GLOBAL SAMPLING FOR SEQUENTIAL FILTERING OVER DISCRETE STATE SPACE

P. Cheung-Mon Chan, E. Moulines

Ecole Nationale Supérieure des Télécommunications, 46, rue Barrault, 75634 PARIS Cédex 13

ABSTRACT
In many situations, it is required to approximate sequence of probability measures over a growing product of finite spaces. This is typically the case in digital communications, where the finite space is the symbol alphabet and the probability measures to be approximated are the posterior distribution of the transmitted symbols given the observations. Whereas it is in general possible to compute explicitly these probability measures, the typical complexity of these computations grows exponentially, precluding real-time implementations.

In this paper, we present an efficient approach for approximating these distributions using a particular implementation of the sequential Monte-Carlo filter (SMC). SMC consists in approximating the sequence of probability measures by the empirical distribution of a finite set \( N \) of trajectories which evolve under a random mechanism.

Since the space is finite, it is possible to consider every offspring of the trajectory of particles; contrary to the classical sequential importance sampling and resampling (SISR) procedure, it is thus not required to develop a sophisticated strategy to build a finite set of trajectories. The procedure is therefore straightforward to implement, and is well-suited for real-time implementations. The approach compares favorably with SMC techniques proposed in the literature and appears to be extremely robust even when the number of particles is small.

Our aim is to estimate recursively in time conditional probabilities of the form \( P(Z_n | \mathcal{Y}_{0:n+\Delta} = \mathcal{Y}_{n+\Delta}) \) where for any sequence \( \{\lambda_i\}_{i>0} \) and any integer \( 0 \leq i < j \), we denote \( \lambda_{ij} \equiv \{\lambda_i, \ldots, \lambda_j\} \) and \( \Delta \) is a non-negative integer. Although these quantities can be computed in closed form, these computations involve high-dimensional integration and are infeasible in practice. To solve this kind of problem, it is possible to resort to sequential Monte-Carlo techniques (see e.g. [12], [11] and [7]). Since the system is a linear Gaussian state-space model conditionally on the indicator sequence \( Z_{0:n} \), efficient recursive procedures - such as the Kalman filter / smoother - can be developed to determine the distribution of the state-variable conditional to the indicator variable and the observations. These algorithms are collectively referred to as mixture Kalman filters a term apparently coined by [5]; closely related ideas have appeared earlier in the automatic control / signal processing and computational statistics literature (see e.g. [1], [13] for early work in this field; see [11] and the references therein for a tutorial of these methods).

Because the indicator variables are discrete and take a finite number of different values, it is possible to consider every possible offspring of a trajectory defined as a sequence of indicators from initial time \( 0 \) to the current time \( t \). This has been observed by [1], [10], [11], [5], among others, who have used this property to construct appropriate importance distributions to improve the accuracy and performance of sequential importance sampling and resampling (SISR) procedures. In this work, we use this key property in a different way; the basic idea consists in considering the population of possible offspring of every trajectory and sampling globally from this population. This algorithm is referred to as the global sampling algorithm. This algorithm is simpler to implement than previously reported SISR schemes because it avoids the use of any additional logic to build the importance distribution which can be costly to implement in real-time setting.

1. INTRODUCTION
In this paper, we consider a special case of state-space model, often referred in the literature to as conditionally Gaussian linear state-space models (CGLSM) which has received a lot of attention in the recent years (see e.g. [1], [2], [3], [4], [5], [6]; see also [7] and the references therein). The main feature of (CGLSM) is that, conditional to the indicator variables, the system becomes linear and Gaussian; More precisely, define

\[
\begin{align*}
S_t &= \Psi_t(Z_{0:t}) \\
\bar{X}_t &= A_S \bar{X}_{t-1} + C_S W_t \\
\bar{Y}_t &= B_S \bar{X}_t + D_S V_t
\end{align*}
\]

where

- \( \{Z_t\}_{t=0} \) are the indicators variables, here assumed to take values in a finite set \( Z = \{a_1, a_2, \ldots, a_D\} \), where \( |Z| \) denotes the cardinal of the set \( Z \); the law of the process \( \{Z_t\}_{t=0} \) is assumed to be known and can take any form;

