

The Monte-Carlo method for filtering with discrete-time observations

Pierre Del Moral*, Jean Jacod† and Philip Protter‡

September 28, 1999

1 Introduction

1) The problem which we address in this work is the following: we have an \mathbb{R}^d -valued state process $X = (X_t)_{t \geq 0}$ which evolves according to a stochastic differential equation of the form

$$dX_t = a(X_t)dt + b(X_t)dW_t, \quad \mathcal{L}(X_0) = \mu, \quad (1.1)$$

where μ is a known distribution on \mathbb{R}^d , and a, b are known functions, and W is a d -dimensional Wiener process. We have noisy observations Y_1, \dots, Y_N at N regularly spaced times, and without loss of generality we will assume that these times are $1, \dots, N$. We wish to compute the conditional expectations

$$\pi_{Y,N}f = E(f(X_N)|Y_0, \dots, Y_N) \quad (1.2)$$

for all reasonable functions f on \mathbb{R}^d .

This is a filtering problem, but although X evolves according to “continuous” time this problem is essentially discrete in time, and the integer N , although perhaps large, is fixed throughout (in contrast with the usual continuous-time non-linear filtering problem).

We will consider below three different ways in which the noisy observations occur. But in all cases, we have no explicit form for the transition semigroup $(Q_t)_{t \geq 0}$ of the Markov process X , so that an explicit form for the filter $\pi_{Y,N}$ is not available, and we will approximate this filter by Monte-Carlo simulations.

The performance is measured in terms of how many “single” random variables are necessary to simulate in order to achieve a given error in our approximation. More precisely, we count as a “single” variable any variable of a fixed dimension which we have to

*LSP, Bât. 1R1, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France (CNRS UMR C5583)

†Laboratoire de Probabilités, Université Paris 6, 4 place Jussieu, 75252 Paris, France (CNRS UMR 7599)

‡Mathematics and Statistics Departments, Purdue University, W.Lafayette, IN 47907-1395, USA. Supported in part by ONR grant # N00014-96-1-0262 and NSF grant # 9401109-INT

simulate. Then if we are allowed to simulate r single variables, we try to find a Monte-Carlo algorithm giving an estimate $\hat{\Pi}_{Y,N}^r f$ for $\pi_{Y,N} f$ in such a way that the sequence $r^{1/\alpha}(\hat{\Pi}_{Y,N}^r f - \pi_{Y,N} f)$ has bounded moments of some order (these are random variables defined on the space on which the simulated variables are defined themselves), for some $\alpha > 0$ as small as possible: it means that if we wish to have a (mean) error less than ε we need a constant times $(1/\varepsilon)^\alpha$ simulated variables.

As said before, we wish to have α as small as possible, together with the following important properties:

- 1) α does not depend on the number N of observations (of course the bounds on the moments will depend on N , and in an exponential way as a matter of fact: so when N is big the method does not work well).
- 2) α does not depend on the observed points (Y_1, \dots, Y_N) , nor on the bounded measurable function f .

2) We single out three observations schemes:

Case A) For all $i = 1, \dots, N$ we have

$$Y_i = h(X_i, \varepsilon_i), \quad (1.3)$$

where the ε_i are i.i.d. q' -dimensional variables, independent of X and with a law having a known density, and h is a known function from $\mathbb{R}^d \times \mathbb{R}^{q'}$ into \mathbb{R}^q .

Case B) The observed values Y_i are the values at times i of an \mathbb{R}^q -valued process $(Y_t)_{t \geq 0}$ which satisfies the following:

$$dY_t = h(X_t)dt + \sigma dW'_t, \quad Y_0 = 0, \quad (1.4)$$

where h is a known function and W' is a q -dimensional Wiener process, independent of W , and σ is an invertible $q \times q$ -matrix.

Case C) The same situation as Case B, except that the equation is

$$dY_t = a'(X_t, Y_t)dt + b'(X_t, Y_t)dW'_t, \quad Y_0 = 0, \quad (1.5)$$

where a' and b' are known functions and W' is as above.

In Cases B and C the initial value $Y_0 = 0$ is just for convenience, and for homogeneity of notation we also set $Y_0 = 0$ in case A. In the three cases, the two sequences $(X_i)_{i \in \mathbb{N}}$ and $(X_i, Y_i)_{i \in \mathbb{N}}$ are Markov chains, with transitions denoted by Q and P respectively (we have $Q(x, dx') = P(x, y; dx', \mathbb{R}^q)$ for all y).

Let us now state the assumptions which will be in force in the paper. We do not seek the weakest possible assumptions here. First, in case (A) and (C) (resp. (B)) we will suppose (E1) (resp. (E2)) below about Equation (1.1):

(E1) The functions a and b are 2 times differentiable with bounded derivatives of all orders up to 2, and the matrix $(bb^*)(x)$ is uniformly non-degenerate.

(E2) The functions a and b are bounded and Lipschitz continuous.

For case (C) we also need an assumption on Equation (1.5), which is:

(E3) The functions a' and b' are 2 times differentiable with bounded derivatives of all orders up to 2, and the matrix $(b'b'^*)(x)$ is uniformly non-degenerate.

For case (A) we also need an assumption on the function h and on the i.i.d. variables ε_i :

(A1) For any $x \in \mathbb{R}^d$ the variable $Y = h(x, \varepsilon_1)$ admits a density $y \mapsto g(x, y)$ and the function g is bounded and *explicitly known*.

Observe that (A1) is satisfied as soon as $Y_i = H(X_i) + \varepsilon_i$ and the ε_i 's have a bounded known density, whatever the known function H is.

For case (B) we need as well an assumption on the the function h occuring in (1.4) and on σ :

(B1) The function h is bounded and Lipschitz continuous, and the matrix σ is invertible.

We will exhibit Monte-Carlo algorithms with the following performances. Below, the constant B_N depends on the quantities a, b, a', b', h, σ involved in the equations, but not on the observed value $(Y_1, \dots, Y_N) = (y_1, \dots, y_N)$ except through a minoration of the continuous version of the density of the variable (Y_1, \dots, Y_N) at the point (y_1, \dots, y_N) (see (2.14) below; the assumptions made ensure the existence of this continuous density). As indicated by our notation, B_N also depends in a crucial way on the number N of steps, and in fact the dependency is exponential: we have $B_N = CC'^N$ for two constants $C > 0$ and $C' > 1$.

The main results are as follows:

Case A: Under (E1) and (A1) we have $E(|\hat{\Pi}_{Y,N}^r f - \pi_{Y,N} f|) \leq \frac{B_N \|f\|}{r^{1/3}}$, where $\|f\|$ is the sup-norm of f .

Case B: Under (E2) and (B1) we have $E(|\hat{\Pi}_{Y,N}^r f - \pi_{Y,N} f|) \leq \frac{B_N \|f\|'}{r^{1/4}}$, where $\|f\|'$ is the Lipschitz norm of f , supposed to be a bounded Lipschitz function.

Case C: Under (E1) and (E3) we have $E(|\hat{\Pi}_{Y,N}^r f - \pi_{Y,N} f|) \leq \frac{B_N \|f\|}{r^{1/(3+q)}}$.

Together with these we will also obtain exponential bounds.

Note that B is a particular case of C, provided (E1) holds and h is of class C^2 : in this case we can also apply the result of case C, thus obtaining a majoration involving $\|f\|$ instead of $\|f\|'$. When $q = 1$ the rate is then the same when f is Lipschitz or simply bounded measurable, but this is no longer true when $q \geq 2$.

3) The paper is organized as follows: Some preliminaries are provided in Section 2. Sections 3, 4 and 5 are devoted to the proofs of the above results in cases A, B and C respectively. In Section 6 we first state some extensions, without proofs. Then we give some numerical results: indeed the theory gives us constants B_N above with the form $B_N = C^N$ for some $C > 1$, which seems to indicate that the method is unfeasible in practice even for N moderately large. To understand what these constants B_N really are, we have made a numerical study in a very simple case: (X, Y) is a 2-dimensional Gaussian diffusion; the simulations show that B_N stays reasonably small when the diffusion is recurrent, while it explodes as N increases in the transient case: this is of course to be expected, although we do not have a proof for this fact.

2 Preliminaries

This section is devoted to recalling some known facts about stochastic differential equations, and to provide some preliminary results. As a rule in this paper, C and C' will denote constants which may depend on the coefficients of the equations (1.1), (1.3), (1.4) and possibly on the function g in (A1), but on nothing else, and they change from line to line.

