FAST COMPUTATION OF SMOOTHED ADDITIVE FUNCTIONALS IN GENERAL STATE-SPACE MODELS

Cyrille Dubarry

TELECOM SudParis
Département CITI
Evry, France

Sylvain Le Corff

LTCI
TELECOM ParisTech
Paris, France

ABSTRACT

Approximating fixed-interval smoothing distributions using particle-based methods is a well-known issue in statistical inference when operating on general state-space hidden Markov models (HMM). In this paper we focus on the computation of path-space smoothed additive functionals. More precisely, this contribution provides new results on the forward filtering backward smoothing (FFBS) and the forward filtering backward simulation (FFBSi) algorithms. We prove that the L_q-mean error convergence rate of both algorithms depends on the number of observations T and the number of particles N only through the ratio T/N. We also derive non-asymptotic exponential deviation inequalities for these algorithms. The FFBS and FFBSi algorithms are compared when applied to parameter estimation in HMM.

***Index Terms*—** Sequential Monte Carlo methods, FFBS, FFBSi, additive functionals, Expectation Maximization.

1. INTRODUCTION

Let $\{X_t, Y_t\}_{t\geq 1}$ be a hidden Markov model taking values in $\mathbb{X}\times\mathbb{Y}$ where $\mathbb{X}$ and $\mathbb{Y}$ are general state-space endowed with countably generated σ-fields. The initial distribution of the Markov chain $\{X_t\}_{t\geq 1}$ is denoted by χ and its transition density by $m(x, x')$, with respect to a probability distribution λ on $\mathbb{X}$. For any $t \geq 1$, the conditional distribution of Y_t given X_t has a density denoted by $g(X_t, \cdot)$ with respect to a given measure on $\mathbb{Y}$.

For any integers $1 \leq s \leq t$ and $u \geq 1$ define the function $\phi_{s:t|u}$ such that

$$\mathbb{E}\left[H(X_{s:t})|Y_{1:u}\right] = \phi_{s:t|u}(H, Y_{1:u}) \ ,$$

where H is a bounded function on $\mathbb{X}^{t-s+1}$. Note that such a function exists by definition of the conditional expectation. In many instances, such as solving inverse problems in non-linear HMM, it is necessary to compute path-space smoothed functionals i.e. quantities of the form $\phi_{1:T|T}(H, y_{1:T})$ where $y_{1:T}$ is a fixed set of observations. In this paper we consider particle methods to approximate $\phi_{1:T|T}(H, y_{1:T})$ when the function H is of the form

$$\sum_{t=1}^{T} h_t(x_t) \ , \tag{1}$$

where $\{h_t\}_{t=1}^{T}$ is a family of bounded functions on $\mathbb{X}$.

Particle filters are efficient solutions to compute approximations of $\phi_{s:t|u}(H, y_{1:u})$. Many different implementations of the particle filters have been proposed in the literature with different computational costs; see [2, 6]. The FFBS algorithm has a computational cost per time step of order $\mathcal{O}\left(N^2\right)$, where N is the number of particles, while it is shown in [4] that the FFBSi algorithm can be implemented with a $\mathcal{O}\left(N\right)$ complexity per time step. In the case of additive functionals, [3] proposed a FFBS-based algorithm which can be implemented forward in time (see [7] for an application). L_q-mean error bounds have been derived for the FFBS algorithm: [3] shows that it is of order $T/\sqrt{N}$ for $q > 2$.

In this paper, we establish that the L_q-mean error convergence rate of both the FFBS and the FFBSi algorithms depends on T and N only through the ratio T/N. These results are obtained for $q \geq 2$ under weaker conditions than in [3]. We also derive exponential deviation inequalities. Section 2 is devoted to our theoretical contribution: we first briefly recall FFBS and FFBSi and then state our new results on the L_q-mean errors and exponential deviation inequalities. In Section 3, we compare these two procedures when applied to inference in non-linear HMM based on an Expectation Maximization (EM) algorithm.