- For any \( t \geq 0 \), \( \Psi_t \) is a function \( \Psi_t : Z^{t+1} \to S \), where \( S \) is a finite state-space;

- \( \{\bar{X}_t\}_{t=0} \) are the \( (n_S \times 1) \) (unobserved) state vectors;

- \( \{\bar{Y}_t\}_{t=0} \) are the \( (n_Y \times 1) \) observations;

- \( \{W_t\}_{t=0} \) and \( \{V_t\}_{t=0} \) are (complex) \( n_w \) and \( n_v \) dimensional (complex) gaussian white noise, \( W_t \sim \mathcal{N}(0, I_{n_w \times n_w}) \) and \( V_t \sim \mathcal{N}(0, I_{n_v \times n_v}) \), where \( I_{p \times p} \) is the \( p \times p \) identity matrix; \( \{\bar{W}_t\}_{t=0} \) is referred to as the state noise, whereas \( \{\bar{V}_t\}_{t=0} \) is the observation noise;

- The sequence of matrices \( \{A_s, s \in S\}, \{B_s, s \in S\}, \{C_s, s \in S\} \) and \( \{D_s, s \in S\} \) are known;

This model has been considered by many authors, following the pioneering work by [8] and [9] (see [10], [11] and [5] for authoritative recent surveys).

Our aim is to estimate recursively in time conditional probabilities of the form \( P(Z_n | \mathcal{Y}_{0:n+\Delta} = \mathcal{Y}_{n+\Delta}) \) where for any sequence \( \{\lambda_i\}_{i>0} \) and any integer \( 0 \leq i < j \), we denote \( \lambda_{ij} \equiv \{\lambda_i, \ldots, \lambda_j\} \) and \( \Delta \) is a non-negative integer. Although these quantities can be computed in closed form, these computations involve high-dimensional integration and are infeasible in practice. To solve this kind of problem, it is possible to resort to sequential Monte-Carlo technique (see e.g. [12], [11] and [7]). Since the system is a linear Gaussian state-space model conditionally on the indicator sequence \( Z_{0:n} \), efficient recursive procedures - such as the Kalman filter / smoother - can be developed to determine the distribution of the state-variable conditional to the indicator variable and the observations. These algorithms are collectively referred to as mixture Kalman filters a term apparently coined by [5]; closely related ideas have appeared earlier in the automatic control / signal processing and computational statistics literature (see e.g. [1], [13] for early work in this field; see [11] and the references therein for a tutorial of these methods).

Because the indicator variables are discrete and take a finite number of different values, it is possible to consider every possible offspring of a trajectory defined as a sequence of indicators from initial time \( 0 \) to the current time \( t \). This has been observed by [1], [10], [11], [5], among others, who have used this property to construct appropriate importance distributions to improve the accuracy and performance of sequential importance sampling and resampling (SISR) procedures. In this work, we use this key property in a different way; the basic idea consists in considering the population of possible offspring of every trajectory and sampling globally from this population. This algorithm is referred to as the global sampling algorithm. This algorithm is simpler to implement than previously reported SISR schemes because it avoids the use of any additional logic to build the importance distribution which can be costly to implement in real-time setting.

This work has been supported by Thalès-Communications under contract XXXX.
2. GLOBAL SAMPLING ALGORITHM

2.1. Notations and definitions

Before going further, some additional definitions and notations are required. Let \( X \) (resp. \( Y \)) be general sets, and \( B(X) \) (resp. \( B(Y) \)) denote a \( \sigma \)-field on \( X \) (resp. \( Y \)). If \( Q := \{ Q(x, B), x \in X, A \in B(Y) \} \) is such that

(i) For each \( B \in B(Y) \), \( Q(\cdot, B) \) is a non-negative measurable function on \( X \),

(ii) For each \( x \in X \), \( Q(x, \cdot) \) is a finite measure on \( (X, B) \),

then we call \( Q \) a finite transition kernel from \((X, B(X))\) to \((Y, B(Y))\) and we denote \( \mu \circ Q \) the measure on the product space \((X \times Y, B(X) \otimes B(Y))\) defined, for all \( A \in B(X), B \in B(Y) \) by

\[
\mu \circ Q(A \times B) = \int_A \mu(dx)Q(x, B) \tag{2}
\]