2-1) Since we cannot simulate exactly a random variable having the same law as X_1 (starting at any $X_0 = x$), we will approximate it by the well known Euler scheme. More precisely, for any integer $m \geq 1$ and any starting point x , we define recursively:

$$X(x, m)_0 = x, \quad X(x, m)_{i+1} = X(x, m)_i + \frac{1}{m}a(X(x, m)_i) + \frac{1}{\sqrt{m}}b(X(x, m)_i)U_i, \quad (2.1)$$

where the variables U_i are i.i.d. d -dimensional centered Gaussian with unit covariance matrix.

Under (E1), if $Q^{(m)}(x, \cdot)$ denotes the law of $X(x, m)_m$, then Q and $Q^{(m)}$ have densities q and $q^{(m)}$ satisfying

$$q(x, x') + q^{(m)}(x, x') \leq Ce^{-C'|x-x'|^2}, \quad (2.2)$$

$$|x - x'| > \frac{2}{m} \Rightarrow |q(x, x') - q^{(m)}(x, x')| \leq \frac{C}{m}e^{-C'|x-x'|^2}. \quad (2.3)$$

(2.2) is a standard fact under (E1), and (2.3) is proved by Bally and Talay [1], [2] (since bb^* is assumed to be uniformly elliptic we do not need to consider the ‘‘perturbed’’ scheme introduced in [2], and the C^2 -regularity is enough). This implies in particular that for any bounded Borel function f we have for another constant still denoted by C :

$$|Q^{(m)}f(x) - Qf(x)| \leq \frac{C}{m}\|f\|. \quad (2.4)$$

Next, if we simply assume (E2) and if we denote by $X(x)$ the solution to (1.1) starting at $X_0 = x$ and if we take $U_i = \sqrt{m}(W_{\frac{i+1}{m}} - W_{\frac{i}{m}})$ in (2.1) with the same Wiener process W as in (1.1), we have (see e.g. [3]):

$$E\left(\sup_{1 \leq i \leq m} |X(x, m)_i - X(x)_{\frac{i}{m}}|^2\right) \leq \frac{C}{m} \quad (2.5)$$

Under the same assumption (E2) we also have

$$E(\sup_{s \leq 1} |X(x)_s - X(y)_s|) + \sup_m E(\sup_{1 \leq i \leq m} |X(x, m)_i - X(y, m)_i|) \leq C|x - y|. \quad (2.6)$$

2-2) In order to study case C, we also need to perform an Euler scheme for Equation (1.5). In fact we do this simultaneously for both equations (1.1) and (1.5): starting at $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^q$, we define $X(x, m)_i$ by (2.1), and set

$$\left. \begin{aligned} Y(x, y, m)_0 &= y, & Y(x, y, m)_{i+1} &= Y(x, y, m)_i + \\ &+ \frac{1}{m} a'(X(x, m)_i, Y(x, y, m)_i) + \frac{1}{\sqrt{m}} b'(X(x, m)_i, Y(x, y, m)_i) U_i', \end{aligned} \right\} \quad (2.7)$$

where the variables U_i' are i.i.d. q -dimensional centered Gaussian with unit covariance matrix. We clearly have the same properties for the pair $(X(x, m)_i, Y(x, y, m)_i)$ as we had above for $X(x, m)_i$. In particular under (E1) and (E3), the pair (X, Y) satisfies an equation of type (1.1) whose coefficients satisfy (E1) on \mathbb{R}^{d+q} . Therefore if $P^{(m)}(x, y; \cdot)$ denotes the law of $(X(x, m)_m, Y(x, y, m)_m)$, the transitions P and $P^{(m)}$ have densities p and $p^{(m)}$ satisfying

$$p(x, y; x', y') + p^{(m)}(x, y; x', y') \leq C e^{-C'(|x-x'|^2 + |y-y'|^2)}, \quad (2.8)$$

$$|x-x'| + |y-y'| > \frac{2}{m} \Rightarrow |p(x, y; x', y') - p^{(m)}(x, y; x', y')| \leq \frac{C}{m} e^{-C(|x-x'|^2 + |y-y'|^2)}. \quad (2.9)$$

2-3) We fix the observation sequence $(y_0 = 0, y_1, y_2, \dots)$, and instead of $\pi_{Y, N}$ in (1.2) we simply write $\pi_n f$.

Our assumptions imply that there is a function G on $\mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^d \times \mathbb{R}^q$ such that $y' \mapsto G(X_i, Y_i; X_{i+1}, y')$ is a version of the density of the law of Y_{i+1} , conditional on (X_i, Y_i, X_{i+1}) : this is trivial in case A with $G(x, y; x', y') = g(x', y')$ (see (A1)), and in case C with $G(x, y; x', y') = p(x, y; x', y')/q(x, x')$. In case B, if φ is the density of the normal law $\mathcal{N}(0, \sigma\sigma^*)$ on \mathbb{R}^q (recall that σ is invertible) and if $k(x, x'; du)$ denotes a version of the law of the variable $\int_0^1 h(X_s) ds$ conditionally on $X_0 = x$ and $X_1 = x'$, then the above property is satisfied with the function $G(x, y; x', y') = \int \varphi(y' - y - u) k(x, x', du)$.

Now, a version of $\pi_N f$ is given by

$$\pi_N f = \frac{\int \mu(dx_0) Q(x_0, dx_1) G(x_0, y_0; x_1, y_1) \dots Q(x_{N-1}, dx_N) G(x_{N-1}, y_{N-1}; x_N, y_N) f(x_N)}{\int \mu(dx_0) Q(x_0, dx_1) G(x_0, y_0; x_1, y_1) \dots Q(x_{N-1}, dx_N) G(x_{N-1}, y_{N-1}; x_N, y_N)}. \quad (2.10)$$

There is a recursive way to write the “filter” π_N , which is as follows. First, we clearly have

$$\pi_0 f = \mu(f) = \int f(x) \mu(dx). \quad (2.11)$$

Next, we introduce the kernels H_j from \mathbb{R}^d into itself by

$$H_j f(x) = \int Q(x, dx') G(x, y_{j-1}, x', y_j) f(x') dx'. \quad (2.12)$$

Then π_N is also given by the following recursive formula, starting with (2.11):

$$\pi_N f = \frac{\pi_{N-1} H_N f}{\pi_{N-1} H_N 1}. \quad (2.13)$$

Note that the denominator of (2.10) is $\mu H_1 \dots H_N 1$, and it is of course natural to assume that it is not equal to 0 (it is the value of the density of (Y_1, \dots, Y_N) at the observed values (y_1, \dots, y_N)). We give a name to this number:

$$\varepsilon = \mu H_1 \dots H_N 1. \quad (2.14)$$

Next, as seen before the function G is bounded in case A (use (A1)) and B (because φ is bounded); in case C we have $G(x, y; x', y') \leq C/q(x, x')$ by (2.8). Then (2.12) implies that there exists a constant K such that

$$H_j 1(x) \leq K \quad \forall x \in \mathbb{R}^d, \quad j = 1, \dots, N. \quad (2.15)$$

Then (2.14) and (2.15) yield

$$\mu H_1 \dots H_j 1 \geq \frac{\mu H_1 \dots H_{j+1} 1}{K} \geq \dots \geq \frac{\varepsilon}{K^{N-j}}.$$

By applying (2.15) several times we also get $\mu H_1 \dots H_{j-1} 1 \leq K^{j-2}$. Then since $\pi_j f = \frac{\mu H_1 \dots H_j f}{\mu H_1 \dots H_j 1}$ by (2.13), we readily deduce the following estimates (for the right side, use the fact that π_{j-1} is a transition measure and (2.15) again):

$$\frac{\varepsilon}{K^{N-1}} \leq \pi_{j-1} H_j 1 \leq K. \quad (2.16)$$

2-4. The approximate filter. We will replace the true filter (given by formula (2.13) for example) by an approximate filter constructed via the Euler approximation mentioned above. More precisely, we replace the kernel H_j of (2.12) by another kernel H_j^m such that for any bounded function f ,

$$\|H_j^m f - H_j f\| \leq \beta_m \|f\| \quad (2.17)$$

for some $\beta_m > 0$ which will go to 0 as $m \rightarrow \infty$ (this kernel H_n^m will be related to the Euler approximation of stepsize $1/m$ in a way which will depend on the cases A, B or C). Then we define the probability measures π_n^m by induction on n as follows:

$$\pi_0^m = \pi_0 = \mu, \quad \pi_N^m f = \frac{\pi_{N-1}^m H_N^m f}{\pi_{N-1}^m H_N^m 1}. \quad (2.18)$$