2. RESULTS

The approximation of $\phi_{1:T|T}(H, y_{1:T})$ when H is of the form (1) is equivalent to the approximation of the marginal smoothing expectations $\phi_{t:t|T}(h_t, y_{1:T})$ for $t \in \{1, \cdots, T\}$. For both FFBS and FFBSi, the forward pass produces a particle approximation of the filtering distributions $\phi_{t:t|t}(\cdot, y_{1:t})$ and the backward pass computes an approximation of the smoothing distributions $\phi_{t:T|T}(\cdot, y_{1:T})$. We then deduce an approximation of the marginals $\phi_{t:t|T}(\cdot, y_{1:T})$. We now detail these algorithms.

2.1. Forward pass

The distribution $\phi_{t:t|t}(\cdot, y_{1:t})$ is approximated using weighted particles $\{(\xi_t^{N,i}, \omega_t^{N,i})\}_{i=1}^N$, $1 \leq t \leq T$ defined as follows. Let $\{\xi_1^{N,i}\}_{i=1}^N$ be i.i.d. (independent and identically distributed) random variables distributed according to the instrumental density ρ_1; set the unnormalized importance weights $\omega_1^{N,i} \stackrel{\text{def}}{=} \omega_1(\xi_1^{N,i})$, where $\omega_1(x) \stackrel{\text{def}}{=} \mathrm{d}\chi/\mathrm{d}\rho_1(x)\, g(x, y_1)$. Given $\{(\xi_{t-1}^{N,i}, \omega_{t-1}^{N,i})\}_{i=1}^N$, the sample $\{(\xi_t^{N,i}, \omega_t^{N,i})\}_{i=1}^N$ is obtained by considering the auxiliary particle filter (see [13, 1]): pairs $\{(I_t^{N,i}, \xi_t^{N,i})\}_{i=1}^N$ of indices and particles are simulated from the instrumental distribution defined by:

$$\pi_{t|t}(i, \mathrm{d}x) \propto \omega_{t-1}^{N,i}\, p_t(\xi_{t-1}^{N,i}, x)\, \lambda(\mathrm{d}x) , \tag{2}$$

on the product space $\{1, \ldots, N\} \times \mathbb{X}$, where p_t is an (unnormalized) transition kernel. For any $i = 1, \ldots, N$ we associate to the particle $\xi_t^{N,i}$ its unnormalized importance weight defined by $\omega_t^{N,i} \stackrel{\text{def}}{=} \omega_t(\xi_{t-1}^{N,I_t^{N,i}}, \xi_t^{N,i})$, where

$$\omega_t(x, x') \stackrel{\text{def}}{=} \frac{m(x, x')g(x', y_t)}{p_t(x, x')} .$$

Keeping the particles genealogy obtained in this forward pass provides an approximation of the smoothing distributions $\phi_{1:T|T}$. This procedure is called the *geneaogical tree* algorithm.

2.2. Backward pass of the FFBS algorithm

From the weighted samples $\{(\xi_t^{N,i}, \omega_t^{N,i})\}_{i=1}^N$, $1 \leq t \leq T$, an approximation $\phi_{t:T|T}^{\mathrm{FFBS},N}$ of the smoothing distribution $\phi_{t:T|T}(\cdot, y_{1:T})$ is constructed recursively backward using

$$\phi_{t:T|T}^{\mathrm{FFBS},N}(H) = \int \mathrm{B}_{\phi_t^N}(x_{t+1}, \mathrm{d}x_t)\phi_{t+1:T|T}^{\mathrm{FFBS},N}(\mathrm{d}x_{t+1:T})H(x_{t:T}) ,$$

where, for all $x \in \mathbb{X}$ and all function $h \geq 0$ on $\mathbb{X}$

$$\mathrm{B}_{\phi_t^N}(x, h) \stackrel{\text{def}}{=} \sum_{i=1}^N \frac{\omega_t^{N,i} m(\xi_t^{N,i}, x)}{\sum_{\ell=1}^N \omega_t^{N,\ell} m(\xi_t^{N,\ell}, x)} h\left(\xi_t^{N,i}\right) ,$$

is a particle approximation of the backward smoothing kernel. From these two equations, we deduce a particle approximation of $\phi_{t:t|T}(\cdot, y_{1:T})$, see [5],

$$\sum_{i=1}^N \omega_{t|T}^{N,i}\, \delta_{\xi_t^{N,i}} ,$$

where the importance weights $\omega_{t|T}^{N,i}$ are updated recursively as follows: for all $i \in \{1, \ldots, N\}$,