Let \( X : (\Omega, \mathcal{F}) \rightarrow (X, B(X)) \) and \( Y : (\Omega, \mathcal{F}) \rightarrow (Y, B(Y)) \) be two random variables (r.v. and \( \mu \) and \( \nu \) two measures on \((X, B(X))\) and \((Y, B(Y))\), respectively. Assume that the probability distribution of \( Y \) is \( \nu \). The indicator variables \( g_i(\cdot; Z, \Gamma) \) of these \( N \) trajectories can be extended to the case where one deals with the conditional distribution of \( Y \) given \( X \) and for the density of \( F_t \) w.r.t. the counting measure. With these conventions and notations, for any \( t \geq 1 \) and for any \( z_{0:t} \in \mathbb{Z}^{t+1} \), straightforward calculations show that

\[
F_t = F_{t-1} \otimes Q_t, \tag{3}
\]

where

\[
Q_t(z_{0:t}; \mathcal{A}) = \int \nu(dx)Q_t(x, z_{0:t-1})\tag{4}
\]

the predictive distribution of the observations given the indicator variables can be evaluated along each trajectory of the indicator variables \( Z_{0:t} = z_{0:t} \) using Kalman filter recursions. Denote \( g_i(\cdot; \mu, \Gamma) \) the density of a complex circular Gaussian random vector with mean \( \mu \) and covariance matrix \( \Gamma \) and for \( A \) a matrix, let \( A^\dagger \) be the transpose-conjugate of \( A \); we have, with \( \nu_t = \Psi_t(z_{0:t}) \) and \( \Psi_t \) defined in (1)

\[
\begin{align*}
&f_{Y_t}^Z(z_{0:t-1}; \tilde{y}_t) = g_{\tilde{y}_t}(\tilde{y}_t; B_{\nu_t} \tilde{z}_t, D_{\nu_t} I_t), \\
&f_{Z_t}^Y(z_{0:t-1}; \tilde{z}_t, \tilde{y}_t) = g_{\tilde{y}_t}(\tilde{y}_t; B_{\nu_t} \tilde{z}_t, D_{\nu_t} I_t) \tag{5}
\end{align*}
\]

where the conditional mean and covariance \( \mu_{t}(z_{0:t-1}, \tilde{y}_{t-1}) \) and \( \Gamma_t(z_{0:t-1}) \) are computed recursively according to

\[
\begin{align*}
&\mu_{t+1}(z_{0:t+1}, \tilde{y}_{t+1}) = U(\hat{y}_t, \hat{\Psi}_t(z_{0:t}), \Psi_{t+1}(z_{0:t+1}, \mu_t(z_{0:t-1}, \tilde{y}_{t-1}), \Gamma_t(z_{0:t-1}))), \\
&\Gamma_t(z_{0:t+1}) = A_{\nu_t}(I - K_B A_{\nu_t}) \Gamma_t(z_{0:t-1}) + C_{\nu_t} C_{\nu_t}^t
\end{align*}
\]

where \( K = \Gamma B (\Gamma B B^t + D_r D_r^t)^{-1} \) is the Kalman gain. The predictive distribution of the observation conditional to the indicator variables may be expressed as

\[
f_{Y_t}^Z(z_{0:t-1}, \tilde{y}_t) = g_{\tilde{y}_t}(\tilde{y}_t; B_{\nu_t} \tilde{z}_t, D_{\nu_t} I_t) \tag{6}
\]

the mapping \( V \) being defined as

\[
(V, \Sigma) = V(s, \tilde{y}_t, \nu_t) = (B_{\nu_t} \tilde{z}_t, D_{\nu_t} I_t)
\]

2.3. Algorithm derivation

In this section, we describe the algorithm to sample from a sequence of distributions satisfying a relation of the form (3). More specifically, let \( Z \) be a finite set. Consider a sequence \((F_t)_{t \geq 1}\) of probability measures such that