The quality of this approximation of π_n is provided by the following result:

Proposition 2.1 *Assume (2.17). For all bounded Borel functions f on \mathbb{R}^d and all $j = 1, \dots, N$, we have*

$$|\pi_j^m f - \pi_j f| \leq \beta_m \|f\| A_j, \quad (2.19)$$

where

$$A_j = \frac{\rho^{j+1} - \rho}{K(\rho - 1)}, \quad \rho = 2 \frac{K^N}{\varepsilon}. \quad (2.20)$$

Proof. a) We will prove (2.19) by induction on j . For $j = 0$ it is trivial with $A_0 = 0$, so we assume that it holds for $j - 1$. We may write

$$\pi_j^m f - \pi_j f = \frac{1}{\pi_{j-1} H_j 1} \left((\pi_{j-1}^m H_j^m f - \pi_{j-1} H_j f) + (\pi_j^m f)(\pi_{j-1} H_j 1 - \pi_{j-1}^m H_j^m 1) \right).$$

Since π_{j-1}^m and π_j^m are probability measures, we obtain by (2.17):

$$\begin{aligned} |\pi_{j-1}^m H_j^m f - \pi_{j-1} H_j f| &\leq \beta_m \|f\|, \\ |\pi_j^m f| &\leq \|f\|. \end{aligned}$$

Taking also $\|H_j f\| \leq K \|f\|$ (see (2.15)) into account and using the inductive hypothesis (2.19) for $j - 1$, we then easily deduce that

$$|\pi_j^m f - \pi_j f| \leq \frac{K^{N-1}}{\varepsilon} 2\beta_m \|f\| (1 + K A_{j-1}).$$

Therefore (2.19) holds, provided we have

$$A_j = 2 \frac{K^{N-1}}{\varepsilon} (1 + K A_{j-1}) = \rho \left(\frac{1}{K} + A_{j-1} \right).$$

Since $A_0 = 0$, this gives (2.20) for A_j . \square

Finally we have another version of this result which proves useful for case B. First, we introduce the following norm on Lipschitz functions on \mathbb{R}^d :

$$\|f\|' = \|f\| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} \quad (2.21)$$

(with $\|f\|' = \infty$ by convention if f is not Lipschitz). Next we assume that we have for some constant K' :

$$\|H_j f\|' \leq K' \|f\|' \quad \forall j = 1, \dots, N. \quad (2.22)$$

Observe that this implies (2.15) with $K \leq K'$. Finally, suppose that for each n we have other kernels H_j^m satisfying

$$\|H_j^m f - H_j f\| \leq \beta_m' \|f\|' \quad (2.23)$$

for some $\beta_m' > 0$. Define again the probability measures π_n^m by induction on n by (2.18).

Proposition 2.2 *Assume (2.22) and (2.23). For all bounded Lipschitz functions f on \mathbb{R}^d and all $j = 1, \dots, N$, we have*

$$|\pi_j^m f - \pi_j f| \leq \beta_m' \|f\|' A_j', \quad (2.24)$$

where

$$A_j' = \frac{\rho'^{j+1} - \rho'}{K'(\rho' - 1)}, \quad \rho' = 2 \frac{K'^N}{\varepsilon}. \quad (2.25)$$

Proof. The proof is exactly the same as for Proposition 2.1: we use the properties

$$\begin{aligned} |\pi_{j-1}^{m_n} H_j^{m_n} f - \pi_{j-1}^m H_j f| &\leq \beta'_m \|f\|', \\ |\pi_j^m f| &\leq \|f\| \leq \|f\|', \end{aligned}$$

and $\|H_j f\|' \leq K' \|f\|'$ by (2.22) and the fact that under (2.22) we have (2.16) with K' instead of K . \square

3 Case A

In this section we study our first case A, which is the easiest one. The assumptions are (E1) and (A1); recall that g is defined in (A1).

A possible procedure goes as follows. We fix an integer n , and we take for m_n the smallest integer $\geq \sqrt{n}$. Then we perform the following steps:

- **Step 1:** We simulate n i.i.d. variables $(X_0^j)_{1 \leq j \leq n}$ according to the law μ , and for each j a variable X_1^j according to the law $Q^{(m_n)}(X_0^j; \cdot)$.

- **Notation:** At the end of Step $k \geq 1$ we will know the variables $(X_k^j)_{1 \leq j \leq n}$, so we can set for every function f on \mathbb{R}^d :

$$W_k^n f = \frac{1}{n} \sum_{i=1}^n f(X_k^i) g(X_k^i, y_k). \quad (3.1)$$

Then we introduce the following random probability measure on \mathbb{R}^d :

$$\Psi_k^n = \begin{cases} \frac{1}{n W_k^n} \sum_{i=1}^n g(X_k^i, y_k) \varepsilon_{X_k^i} & \text{if } W_k^n > 0 \\ \varepsilon_0 & \text{otherwise} \end{cases} \quad (3.2)$$

(the measure ε_0 above could indeed be replaced by any probability measure on \mathbb{R}^d).

- **Step $k \geq 2$:** We simulate n i.i.d. variables $(X_{k-1}^j)_{1 \leq j \leq n}$ according to the law Ψ_{k-1}^n . Then for each $j = 1, \dots, n$ we simulate the random variable X_k^j according to the law $Q^{(m_n)}(X_{k-1}^j; \cdot)$.

We stop at the end of Step N , and our approximation of $\pi_N f$ will be $\Psi_N^n f$: as always in this paper, π_N is defined by (2.13) with a given observation (y_1, \dots, y_N) , while the expectations below refer to the probability underlying our Monte-Carlo simulations.

Theorem 3.1 *Assume (E1) and (A1). For all bounded Borel functions f on \mathbb{R}^d , all $k = 1, \dots, N$ and all n , we have*

$$\left. \begin{aligned} E(|\Psi_k^n f - \pi_k f|) &\leq \frac{\bar{F}_k}{\sqrt{n}} \|f\|, \\ P(|\Psi_k^n f - \pi_k f| \geq \delta) &\leq \bar{F}_k \exp\left(-n \left(\frac{\delta}{\bar{F}_k \|f\|}\right)^2\right), \end{aligned} \right\} \quad (3.3)$$

where (with $\rho = 2K^N/\varepsilon$: recall that ε is given in (2.14):

$$F_k = C \frac{(4\rho)^{k+1} - 4\rho}{4\rho - 1}, \quad \bar{F}_k = 3^{k+1}, \quad \hat{F}_k = C(8\rho)^k \quad (3.4)$$

and C depends only on a , b and g .

Remark 1: Since $E(U) = \int_0^\infty P(U \geq u)du$ holds for any nonnegative variable U , the first property in (3.3) readily follows from the second one and from the fact that $\int_0^\infty e^{-y^2/2}dy = \sqrt{\pi/2}$, except that the constant F_k has to be substituted with $\bar{F}_k \hat{F}_k \sqrt{\pi}$, which is much bigger: so if one is interested only in the rates of convergence (i.e. the behavior as $n \rightarrow \infty$), then there is no need to prove the first half of (3.3).

Remark 2: More generally, $E(U^p) = \int_0^\infty P(U \geq u^{1/p})du$. So the second property in (3.3) yields

$$E(|\Psi_k^n f - \pi_k f|^p)^{1/p} \leq \frac{F_k(p)}{\sqrt{n}} \|f\| \quad (3.5)$$

with $F_k(p) = C_p \hat{F}_k \bar{F}_k^{1/p}$ for some constant C_p depending on p only; exactly as in Remark 1, the constants $F_k(p)$ thus obtained are far too big, and a direct proof would yield smaller constants.