$$\omega_{T|T}^{N,i} \stackrel{\text{def}}{=} \frac{\omega_T^{N,i}}{\Omega_T} , \quad \Omega_T \stackrel{\text{def}}{=} \sum_{i=1}^N \omega_T^{N,i} ,$$

$$\omega_{t|T}^{N,i} \stackrel{\text{def}}{=} \sum_{j=1}^N \omega_{t+1|T}^{N,j} \frac{\omega_t^{N,i} m(\xi_t^{N,i}, \xi_{t+1}^{N,j})}{\sum_{\ell=1}^N \omega_t^{N,\ell} m(\xi_t^{N,\ell}, \xi_{t+1}^{N,j})} , \ 1 \leq t < T .$$

2.3. Backward pass of the FFBSi algorithm

Consider, for $t \in \{1, \ldots, T-1\}$, the Markov transition matrix $\{\Lambda_t^N(i,j)\}_{i,j=1}^N$ over the state-space $\{1, \ldots, N\}$ given, for all $(i,j) \in \{1, \ldots, N\}^2$, by

$$\Lambda_t^N(i,j) \stackrel{\text{def}}{=} \frac{\omega_t^{N,j} m(\xi_t^{N,j}, \xi_{t+1}^{N,i})}{\sum_{\ell=1}^N \omega_t^{N,\ell} m(\xi_t^{N,\ell}, \xi_{t+1}^{N,i})} . \tag{3}$$

The transition probabilities defined in (3) induce an inhomogeneous Markov chain $\{J_u\}_{u=1}^T$ evolving backward in time with a joint distribution given, for $j_{1:T} \in \{1, \ldots, N\}^T$, by

$$\mathbb{P}\left[J_{1:T} = j_{1:T} \,\middle|\, \mathcal{F}_T^N\right] = \frac{\omega_T^{N,j_T}}{\Omega_T} \prod_{t=1}^{T-1} \Lambda_t^N(j_{t+1}, j_t) , \tag{4}$$

where $\mathcal{F}_T^N \stackrel{\text{def}}{=} \sigma\left\{(\xi_s^{N,i}, \omega_s^{N,i})_{i=1}^N; 1 \leq s \leq T\right\}$. FFBSi approximates the smoothing distribution $\phi_{1:T|T}$ by

$$\phi_{1:T|T}^{\mathrm{FFBSi},N}(H) \stackrel{\text{def}}{=} N^{-1} \sum_{\ell=1}^N H\left(\xi_1^{N,J_1^\ell}, \ldots, \xi_T^{N,J_T^\ell}\right) , \tag{5}$$

where $\{J_{1:T}^\ell\}_{\ell=1}^N$ are N paths drawn independently (given $\mathcal{F}_T^N$) according to (4). The approximation (5) was introduced in [10, Algorithm 1, p.158]. [4] proposes an implementation that reduces the computational complexity from $\mathcal{O}(N^2T)$ to $\mathcal{O}(NT)$ using a specific form of acceptance-rejection method. This version of the FFBSi algorithm is used in Section 3 to illustrate the new results provided in this contribution.

2.4. Bounds on the smoothing errors

Consider the following assumptions

(A1) a) $0 < g(\cdot, y_t) \leq \sup_{1\leq t\leq T} \sup_x g(x, y_t) < +\infty$.
b) $\sup_{1\leq t\leq T, x\in\mathbb{X}} \int p_t(x, x')\lambda(\mathrm{d}x') < +\infty$.
c) $\sup_{2\leq t\leq T} \sup_{x,x'} \omega_t(x, x') + \sup_x \omega_1(x) < +\infty$.