1. for all \( t \in \mathbb{N} \), \( F_t \) is a probability measure on the product probability space \((Z, P(Z))^t \)

2. for all \( t \geq 1 \), there exists a finite transition \( Q_t : (Z, P(Z)) \rightarrow (Z, P(Z)) \) such that

\[
F_t = F_{t-1} \otimes Q_t \tag{5}
\]

For each \( t \geq 0 \), the probability measure \( F_t \) is approximated by the empirical distribution associated to \( N \) trajectories that are denoted \( \Lambda^{(0)} = \{\Lambda_j^{(0)}, \ldots, \Lambda_j^{(N)}\} \), \( j \in \{1, \ldots, N\} \) and taking values in the set \((Z, P(Z))\). The main difference with classical particle filters stem from the way these trajectories are updated and selected. The global sampling strategy proceeds as follows:

Initialisation: for each \( t \geq 1 \), we draw independently \( N \) r.v. denoted \( \Lambda^{(1,0)}, \ldots, \Lambda^{(N,0)} \) from distribution \( F_0 \). Denote by \( F_0^N \) the empirical distribution of \( \Lambda^{(1,0)}, \ldots, \Lambda^{(N,0)} \) \( F_0^N \) is defined as

\[
F_0^N = \frac{1}{N} \sum_{j=1}^N \delta_{\Lambda_j^{(1,0)}}
\]

Updating: At time \( t \geq 1 \), we have \( N \) trajectories \( \Lambda^{(1,1)}, \ldots, \Lambda^{(N,1)} \); denote by \( F_{t-1}^N \) the empirical distribution of these \( N \) trajectories:

\[
F_{t-1}^N = \frac{1}{N} \sum_{j=1}^N \delta_{\Lambda_j^{(t-1)}} \tag{6}
\]
We compute, for \((i,j) \in \{1, \ldots, N\} \times \{1, \ldots, |Z|\}\) the weights
\[ w_{i,j} \propto Q_t \left( A(t-1), \{a_j\} \right). \]
We then draw \(N\) trajectories out of these \(N|Z|\) trajectories according to the weights \(w_{i,j}\). The sampling is done with replacement: the same trajectory can be included several times in the final updated sample. More precisely, we draw \((I_k, J_k) \in \{1, \ldots, N\} \times \{1, \ldots, |Z|\}\) such that, for all \((i,j) \in \{1, \ldots, N\} \times \{1, \ldots, |Z|\}\), the number of times \(N_{i,j}\) that the prolonged trajectory indexed by \((i,j)\) is duplicated in the updated sample, is
\[ N_{i,j} = \left\{ k \in \{1, \ldots, N\}, (I_k, J_k) = (i,j) \right\} \]
satisfy the two following conditions: \(\sum_{i=1}^N \sum_{j=1}^{|Z|} N_{i,j} = N\) and \(E\left\{ N_{i,j} \left| A(i,t-1), \ldots, A(N,t-1) \right. \right\} = N w_{i,j}\). Define the updated set of particles as
\[ A^{(k,t)} = A(i_{k,t-1}, a_{J_{k,t}}). \]
Define the empirical distribution \(P_t^N\) on \((Z, \mathcal{P}(Z))\) by
\[ P_t^N = \frac{1}{N} \sum_{k=1}^N \delta_{X(k,t)} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{|Z|} N_{i,j} \delta_{A(i,t-1), a_i}. \]
Note that, under the conditions stated above, the sampling is unbiased in the sense that
\[ E\left\{ P_t^N\left( A(1,t-1), \ldots, A(N,t-1) \right) \right\} = P_{t-1} \otimes Q_t. \]
The most obvious choice to satisfy the unbiasedness conditions consists in drawing the \(N\) i.i.d. independently conditionally to \((A(1,t-1), \ldots, A(N,t-1))\) according to a multinomial distribution with parameter \(w_{i,j}\) \((i,j) \in \{1, \ldots, N\} \times \{1, \ldots, |Z|\}\). In the literature, this is referred to as the simple random sampling. A better alternative consists in using residual sampling, as described in [10, Section 4].

In the sequel, we refer to the procedure outlined above as the global sampling algorithm.