Proof. 1) For further reference, let us recall some well known estimates for i.i.d. random variables. Suppose that ξ_1, \dots, ξ_n are random variables which conditionally on a σ -field \mathcal{G} are independent and centered. Then we have the following two Kolmogorov exponential inequalities (see e.g. Lemma 1.5 and 1.6 in [4]), where δ is arbitrary positive:

$$|\xi_i| \leq a, \quad E(|\xi_i|^2|\mathcal{G}) \leq b \quad \Rightarrow \quad P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| \geq \delta\right) \leq \begin{cases} 2 \exp(-n \frac{\delta^2}{2a^2}), \\ 2 \exp\left(-n \frac{\delta^2}{2b} \left(2 - e^{a\delta/b}\right)\right). \end{cases}$$

In the case where $\xi_i = \eta_i - E(\eta_i|\mathcal{G})$, we deduce that

$$|\eta_i| \leq a, \quad E(|\eta_i|^2|\mathcal{G}) \leq b \quad \Rightarrow \quad P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right| \geq \delta\right) \leq \begin{cases} 2 \exp(-n \frac{\delta^2}{8a^2}), \\ 2 \exp\left(-n \frac{\delta^2}{2b} \left(2 - e^{2a\delta/b}\right)\right) \end{cases} \quad (3.6)$$

because with the above notation we have $|\xi_i| \leq 2a$ and $E(|\xi_i|^2|\mathcal{G}) \leq b$. Taking the latter into account, we also have

$$E(|\eta_i|^2|\mathcal{G}) \leq b \quad \Rightarrow \quad E\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i\right|^2\right) \leq \frac{b}{n}. \quad (3.7)$$

2) We can introduce another random measure on \mathbb{R}^d by $\Psi_k^m = \frac{1}{n} \sum_{i=1}^n \varepsilon_{X_k^{i,i}}$ (setting $X_0^i = X_0$). We denote by H_j^n the kernel defined by (2.12) with Q substituted by Q^{m_n} ,

and π_j^n is defined by the recursive formula (2.18). In virtue of (2.4) the estimate (2.17) holds with $\beta_n = \alpha \|g\|/m_n \leq \alpha K/\sqrt{n}$, where α is some constant.

In fact we will prove together the following estimates:

$$E(|\Psi_k^n f - \pi_k^n f|) \leq \frac{G_k}{\sqrt{n}} \|f\|, \quad k = 1, \dots, N, \quad (3.8)$$

$$P(|\Psi_k^n f - \pi_k^n f| \geq \delta) \leq 2I_k \exp\left(-n \frac{\delta^2}{8J_k \|f\|^2}\right), \quad k = 1, \dots, N, \quad (3.9)$$

$$E(|\Psi_k^m f - \pi_k^m f|) \leq \frac{G'_k}{\sqrt{n}} \|f\|, \quad k = 0, \dots, N-1. \quad (3.10)$$

$$P(|\Psi_k^m f - \pi_k^m f| \geq \delta) \leq 2I'_k \exp\left(-n \frac{\delta^2}{8J'_k \|f\|^2}\right), \quad k = 0, \dots, N-1. \quad (3.11)$$

for some $G_k, G'_k, I_k, I'_k, J_k$ and J'_k to be computed later.

3) Let us begin with some preliminary estimates. Below f is a bounded Borel function on \mathbb{R}^d . If $k \geq 1$ we have $\Psi_k^m f - \Psi_k^n f = \frac{1}{n} \sum_{i=1}^n \xi_i$, where $\xi_i = f(X_k^i) - \Psi_k^n f$. These variables satisfy the assumptions for (3.6) and (3.7), with $\eta_i = f(X_k^i)$ and w.r.t. the σ -field \mathcal{G} generated by the variables $X_k^i : i = 1, \dots, n$ and with $a = \|f\|$ and $b = \|f\|^2$. Hence

$$\left. \begin{aligned} E(|\Psi_k^m f - \Psi_k^n f|^2) &\leq \frac{\|f\|^2}{n}, \\ P(|\Psi_k^m f - \Psi_k^n f| \geq \delta) &\leq 2 \exp\left(-n \frac{\delta^2}{8\|f\|^2}\right). \end{aligned} \right\} \quad (3.12)$$

Next, the kernel H_k^n takes the form $H_k^n f(x) = Q^{(m_n)}(f g_k)(x)$, where $g_k(x') = g(x', y_k)$. We have $W_k^n f - \Psi_{k-1}^m H_k^n f = \frac{1}{n} \sum_{i=1}^n \xi_i$, where

$$\xi_i = f(X_k^i) g_k(X_k^i) - Q^{(m_n)}(f g_k)(X_{k-1}^i).$$

These variables satisfy the assumptions for (3.6) and (3.7), with $\eta_i = f(X_k^i) g_k(X_k^i)$ and w.r.t. the σ -field \mathcal{G} generated by the variables $X_{k-1}^i : i = 1, \dots, n$ and with $a = K \|f\|$ and $b = K^2 \|f\|^2$ (recall that here $K = \|g\|$). Hence

$$\left. \begin{aligned} E(|W_k^n f - \Psi_{k-1}^m H_k^n f|^2) &\leq \frac{\|f\|^2 K^2}{n}, \\ P(|W_k^n f - \Psi_{k-1}^m H_k^n f| \geq \delta) &\leq 2 \exp\left(-n \frac{\delta^2}{8\|f\|^2 K^2}\right). \end{aligned} \right\} \quad (3.13)$$

4) Now we proceed to prove (3.8), (3.9), (3.10) and (3.11) by induction on k .

Since $\pi_0^n = \mu$ and the variables X_0^i are i.i.d. with law μ , we can again apply (3.7) and (3.6) to the variables $\xi_i = f(X_0^i) - \mu(f)$, thus obtaining (3.10) and (3.11) for $k = 0$ with $G'_0 = I'_0 = J'_0 = 1$.

Now let $k \geq 1$ and assume that (3.10) and (3.11) hold for $k-1$. Recalling (2.19) and (2.17) and $H_k^n 1 \leq K$ and $\beta_n \leq \alpha K/\sqrt{n}$, we get

$$\begin{aligned} |\pi_{k-1}^n H_k^n 1 - \pi_{k-1} H_k 1| &\leq \pi_{k-1}^n |H_k^n 1 - H_k 1| + |\pi_{k-1}^n H_k 1 - \pi_{k-1} H_k 1| \\ &\leq \beta_n (1 + K A_{k-1}) \leq \frac{\alpha K A_k}{\rho \sqrt{n}}, \end{aligned}$$

where we used the property $1 + KA_{k-1} = A_k/\rho$ coming from (2.20). Since (2.16) gives $\pi_{k-1}H_k1 \geq 2K/\rho$, it follows that

$$\sqrt{n} \geq \alpha A_k \quad \Rightarrow \quad \pi_{k-1}^n H_k^n 1 \geq \frac{K}{\rho}. \quad (3.14)$$

In view of (2.13) and (3.2), we have

$$\Psi_k^n f - \pi_k^n f = \frac{1}{\pi_{k-1}^n H_k^n 1} \left(\frac{W_k^n f}{W_k^n 1} (\pi_{k-1}^n H_k^n 1 - W_k^n 1) + (W_k^n f - \pi_{k-1}^n H_k^n f) \right)$$

if $W_k^n 1 > 0$ and $\Psi_k^n f - \pi_k^n f = f(0) - \pi_k^n f$ otherwise. Therefore if $Lf = W_k^n f - \pi_{k-1}^n H_k^n f$, we obtain as soon as the condition in (3.14) is met and since obviously $|W_k^n f| \leq \|f\| W_k^n 1$:

$$\left. \begin{aligned} |\Psi_k^n f - \pi_k^n f| &\leq \begin{cases} \frac{\rho}{K} (|Lf| + \|f\| |L1|) & \text{if } W_k^n 1 > 0 \\ 2\|f\| & \text{otherwise} \end{cases} \\ |Lf| &\leq |W_k^n f - \Psi_{k-1}^n H_k^n f| + |\Psi_{k-1}^n H_k^n f - \pi_{k-1}^n H_k^n f|. \end{aligned} \right\} \quad (3.15)$$

Now, (3.13) and (3.10) for $k-1$ and $\|H_k^n f\| \leq K\|f\|$ yield

$$\left. \begin{aligned} E(|Lf|) &\leq \frac{\|f\|K}{\sqrt{n}} (1 + G'_{k-1}), \\ P(|Lf| \geq \delta) &\leq 2e^{-n\delta^2/32\|f\|^2 K^2} + 2I'_{k-1} e^{-n\delta^2/32J'_{k-1}\|f\|^2 K^2}. \end{aligned} \right\} \quad (3.16)$$