(A2) a) There exist $0 < \sigma_- < \sigma_+ < +\infty$ such that for any $(x, x') \in \mathbb{X}^2$, $\sigma_- \leq m(x, x') \leq \sigma_+$.
b) There exists $c_- > 0$ s.t. $\int \chi(\mathrm{d}x)g(x, y_1) \geq c_-$ and

$$\inf_{2\leq t\leq T} \inf_x \int m(x, x')g(x', y_t)\lambda(\mathrm{d}x') \geq c_- .$$

For the FFBS algorithm, the smoothing error can be decomposed into a sum of two terms. The first one is a martingale whose L_q-mean error is upper-bounded by $\sqrt{T/N}$ and the L_q-mean error of the second one is bounded by T/N. For the exponential deviation inequalities of the FFBS algorithm, the martingale term can be dealt with using the Azuma-Hoeffding inequality while the second term needs a specific exponential deviation inequality for ratios of random variables. Finally, the difference between the FFBS and the FFBSi approximations can be rewritten as a martingale, thus yielding the results for the FFBSi. The proof of Theorem 1 is given in [8]. Define the FFBS and the FFBSi smoothing errors by

$$\Delta_T^{\mathrm{algo},N}(H, y_{1:T}) \stackrel{\mathrm{def}}{=} \phi_{1:T|T}(H, y_{1:T}) - \phi_{1:T|T}^{\mathrm{algo},N}(H, y_{1:T}) ,$$

where algo is FFBS or FFBSi.

Theorem 1. *Let H be a function of the form (1) for some bounded functions $\{h_t\}_{t=1}^T$. Assume A1–2. Then, for all $q \geq 2$, there exists a constant C s.t. for all T, $N \in \mathbb{N}^\star$ and $\varepsilon > 0$:*

$$\left\| \Delta_T^{\mathrm{algo},N}(H, y_{1:T}) \right\|_q \leq C\sigma_T \left(\sqrt{\frac{T}{N}} + \frac{T}{N} \right) ,$$

$$\mathbb{P}\left\{ \left| \Delta_T^{\mathrm{algo},N}(H, y_{1:T}) \right| > \varepsilon \right\} \leq 2\exp\left(-\frac{CN\varepsilon^2}{\sigma_T^2 T} \right) + 8\exp\left(-\frac{CN\varepsilon}{\sigma_T T} \right) ,$$

where $\sigma_T \stackrel{\mathrm{def}}{=} \max_{1 \leq t \leq T} \{\mathrm{osc}(h_t)\}$ and algo is FFBS or FFBSi.

3. APPLICATIONS

3.1. Linear Gaussian model

The performance of the FFBS and FFBSi algorithms - implemented resp. as in [3] and [4] and referred to as Forward-FFBS and Fast-FFBSi - are compared when applied to the estimation of $\mathcal{I}_T \stackrel{\mathrm{def}}{=} T^{-1} \sum_{t=1}^T \mathbb{E}\left[X_t | Y_{1:T}\right]$. It is illustrated with the following linear Gaussian model:

$$X_{t+1} = \phi X_t + \sigma_u U_t , \qquad Y_t = X_t + \sigma_v V_t ,$$

where $X_1 \sim \mathcal{N}\left(0, \frac{\sigma_u^2}{1-\phi^2}\right)$, $\{U_t\}_{t\geq 1}$ and $\{V_t\}_{t \geq 1}$ are independent sequences of i.i.d. standard gaussian random variables (independent of X_1). The parameters $(\phi, \sigma_u, \sigma_v)$ are assumed to be known. Data were generated using the model with $\phi = 0.9$, $\sigma_u = 0.6$ and $\sigma_v = 1$.

Table 1 provides the empirical variance of the estimation of $\mathcal{I}_T$ given by the genealogical tree and the FFBSi methods over 250 independent Monte Carlo experiments. We display in Figure 1 the empirical variance for different values of N as a function of T for both estimators. These estimates are represented by dots and a linear regression (resp. quadratic regression) is also provided for the FFBSi algorithm (resp. for the genealogical tree method).

Genealogical tree

T \ N	300	500	750	1000	1500
300	137.8	119.4	63.7	46.1	36.2
500	290.0	215.3	192.5	161.9	80.3
750	474.9	394.5	332.9	250.5	206.8
1000	673.7	593.2	505.1	483.2	326.4
1500	1274.6	1279.7	916.7	804.7	655.1

FFBSi

T \ N	300	500	750	1000	1500
300	5.1	3.1	2.3	1.4	1.0
500	9.7	5.1	3.7	2.6	2.2
750	11.2	7.1	4.9	3.7	2.6
1000	16.5	10.5	6.7	5.1	3.4
1500	25.6	14.1	7.8	6.8	5.1

Table 1: Empirical variance for different values of T and N.