3. JOINT CHANNEL EQUALIZATION AND SYMBOL DETECTION ON A FLAT RAYLEIGH-FADING CHANNEL

Consider a communication system signaling through a flat-fading channel with additive ambient noise. The input binary information bits \(\{Z_t\}_{t=0}^T\) are encoded using some channel code and combined with a symbol mapper, yielding complex data symbols \(\{S_t\}_{t=0}^T\), which take values from a finite alphabet \(S\). The input-output relationship of the flat-fading channel is described by
\[ Y_t = \alpha_S S_t + V_t \]
where \(Y_t, \alpha, S_t, V_t\) denote the received signal, the fading channel coefficient, the transmitted symbol and the additive noise at time \(t\), respectively. It is assumed in the sequel that
- the processes \(\{\alpha_t\}_{t=0}^T, \{S_t\}_{t=0}^T, \{V_t\}_{t=0}^T\) are mutually independent,
- the noise \(\{V_t\}_{t=0}^T\) is i.i.d. zero-mean complex gaussian \(V_t \sim N_c(0, \sigma_v^2)\),
- the fading \(\{\alpha_t\}_{t=0}^T\) is a zero-mean complex ARMA\((L, L)\) process, \(\alpha_t = \frac{\theta(B)}{\phi(B)} \eta_t\) where
\[ \phi(B) = 1 + \phi_1 B + \ldots + \phi_L B^L \]
\[ \theta(B) = \theta_0 + \theta_1 B + \ldots + \theta_L B^L. \]
and \(\{\eta_t\}\) is a white complex gaussian noise with zero-mean and unit-variance.

Define:
\[ X_t = \Phi X_{t-1} + g W_t \]
where \(\Phi = \text{the companion matrix of the polynomial } \phi(z)\) and \(g = (1, 0, \ldots, 0)^T\), i.e.
\[ \Phi = \begin{bmatrix} -\phi_1 & -\phi_2 & \cdots & -\phi_L & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \]
The fading coefficient can then be expressed as \(\alpha_t = \theta^T X_t\), where \(\theta = (\theta_0, \theta_1, \ldots, \theta_L)^T\) which implies the following state-space formulation
\[ X_t = S_t \theta^T X_{t-1} + V_t \]
\[ Y_t = S_t \theta^T X_{t-1} + V_t \]
This particular problem has been considered, among others by [13], [14], [15], [17]. To allow comparison with previously reported work, we consider the example studied in [14, section VIII]. In this example, the fading process is modelled by the output of a Butterworth filter of order \(L = 3\) whose cutoff frequency is 0.05, corresponding to a normalized Doppler frequency \(f_d T = 0.05\) with respect to the symbol rate \(1/T\), which is a fast fading scenario. We assume that a BPSK modulation is used, i.e. \(S_t \in \{-1, +1\}\) with differential encoding with no channel code.

The performance of the global sampling receiver has been compared with the following receiver schemes: (i) Known channel lower bound, (ii) Genie-aided lower bound (defined in [14, Sections IV & V]), (iii) Differential detector and (iv) the Mixture Kalman Filter (MKF) detector described in [14, Sections IV & V]. The MKF detector uses the Sequential Importance Sampling with Resampling (SISR) algorithm to draw samples in the indicator space and implements a Kalman filter for each trajectory in order to compute its trial sampling density and its importance weight. Resampling is performed when the ratio between the effective sample size defined in [14, Eq. (45)] and the actual sample size \(N\) is lower than a threshold \(\beta\).

Figure 3 shows the BER performance of each receiver versus the SNR. The SNR is defined as \(\text{var}(\alpha_t)/\text{var}(V_t)\) and the BER is obtained by averaging the error rate over \(10^5\) symbols. The BER performance of the global sampling receiver is shown for estimation delays \(\delta = 0\) and \(\delta = 1\). Also shown are the BER curves for the known channel lower bound, the genie-aided lower bound, the differential detector and the MKF detector with estimation delays \(\delta = 0\) and \(\delta = 1\) and resampling thresholds \(\beta = 0.1\) and \(\beta = 1\) (systematic resampling). The number of particles for both the global sampling receiver and the MKF detector is set to 50.