Now we will prove (3.8) and (3.10) for k . Assume first that $\sqrt{n} \geq \alpha K$. In this case, (3.14) yields that $W_k^n 1 = 0 \Rightarrow |L1| \geq K/\rho$: thus (3.16) yields $P(W_k^n 1 = 0) \geq \frac{\rho}{\sqrt{n}}(1 + G'_{k-1})$. Then putting together (3.15) and the first line of (3.16), we get (3.8) for k , provided $G_k \geq 4\rho(1 + G'_{k-1})$. On the other hand the left side of (3.8) is smaller than $2\|f\|$, so (3.8) will hold for all n having $\sqrt{n} < \alpha A_k$ as soon as $G_k \geq 2\alpha A_k$. Further, under (3.8) we readily deduce (3.10) from (3.12) for k , with $G'_k = 1 + G_k$. Therefore (3.8) and (3.10) will be true for all k, n , as soon as

$$G_k \geq (4\rho(1 + G'_{k-1})) \bigvee (\alpha A_k), \quad G'_k = 1 + G_k, \quad G'_0 = 1.$$

At this point, one easily checks that the above properties are satisfied if

$$G_k = (2 \bigvee \alpha) \frac{(4\rho)^{k+1} - 4\rho}{4\rho - 1}. \quad (3.17)$$

Next we prove (3.9) and (3.11) for k . Assume first that $\sqrt{n} \geq \alpha A_k$. Then (3.14) and (3.15) yield

$$P(|\Psi_k^n f - \pi_k^n f| \geq \delta) \leq P(|Lf| \geq \frac{\delta K}{2\rho}) + P(|L1| \geq \frac{\delta K}{2\rho\|f\|}) + P(W_k^n 1 = 0) 1_{\{2\|f\| \geq \delta\}}.$$

Recall also that $P(W_k^n 1 = 0) \leq P(|L1| \geq K/\rho) \leq P(|L1| \geq \delta K/2\rho\|f\|)$ if $2\|f\| \geq \delta$. Thus (3.16) yields that (3.9) holds with $I_k \geq 3(1 + I'_{k-1})$ and $J_k \geq (4\rho)^2 J'_{k-1}$. Since the left side

of (3.9) is 0 as soon as $\delta > 2\|f\|$ and is always smaller than 1, we see that (3.9) holds for all n with $\sqrt{n} < \alpha A_k$ as soon as $I_k \geq 2$ and $J_k \geq (\alpha A_k)^2/2$. Further, taking into account (3.12) we deduce from (3.9) that (3.11) also holds for k , as soon as $I'_k = 1 + I_k$ and $J'_k = 4J_k$. Therefore (3.9) and (3.11) will be true for all k, n , as soon as

$$\begin{aligned} I_k &\geq (3(1 + I'_{k-1}) \vee 2), & J_k &\geq ((4\rho)^2 J'_{k-1}) \vee \frac{(\alpha A_k)^2}{2}, \\ I'_k &= 1 + I_k, & J'_k &= 4J_k, & I'_0 &= J'_0 = 1. \end{aligned}$$

At this point, one easily checks that the above property are satisfied if

$$J_k = \left(1 \vee \frac{\alpha^2}{K^2}\right) (2^6 \rho^2)^k, \quad I_k = 3^{k+1} - 3. \quad (3.18)$$

5) So far we have (3.8) and (3.9) for all k , with constants given by (3.17) and (3.18). It remains to apply (2.19). Since $\beta_n \leq \alpha K/\sqrt{n}$, the first part of (3.3) is obvious with F_k given in (3.4), because the function $x \mapsto \frac{x^{k+1}-x}{x-1}$ is increasing over $(0, \infty)$, and with $C = 2 + 2\alpha$.

In order to obtain the second part of (3.3) we observe that when $\beta_n \|f\| A_k \leq \delta/2$, i.e. when $n\delta^2/\|f\|^2 \geq (2\alpha K A_k)^2$, then the left side of (3.3) is smaller than the left side of (3.9) written with $\delta/2$; otherwise it is certainly smaller than 1: so the result will hold with \hat{F}_k and \bar{F}_k given by (3.4), in view of (3.18). \square

Now, if we are allowed to simulate at most r “single” random variables, we give our final result under the assumptions of this section. For a given n , the previous procedure necessitates $n(1 + m_n)N$ single simulations (counting for 1 each variable X_0^i and X_k^i and for m_n each variable X_k^i for $k = 1, \dots, N$ because of the Euler scheme which we use). So we choose $n = n(r)$ to be the biggest integer $n(1 + m_n)N \leq r$, and the simulated filter is

$$\hat{\Pi}_N^r = \Psi_N^{n(r)}. \quad (3.19)$$

Then (3.3) gives, as soon as $r \geq 6N$ (hence $n(r) \geq (r/2N)^{2/3}$):

$$\left. \begin{aligned} E(|\hat{\Pi}_N^r f - \pi_N f|) &\leq \frac{(2N)^{1/3} F_N}{r^{1/3}} \|f\| \\ P(|\hat{\Pi}_N^r f - \pi_N f| \geq \delta) &\leq \hat{F}_N \exp\left(-r^{2/3} \left(\frac{\delta}{F_N (2N)^{1/3} \|f\|}\right)^2\right). \end{aligned} \right\} \quad (3.20)$$

We also observe that we have strong consistency for our estimates $\hat{\Pi}_N^r$: indeed, taking $\delta = r^{-1/6}$ in the second estimate (3.20) yields that

$$\sum_{r \geq 1} P(|\hat{\Pi}_N^r f - \pi_N f| \geq \frac{1}{r^{1/6}}) < \infty,$$

so that $\hat{\Pi}_N^r f$ converges almost surely to $\pi_N f$ as $r \rightarrow \infty$ by Borel-Cantelli lemma.

Finally, let us emphasize that the constants obtained above are perhaps (very) big, but they also are “uniform” in the observations (y_1, \dots, y_N) as soon as $\mu H_1 \dots H_N 1 \geq \varepsilon$ (see (2.14)) because they are obviously decreasing as ε increases.

4 Case B

In this section we study our second case B. The assumptions are (E2) and (B1). The function G is not explicitly known in this case, but the kernel H_j is given by the following formula, where as above φ is the density of the normal law $\mathcal{N}(0, \sigma\sigma^*)$ in \mathbb{R}^q :

$$H_j f(x) = E \left(f(X(x)_1) \varphi \left(y_j - y_{j-1} - \int_0^1 h(X(x)_s) ds \right) \right). \quad (4.1)$$

Thus in view of (B1) and (2.6) and since the function g is also bounded and Lipschitz, we have (2.22) with some constant K' depending on a , b and h .

A possible procedure goes as follows. We fix an integer n , and we perform the following steps:

- **Step 1:** We simulate n i.i.d. variables $(X_0^i)_{1 \leq i \leq n}$ according to the law μ , and we set $X_0^n = X_0^i$. Then for each i we simulate the variables $(X_{1,j}^i)_{1 \leq j \leq n}$ along the Euler scheme (2.1) with stepsize $1/n$ and starting point X_0^i , i.e. with the same law as the sequence $(X(X_0^i, n)_j)_{1 \leq j \leq n}$ if we use the notation (2.1).

- **Notation:** At the end of Step $k \geq 1$ we will know the variables $(X_{k,j}^i)_{1 \leq i \leq n, 1 \leq j \leq n}$, so we can set for every function f on \mathbb{R}^d :

$$W_k^n f = \frac{1}{n} \sum_{i=1}^n f(X_{k,n}^i) \varphi \left(y_k - y_{k-1} - \frac{1}{n} \sum_{j=1}^n h(X_{k,j}^i) \right). \quad (4.2)$$

Then we introduce the following random probability measure on \mathbb{R}^d :

$$\Psi_k^n = \frac{1}{n W_k^n} \sum_{i=1}^n \varphi \left(y_k - y_{k-1} - \frac{1}{n} \sum_{j=1}^n h(X_{k,j}^i) \right) \varepsilon_{X_{k,n}^i} \quad (4.3)$$

(observe that here $\varphi > 0$, hence $W_k^n > 0$ identically).