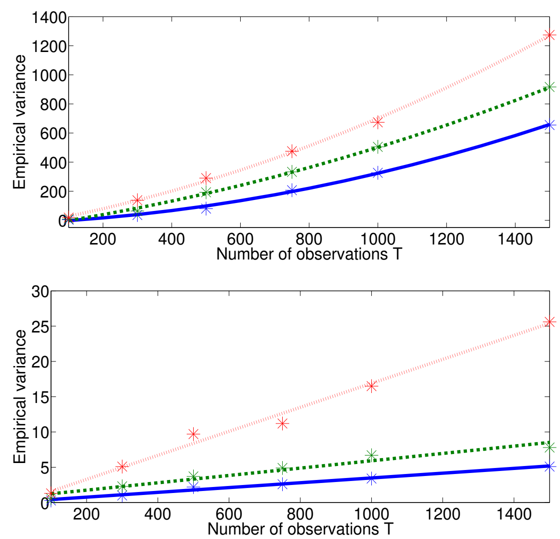

Fig. 1: Empirical variance for $N = 300$ (dotted line), $N = 750$ (dashed line) and $N = 1500$ (bold line).

3.2. EM-based inference in non-linear HMM

Inference in non-linear HMM can be solved by EM-based algorithms: the E-step is in general, not explicit, and has to be approximated by Monte Carlo methods [1]. In HMM, when the complete likelihood belongs to the exponential family, the E-step consists in computing $\phi_{1:T|T}(H, y_{1:T})$ where H is of the form $\sum_{t=1}^{T-1} h_t(x_t, x_{t+1})$. Note that if the state transition density m does not depend on the parameter of interest, H takes the form (1). In the example below, we approximate the E-step by using the FFBSi algorithm. Using Theorem 1 and the results in [9, section 5], the convergence of this FFBSi-EM algorithm to the same limit points as EM is addressed in [11].

The Simultaneous Localization And Mapping (SLAM) problem arises when a robot seeks to acquire a map estimate and at the same time wishes to localize itself. We illustrate the good performance of the Fast-FFBSi algorithm for SLAM.

The robot pose is represented by $X_t = (X_{t,1}, X_{t,2}, X_{t,3})^T$ where $(X_{t,1}, X_{t,2})$ are the robot's cartesian coordinates and $X_{t,3}$ is its heading direction. The controls $\mathbf{u_t} = (v_t, \psi_t)^T$ given to the robot are supposed to be known; ψ_t is the robot's steering angle and v_t its velocity. The state transition model can be written:

$$X_t = X_{t-1} + \begin{pmatrix} (v_t + \epsilon_{t,1})\, d_t \cos(X_{t-1,3} + (\psi_t + \epsilon_{t,2})) \\ (v_t + \epsilon_{t,1})\, d_t \sin(X_{t-1,3} + (\psi_t + \epsilon_{t,2})) \\ (v_t + \epsilon_{t,1})\, d_t \frac{\sin((\psi_t + \epsilon_{t,2}))}{B} + \epsilon_{t,3} \end{pmatrix}$$

where B is the robot's wheelbase, $(\epsilon_{t,1}, \epsilon_{t,2}, \epsilon_{t,3}) \sim \mathcal{N}_3(0, \Sigma)$ with known 3×3 covariance matrix Σ and d_t is the time between two successive poses. The robot's environment is represented by a 2-dimensional map where landmarks are selected in the robot's neighborhood. The total number of landmarks q is assumed to be known. At each time step the robot is able to determine which landmarks are observed. Let $\theta_{\cdot,j}$ (in $\mathbb{R}^2$) be the cartesian coordinates of the $j-$th landmark.