From this figure, it can be seen that with 50 particles, there is no significant performance difference between the proposed receiver...
and the MKF detector with the same estimation delay and $\beta = 0.1$ or $\beta = 1$, the global sampling receiver achieves essentially the genie-aided bound over the considered SNR.

Figure 3 shows the BER performance of the global sampling receiver versus the number of particles at SNR=20 dB and $\phi = 1$. Also shown in this figure are the BER performance for the MKF detector with $\beta = 0.1$ and $\beta = 1$, respectively. It can be seen from this plot that when the number of the particles is decreased from fifty to ten, the BER of the MKF receiver with $\beta = 0.1$ increases by 67% whereas the BER of the global sampling receiver increases by 11% only. In fact, figure 3 also shows that, for this particular example, the BER performance of the global sampling receiver is identical to the BER performance of an MKF with the same number of particles and a resampling threshold set to $\beta = 1$ (systematic resampling). This suggests that, contrary to what is usually argued in the literature [14], [11] systematic resampling of the particle seems to be, more robust when the number of particles is decreased.

4. REFERENCES

[1] G. Rigal, *Filtrage non-linéaire, résolution partielle et applications au traitement du signal*, Ph.D. thesis, Paul Sabatier, 1993.

[2] N. Shephard, “Partial non-Gaussian state space,” *Biometrika*, vol. 81, pp. 115–131, 1994.

[3] C. Carter and R. Kohn, “Markov chain Monte-Carlo in conditionally gaussian state space models,” *Biometrika*, vol. 83, no. 3, pp. 589–601, 1996.

[4] A. Doucet, A. Logothetis, and V. Krishnamurthy, “Stochastic sampling algorithms for state estimation of jump markov linear systems,” *IEEE Transactions on Automatic Control*, vol. 45, pp. 188–202, 2000.

[5] R. Chen and J. Liu, “Mixture Kalman filter,” *Journal Royal Statistical Society Series B*, vol. 62, no. 3, pp. 493–508, 2000.

[6] A. Doucet, N. Gordon, and V. Krishnamurthy, “Particle filters for state estimation of jump Markov linear systems,” *IEEE Trans on Signal Processing*, vol. 49, pp. 613–624, 2001.

[7] A. Doucet, N. d. Freitas, and N. Gordon, “An introduction to sequential monte carlo methods,” in *Sequential Monte-Carlo Methods in Practice*, A. Doucet Gordon, N.de Freitas, and N., Eds., pp. 3–13. New-York, 2001.

[8] H. Akashi and H. Kumamoto, “Random sampling approach to state estimation in switching environment,” *Automatica*, vol. 13, pp. 429–434, 1977.

[9] J. Tugnait, “Adaptive estimation and identification for discrete systems with markov jump parameters,” *IEEE Transactions on Automatic Control*, vol. 27, pp. 1054–1065, 1984.

[10] J. Liu and R. Chen, “Sequential Monte-Carlo methods for dynamic systems,” *Journal American Statistical Association*, vol. 93, pp. 1032–1044, 1998.

[11] A. Doucet, S. Godsill, and C. Andrieu, “On sequential Monte-Carlo sampling methods for bayesian filtering,” *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

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

[13] J. Liu and R. Chen, “Blind deconvolution via sequential imputations,” *Journal American Statistical Association*, vol. 430, pp. 567–576, 1995.

[14] R. Chen, X. Wang, and J. Liu, “Adaptive joint detection and decoding in flat-fading channels via mixture kalman filtering,” *IEEE Trans on Information Theory*, vol. 46, pp. 2079–2094, 2000.

[15] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, “Particle filtering for demodulation in fading channels with non-gaussian additive noise,” *IEEE Trans on Communications*, vol. 49, pp. 579–582, 2001.

[16] F. Bensalem, “Récepteur particulier pour canaux mobiles évanescents,” in *JDA'01 Journées Doctorales d'Automatique*, Toulouse, 2001.

[17] F. Bensalem, *Réception Particulaire Pour Canaux Multi-Trajets évanescents En Communication Radiomobile*, Ph.D. thesis, Toulouse, 2002.