- **Step $k \geq 2$:** We simulate n i.i.d. variables $(X_{k-1}^j)_{1 \leq j \leq n}$ according to the law Ψ_{k-1}^n . Then for each $i = 1, \dots, n$ we simulate the random variables $(X_{k,j}^i)_{1 \leq j \leq n}$ according to the law of the sequence $(X(X_{k-1}^i, n)_j)_{1 \leq j \leq n}$.

We stop at the end of Step N . Our approximation of $\pi_N f$ will be $\Psi_N^n f$. Below, recall that $\|f\|'$ is the Lipschitz norm given by (2.21).

Theorem 4.1 *Assume (E2) and (B1). For all bounded Lipschitz functions f on \mathbb{R}^d , all $k = 1, \dots, N$ and all n , we have*

$$\left. \begin{aligned} E(|\Psi_k^n f - \pi_k f|) &\leq \frac{F'_k}{\sqrt{n}} \|f\|', \\ P(|\Psi_k^n f - \pi_k f| \geq \delta) &\leq \bar{F}'_k \exp \left(-n \left(\frac{\delta}{\bar{F}'_k \|f\|'} \right)^2 \right), \end{aligned} \right\} \quad (4.4)$$

where (with $\rho' = 2K'^N/\varepsilon$):

$$F'_k = C \frac{(4\rho')^{k+1} - 4\rho'}{4\rho' - 1}, \quad \bar{F}'_k = 3^{k+1}, \quad \hat{F}'_k = C(8\rho)^k, \quad (4.5)$$

and C depends only on a, b, h and σ .

Proof. 1) We prove first that (2.23) holds with $\beta'_m = \alpha K'/\sqrt{m}$ for some constant α , provided we choose for H_k^m the kernel defined by

$$H_k^m f(x) = E \left(f(X(x, m)_m) \varphi \left(y_k - y_{k-1} - \frac{1}{m} \sum_{j=1}^m h(X(x, m)_j) \right) \right), \quad (4.6)$$

and π_k^m is defined by the recursive formula (2.18). We have

$$|H_j^m f(x) - H_j f(x)| \leq E(|U_m(x) - V_m(x)|) + E(|V_m(x) - V(x)|),$$

where

$$U_m(x) = f(X(x, m)_m) \varphi \left(y_n - y_{n-1} - \frac{1}{m} \sum_{j=1}^m h(X(x, m)_j) \right),$$

$$V_m(x) = f(X(x)_1) \varphi \left(y_n - y_{n-1} - \frac{1}{m} \sum_{j=1}^m h(X(x)_{j/m}) \right),$$

$$V(x) = f(X(x)_1) \varphi \left(y_n - y_{n-1} - \int_0^1 h(X(x)_s) ds \right).$$

Since φ is bounded and f is bounded and Lipschitz, we have

$$|U_m(x) - V_m(x)| \leq C \|f\|' \sup_{1 \leq j \leq m} |X(x, m)_j - X(x)_{j/m}|,$$

and thus (2.5) yields

$$E(|U_m(x) - V_m(x)|) \leq C \|f\|' \frac{1}{\sqrt{m}}.$$

We also have

$$E(|V_m(x) - V(x)|) \leq C \|f\| \sum_{j=1}^m \int_{\frac{j-1}{m}}^{\frac{j}{m}} E(|X(x)_s - X(x)_{\frac{j}{m}}|) ds,$$

and by well known results about Euler schemes (see [3]) (E2) implies that this expression is smaller than C/\sqrt{m} . Putting these facts together shows that (2.23) holds with $\beta'_m = \alpha K'/\sqrt{m}$.

2) Now the result is proved exactly as in Theorem 3.1: we introduce another random measure on \mathbb{R}^d by $\Psi_k^m = \frac{1}{n} \sum_{i=1}^n \varepsilon_{X_k^i}$ (setting $X_0^i = X_0^i$), and the proof goes exactly the

same way: we have to replace β_n by β'_n , and K and ρ by K' and ρ' , and use Proposition 2.2 instead of Proposition 2.1. We have to take

$$\xi_i = f(X_{k,m_n}^i) \varphi \left(y_k - y_{k-1} - \frac{1}{n} \sum_{j=1}^{m_n} h(X_{k,j}^i) \right) - H_k^n(X_{k-1}^i)$$

for the proof of (3.13), and the fact that $\|f\|'$ comes in instead of the smaller number $\|f\|$ is due to the use of Proposition 2.2 (since (2.17) does not hold here). \square

Now, if we are allowed to simulate at most r “single” random variables, we give our final result under the assumptions of this section. For a given n , the previous procedure necessitates $n(1+n)N$ single simulations. So we choose $n = n(r)$ to be the biggest integer $n(1+n)N \leq r$, and the simulated filter is again given by (3.19). Then (4.4) gives, as soon as $r \geq 8N$:

$$\left. \begin{aligned} E(|\hat{\Pi}_N^r f - \pi_N f|) &\leq \frac{(2N)^{1/4} F'_N}{r^{1/4}} \|f\|' \\ P(|\hat{\Pi}_N^r f - \pi_N f| \geq \delta) &\leq \hat{F}'_N \exp \left(-r^{1/2} \left(\frac{\delta}{F'_N (2N)^{1/4} \|f\|'} \right)^2 \right). \end{aligned} \right\} \quad (4.7)$$

Here also we have strong consistency for our estimates $\hat{\Pi}_N^r$.

5 Case C

Now we study our last case C. The assumptions are (E1) and (E3). We have seen that in this case the function G of subsection 2-3 is $G(x, y; x', y') = p(x, y; x', y')/q(x, x')$, so that

$$H_k f(x) = \int p(x, y_{k-1}, x', y_k) f(x') dx'. \quad (5.1)$$

We choose a Borel bounded function φ from \mathbb{R}^q into $[0, \infty)$ such that

$$\int \varphi(y) dy = 1, \quad \gamma := \int |y| \varphi(y) dy < \infty. \quad (5.2)$$

We can then propose the following procedure: we fix an integer $n \geq 1$. Then m_n denotes the smallest integer $\geq n^{1/(2+q)}$, and we set

$$\varphi_n(y) = n^{q/(2+q)} \varphi(y n^{1/(2+q)}). \quad (5.3)$$

Then we perform the following steps:

- **Step 1:** we simulate n i.i.d. variables $(X_0^i)_{1 \leq i \leq n}$ according to the law μ , and for each i a variable (X_1^i, Y_1^i) according to the law $P^{(m_n)}(X_0^i, y_0; \cdot)$.

• **Notation:** at the end of Step k we will know the variables $(X_k^i, Y_k^i)_{1 \leq i \leq n}$, so we can set for every function f on \mathbb{R}^d :

$$W_k^n f = \frac{1}{n} \sum_{i=1}^n f(X_k^i) \varphi_n(Y_k^i - y_k). \quad (5.4)$$

Then we introduce the following random probability measure on \mathbb{R}^d :

$$\Psi_k^n = \begin{cases} \frac{1}{nW_k^n} \sum_{i=1}^n \varphi_n(Y_k^i - y_k) \varepsilon_{X_k^i} & \text{if } W_k^n > 0 \\ \varepsilon_0 & \text{otherwise} \end{cases} \quad (5.5)$$

• **Step $k \geq 2$:** We simulate n i.i.d. variables $(X_{k-1}^i)_{1 \leq i \leq n}$ according to the law Ψ_{k-1}^n . Then for each $j = 1, \dots, n$ we simulate the random variable (X_k^j, Y_k^j) according to the law $P^{(m_n)}(X_{k-1}^j, y_{k-1}; \cdot)$.

We stop at the end of Step N . Our approximation of $\pi_N f$ will be $\Psi_N^n f$.

Theorem 5.1 *Assume (E1) and (E3). For all bounded Borel functions f on \mathbb{R}^d , all $k = 1, \dots, N$ and all n , we have*

$$\left. \begin{aligned} E(|\Psi_k^n f - \pi_k f|) &\leq \frac{F_k}{n^{1/(2+q)}} \|f\|, \\ P(|\Psi_k^n f - \pi_k f| \geq \delta) &\leq \bar{F}_k \exp\left(-n^{2/(2+q)} \left(\frac{\delta}{\bar{F}_k \|f\|}\right)^2\right), \end{aligned} \right\} \quad (5.6)$$

where F_k , \bar{F}_k and \hat{F}_k are given by (3.4) with the same ρ and another constant C depending on a, b, a', b' and φ .

Proof. Here again the proof will be similar to the proof of Theorem 3.1.

