. Gordon and V. Krishnamurthy, "Particle Filters for State Estimation of Jump Markov Linear Systems," *IEEE Trans. on Signal Processing,* vol. 49, no.3, pp. 613-624, 2001.

[5] A. Doucet, J.F.G. de Freitas and N.J. Gordon (eds.), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag: New-York, 2001.

[6] P. Fearnhead, Sequential Monte Carlo methods in filter theory, PhD thesis, University of Oxford, 1998.

[7] R.A. Iitis, "Joint Estimation of PN Code Delay and Multipath Using the Extended Kalman Filter," *IEEE Trans. on Communications*, vol. 38, pp. 1677-1685, 1990.

[8] G. Kitagawa, "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *J. Comp. Graph. Stat.*, vol. 5, no. 1, pp. 1-25, 1996.

[9] E. Punskaya, C. Andrieu, A. Doucet and W.J. Fitzgerald, "Multiuser CDMA detection under fading conditions using particle filtering," technical report, Cambridge University Engineering Department, CUED/F-INFENG/TR. 413, 2001.

[10] E. Punskaya, C. Andrieu, A. Doucet and W.J. Fitzgerald, "Particle Filtering for Demodulation in Fading Channels with Non-Gaussian Additive Noise," *IEEE Trans. on Communications*, vol. 49, no. 4, pp. 579-582, 2001.

[11] E. Punskaya, C. Andrieu, A. Doucet and W.J. Fitzgerald, "Particle Filtering for Multiuser Detection in Fading CDMA Channels," *Proc IEEE SSP*, 2001.

[12] J.K. Tugnait and A.H. Haddad, "A Detection-estimation Scheme for State Estimation in Switching Environment," *Automatica*, vol. 15, pp. 477-481, 1979.