At time t, the perceives q_t landmarks and for each observed landmark i, the observation model relates the measurement $Y_{t,i}$ to the robot pose:

$$Y_{t,i} = \begin{pmatrix} \sqrt{(\theta_{1,i} - X_{t,1})^2 + (\theta_{2,i} - X_{t,2})^2} \\ \arctan \frac{\theta_{2,i} - X_{t,2}}{\theta_{1,i} - X_{t,1}} - X_{t,3} \end{pmatrix} + \delta_{t,i} \ ,$$

where the noise vectors $\left(\delta_{t,i}\right)_{t,i} \overset{\text{i.i.d}}{\sim} \mathcal{N}_2(0, R)$ with known 2×2 covariance matrix R. Since the state transition density does not depend on the map, H is of the form (1). To make the exposition short and easier, we consider a quite simple model for SLAM; more realistic models (relaxing e.g. some conditions on the association errors or on the Gaussian noise) are considered in [12].

Data are generated with $R = \mathrm{diag}(\sigma_r^2, \sigma_b^2)$ where $\sigma_r = 0.5$m and $\sigma_b = \frac{\pi}{60}$rad. The controls $\mathbf{u_t}$ were given by a set of waypoints, $B = 2$m and the robot path was sampled using $\Sigma = \mathrm{diag}(\sigma_v^2, \sigma_\phi^2, \sigma_\psi^2)$ with $\sigma_v = 0.6\text{m.s}^{-1}$, $\sigma_\phi = 10^{-2}$rad and $\sigma_\psi = \frac{\pi}{30}$rad. Landmarks positions are given in Figure 2. The map is estimated using FFBSi-EM based on $T = 104$ observations obtained along the path of the robot; see figure 2. At each iteration, the observation model is linearized so that the complete-data log-likelihood belongs to the exponential family. When a landmark is seen for the first time, its position is initialized with the received observation and the estimated pose (empirical mean of the particles). All runs were performed using $N = 100$ particles. For the forward pass of Fast-FFBSi, p_t is a Gaussian approximation of the optimal proposal kernel (i.e. the distribution of X_t given the previous pose, the last control and the last observation). Figure 2 displays the mean error estimate over all landmarks (Euclidian distance between the true position and the estimated one) as a function of the FFBSi-EM iterations. Plots are based on 50 independent runs summarized by the median (bold line) and by the upper and lower quartiles (dotted lines).

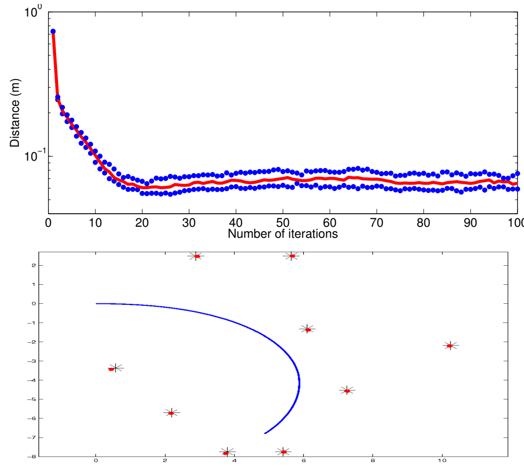

Fig. 2: [top] Evolution of the error for all landmarks. [bottom] True map (star), Estimated map (circle), and the robot path.

4. REFERENCES

[1] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.

[2] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[3] P. Del Moral, A. Doucet, and S. Singh. A Backward Particle Interpretation of Feynman-Kac Formulae. *ESAIM M2AN*, 44(5):947–975, 2010.

[4] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *accepted for publications in Ann. Appl. Probab.*, 2010.

[5] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, 10:197–208, 2000.

[6] A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. *Oxford handbook of nonlinear filtering*, 2009.

[7] A. Doucet, G. Poyiadjis, and S.S. Singh. Particle approximations of the score and observed information matrix in state-space models with application to parameter estimation. *Biometrika*, 2010.

[8] C. Dubarry and S. Le Corff. Non-asymptotic deviation inequalities for smoothed additive functionals in non-linear state-space models with applications to parameter estimation. *Submitted*, 2010.

[9] G. Fort and E. Moulines. Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.*, 31(4):1220–1259, 2003.

[10] S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for non-linear time series. *J. Am. Statist. Assoc.*, 50:438–449, 2004.

[11] S. Le Corff and G. Fort. Online EM based algorithms for inference in Hidden Markov Models. Work in progress, 2011.

[12] S. Le Corff, G. Fort, and E. Moulines. Online EM algorithm to solve the SLAM problem. Submitted, 2011.

[13] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.*, 94(446):590–599, 1999.