1) We denote by H_k^n the kernel defined by

$$H_k^n f(x) = \int p^{(m_n)}(x, y_{k-1}, x', y') f(x') \varphi_n(y' - y_k) dx' dy', \quad (5.7)$$

and π_k^n is defined by the recursive formula (2.18).

In view of (5.1), (5.2) and (5.3) (which yields $\int \varphi_n(y) dy = 1$) we get

$$\begin{aligned} H_k^n f(x) - H_k f(x) &= \int (p^{(m_n)}(x, y_{k-1}; x', y') - p(x, y_{k-1}; x', y')) f(x') \varphi_n(y' - y_k) dx' dy' \\ &\quad + \int (p(x, y_{k-1}; x', y') - p(x, y_{k-1}; x', y_k)) f(x') \varphi_n(y' - y_k) dx' dy'. \end{aligned} \quad (5.8)$$

The first integral above can be divided into two parts: first, the integral over $A_n = \{(x', y') : |x - x'| + |y - y_{k-1}| > 2/m_n\}$, and this bit is smaller than $C\|f\|/m_n$ by (2.9) and $\int \varphi_n(y) dy = 1$; second, the integral over the complement A_n^c which is smaller than

$$C\|f\| \int_{\{x': |x' - x| \leq 2/m_n\}} dx' \int \varphi_n(y' - y_k) dy' \leq C\|f\|/m_n^q$$

by (2.8). Recalling that the derivative $\partial p(x, y; x', y')/\partial y'$ also satisfies (2.8) under (E1) and (E3), and since $\int |y|\varphi_n(y)dy = \gamma n^{-1/(2+q)}$, the second integral in (5.8) is smaller than $C\|f\|n^{-1/(2+q)}$. So in view of the definition of m_n we finally get

$$|H_k^n f(x) - H_k f(x)| \leq C\|f\| \frac{1}{n^{1/(2+q)}}.$$

In other words, we have

$$\|H_k^n f - H_k f\| \leq \beta_n \|f\| \quad \text{with} \quad \beta_n = \frac{C}{n^{1/(2+q)}}. \quad (5.9)$$

2) We introduce another random measure on \mathbb{R}^d by $\Psi_k^m = \frac{1}{n} \sum_{i=1}^n \varepsilon_{X_k^i}$ (setting $X_0^i = X_0^i$). We will prove the following estimates

$$E(|\Psi_k^n f - \pi_k^n f|) \leq \frac{G_k}{n^{1/(2+q)}} \|f\|, \quad k = 1, \dots, N, \quad (5.10)$$

$$P(|\Psi_k^n f - \pi_k^n f| \geq \delta) \leq 2I_k \exp\left(-n^{\frac{2}{2+q}} \frac{\delta^2}{J_k \|f\|^2}\right), \quad k = 1, \dots, N, \quad (5.11)$$

$$E(|\Psi_k^m f - \pi_k^n f|) \leq \frac{G'_k}{n^{1/(2+q)}} \|f\|, \quad k = 0, \dots, N-1. \quad (5.12)$$

$$P(|\Psi_k^m f - \pi_k^n f| \geq \delta) \leq 2I'_k \exp\left(-n^{\frac{2}{2+q}} \frac{\delta^2}{J'_k \|f\|^2}\right), \quad k = 0, \dots, N-1. \quad (5.13)$$

for some $G_k, G'_k, I_k, I'_k, J_k$ and J'_k to be computed later.

3) Here again we give preliminary estimates. First (3.12) is proved exactly as in Theorem 3.1. Next, $W_k^n f - \Psi_{k-1}^m H_k^n f = \frac{1}{n} \sum_{i=1}^n \xi_i$, where

$$\xi_i = f(X_k^i) \varphi_n(X_k^i - y_k) - H_k^n f(X_{k-1}^i).$$

These variables satisfy the assumptions for (3.6), with $\eta_i = f(X_k^i) \varphi_n(X_k^i - y_k)$ and w.r.t. the σ -field \mathcal{G} generated by the variables $X_{k-1}^i : i = 1, \dots, n$ and with $a = a_f := C_1 \|f\| n^{q/(2+q)}$ and $b = b_f := C_2 \|f\|^2 n^{q/(2+q)}$ for some constants C_1 and C_2 : indeed, the first of these assumptions is obvious by (5.3), and for the second one we may write

$$E(|\eta_i|^2 | \mathcal{F}'_{k-1}) \leq C \|f\|^2 n^{q/(2+q)} E(\varphi_n(Y_k^i - y_k) | \mathcal{G}) = C \|f\|^2 n^{q/(2+q)} H_k^n 1(X_{k-1}^i)$$

and $H_k^n 1 \leq K + \beta_n \leq C$ (see (2.15) and (5.9)). Then instead of (3.13) we deduce from (3.7) and from the second inequality in (3.7) that for yet another constant C :

$$\left. \begin{aligned} E(|W_k^n f - \Psi_{k-1}^m H_k^n f|^2) &\leq C \frac{\|f\|^2 K^2}{n^{2/(2+q)}} \\ P(|W_k^n f - \Psi_{k-1}^m H_k^n f| \geq \delta) &\leq 2 \exp\left(-n^{\frac{2}{2+q}} \frac{\delta^2}{C \|f\|^2}\right) \end{aligned} \right\} \quad (5.14)$$

(up to replacing the constant C_2 in the above definition of b_f , we can always assume that $2a_f \delta / b_f \leq 1/2$, hence $2 - e^{2a\delta/b} \geq 2 - \sqrt{e} > 0$, for all functions f and all $\delta > 0$ such that

$\delta < 2\|f\|$: in this case the second inequality comes from (3.6); since on the other hand its left side equals 0 when $\delta > \|f\|$, it always holds).

4) Now we proceed to prove (5.10), (5.11), (5.12) and (5.13) by induction on k . The last two estimates hold for $k = 0$ with $G'_0 = I'_0 = J'_0 = 1$, as in Theorem 3.1.

Now let $k \geq 1$ and assume that (5.12) and (5.13) hold for $k - 1$. Using (5.9) instead of $\beta_n \leq \alpha K/\sqrt{n}$, (3.14) gets substituted with (α being still another positive constant)

$$n^{\frac{1}{2+q}} \geq \alpha A_k \quad \Rightarrow \quad \pi_{k-1}^n H_k^n 1 \geq \frac{K}{\rho}. \quad (5.15)$$

Then under (5.15) we still have (3.15), hence (3.12) and (5.14) allow to replace (3.16) by

$$\left. \begin{aligned} E(|Lf|) &\leq \frac{\|f\|K}{\sqrt{n}}(C + G'_{k-1}), \\ P(|Lf| \geq \delta) &\leq 2e^{-n^{2/(2+q)}\delta^2/4C\|f\|^2K^2} + 2I'_{k-1}e^{-n^{2/(2+q)}\delta^2/4J'_{k-1}\|f\|^2K^2}. \end{aligned} \right\} \quad (5.16)$$

Then as in Theorem 3.1 we see that (5.10) and (5.12) for k , provided

$$G_k \geq (4\rho(C + G'_{k-1})) \vee (\alpha A_k), \quad G'_k = 1 + G_k, \quad G'_0 = 1,$$

which are satisfied by

$$G_k = \left((1 + C) \vee \alpha \right) \frac{(4\rho)^{k+1} - 4\rho}{4\rho - 1}. \quad (5.17)$$

Similarly, we obtain that (3.9) and (3.11) hold for k , provided

$$\begin{aligned} I_k &\geq (3(1 + I'_{k-1}) \vee 2), & J_k &\geq \left((4\rho)^2 (J'_{k-1} \vee C) \right) \vee \frac{(\alpha A_k)^2}{2}, \\ I'_k &= 1 + I_k, & J'_k &\geq 4(J_k \vee C), & I'_0 &= J'_0 = 1. \end{aligned}$$

The above properties are satisfied if

$$J_k = \left(C + 1 \vee \frac{\alpha^2}{K^2} \right) (2^6 \rho^2)^k, \quad I_k = 3^{k+1} - 3. \quad (5.18)$$

5) So far we have (5.10) and (3.9') for all k , with constants given by (5.17) and (5.18). It remains to apply (2.19) and the fact that $\beta_n = C/n^{1/(2+q)}$ by (5.9): Again exactly as in Theorem 3.1 we conclude that (5.6) holds with the prescribed constants. \square

Now, if we are allowed to simulate at most r ‘‘single’’ random variables, we give our final result under the assumptions of this section. For a given n , the previous procedure necessitates $n(1 + m_n)N$ single simulations. So we choose $n = n(r)$ to be the biggest integer $n(1 + m_n)N \leq r$, and the simulated filter is again given by (3.19). Then (5.6) gives, as soon as $r \geq 8N$:

$$\left. \begin{aligned} E(|\hat{\Pi}_N^r f - \pi_N f|) &\leq \frac{(2N)^{1/(3+q)} F_N}{r^{1/(3+q)}} \|f\| \\ P(|\hat{\Pi}_N^r f - \pi_N f| \geq \delta) &\leq \hat{F}_N \exp \left(-r^{2/(3+q)} \left(\frac{\delta}{\hat{F}_N (2N)^{1/(3+q)} \|f\|} \right)^2 \right), \end{aligned} \right\} \quad (5.19)$$

and here again we have strong consistency for our estimates $\hat{\Pi}_N^r$.

6 Comments and numerical results

6-1) The regularity assumptions in (E1), (E2) and (E3) are reasonably weak, but the non-degeneracy ones could obviously be weakened: taking advantage of the result of Bally and Talay we could replace the uniform ellipticity by weak Hörmander's conditions, at the expense of much stronger regularity on the coefficients.

Another interesting point is when σ degenerates in (1.3), for example when $\sigma = 0$. Even then our filtering problem is not trivial, since the knowledge of $\int_0^1 h(X_t) dt$ does not entail full knowledge of X_1 in general. In this situation we can treat case B as a particular case of C with degenerate coefficient $b'b'^*$.

Still another interesting situation is when $b = 0$, i.e. X is a deterministic (not observable) function. The above method does not work since in the “simulation” of X we get a single deterministic path. However one could add a “small noise”, i.e. consider (1.1) with $b(x) = \varepsilon$ and run the algorithm: this is a standard regularization technique used in practice.

In a different direction, it is interesting to see what happens when f is not bounded. Because of the bounds (2.2) the same results will hold when f has $|f(x)| \leq \gamma(1+|x|^p)$, with $\|f\|$ replaced by a number depending on γ and p , and except of course for the exponential bounds (the second halves of (3.20), (4.7) and (5.19)).

6-2) Our second comment is that it makes no practical sense here to achieve an “infinite” precision on the approximate filter. In practice one usually wants to deduce the value $f(X_N)$ for some given function f (possibly a family of such functions) from the observations (y_1, \dots, y_N) . Then even under the true filter π_N the order of magnitude of the error made is the standard deviation $\Sigma_N = \Sigma_n(y_1, \dots, y_N)$ of f under π_N . Thus in this situation it is somewhat meaningless to take the number r of simulated variables such that $|\hat{\Pi}_N^r f - \pi_N f|$ is much smaller than Σ_N .

6-3) An obvious drawback of our method is the form of the constants F_N, \dots . For example F_N is roughly of the form $F_N = CC'^N$ where $C' > 1$. If the “true” value of F_N is really exponential in N , then clearly the method is not feasible.

In order to get an idea of the true value of F_N we have conducted a series of simulations in the following very special case, where $d = q = 1$ and

$$\left. \begin{aligned} dX_t &= \alpha X_t dt + dW_t, & X_0 &= 1, \\ dY_t &= X_t dt + \sigma dW'_t, & Y_0 &= 0. \end{aligned} \right\} \quad (6.1)$$

The true filter π_N is known (and given by the Kalman-Bucy discrete-time filter). We have run the algorithm given for case C, for the function $f(x) = x$ (this is not bounded, but see the comments in 6-2 above), and for an observation set (y_1, \dots, y_N) obtained by a first preliminary simulation of a particular path of the pair (X, Y) .

We have considered three cases, and in each case the algorithm is run for all N between 1 and 90:

Case 1: $\alpha = -.5$, $\sigma = 1$, so X is a recurrent Ornstein-Uhlenbeck process. The standard deviation of f under π_N is $\Sigma_1 = .74$, $\Sigma_2 = .79$ and $\Sigma_N = .8$ for all $N \geq 3$ (the two first values are smaller than the others because in our model we exactly know X_0).

Case 2: $\alpha = -.5$, $\sigma = .1$: this is like the above, except that the smaller value of σ implies that we get “more information” on X from Y , and this is apparent from the standard deviation for the filter which is now $\Sigma_1 = .49$ and $\Sigma_N = .51$ for all $N \geq 2$.

Case 3: $\alpha = .1$, $\sigma = 1$, so X is a transient Ornstein-Uhlenbeck process. The standard deviation of f under π_N is $\Sigma_1 = .94$, $\Sigma_2 = 1.06$ and $\Sigma_N = 1.07$ for all $N \geq 2$.

The function φ of (5.2) is the normalized indicator function of an interval $[-u, u]$ where u has been empirically adjusted so that in (5.5) one does not encounter the second stated possibility. This led us to take $u = 20$ in cases 1 and 2, and $u = 40$ in case 3.

The algorithm has been run for the following values of n in Theorem 5.1: $n = 100, 500, 1000, 2000$, using of course the Euler approximation for X and Y . For each of these values it has been run 20 times, and the following tables give for some values of N the average absolute value of the normalized error $n^{1/3}|\Psi_N^n f - \pi_N f|$ over these 20 runs (recall that here $\pi_N f$ is known).

Case 1

N	n=100	n=500	n=1000	n=2000
5	1.3	2.1	3.3	4.4
10	2.7	4.9	5.0	7.0
20	1.1	4.6	6.0	5.4
30	1.1	2.1	3.1	4.1
40	1.1	1.9	3.0	3.1
50	1.9	4.0	5.2	7.6
60	1.4	1.5	3.4	2.5
70	5.9	9.8	10.6	13.9
80	3.0	4.9	5.7	8.5
90	3.0	6.1	9.3	10.4

Case 2

N	n=100	n=500	n=1000	n=2000
5	9.6	18.6	24.0	27.3
10	0.9	2.6	3.5	4.5
20	6.7	10.9	12.5	16.8
30	4.7	7.4	9.1	10.8
40	7.2	14.2	20.2	21.7
50	2.1	3.3	4.9	7.2
60	2.4	3.3	4.3	5.6
70	4.6	8.7	9.7	10.9
80	2.7	4.9	4.9	7.8
90	7.0	11.8	17.5	21.3

Case 3				
N	n=100	n=500	n=1000	n=2000
5	4	3	12	20
10	14	21	41	74
20	40	145	214	238
30	56	444	359	164
40	146	984	893	369
50	375	1970	1581	907
60	1052	2952	4457	2760
70	2910	4973	6284	7883
80	7902	13541	17033	21516
90	21525	36873	46500	58475

These numerical results are all surprising: in the recurrent cases, the constant F_N seems indeed fairly independent from N , presumably because the process X forgets rather quickly its past values. In the transient case, the constant F_N seems to behave somewhat exponentially in N , as expected. There is however something unexpected, namely that the results are worse in case 2 than in case 1, while the observation noise is smaller.

Note also that since the standard deviation of f under π_N is of the order of magnitude equal to 1, there is no reason to increase the precision of our estimate of $\pi_N f$ much further than 1. Recalling that the above tables give the error multiplied by $n^{1/3}$, and looking at the data for $N = 90$ for example, we see that we could stop at $n = 100$ already for case 1, while for the other two cases $n = 2000$ is not enough, and in case 3 the results are clearly totally inefficient: n should be much larger (and hopefully when n gets large the normalized error stabilizes...).

References

- [1] V. Bally, D. Talay: The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. *Probab. Th. Related Fields*, **104**, 43-60, (1996).
- [2] V. Bally, D. Talay: The law of the Euler scheme for stochastic differential equations: II. Approximation of the density, *Monte Carlo Methods and Applications*, **2**, 93-128, (1996).
- [3] T.G. Kurtz, P. Protter: Wong Zakai corrections, random evolutions, and simulation schemes for SDE's, *Stochastic Analysis*, ed. Mayer-Wolf, Merzbach and Schwarz, Academic Press: New York, 1991, 331-146.
- [4] M. Ledoux, M. Talagrand: *Probability in Banach Spaces, Isoperimetry and Processes*, Springer-Verlag, 23, (1991)