

# USING SYSTEMATIC SAMPLING SELECTION FOR MONTE CARLO SOLUTIONS OF FEYNMAN–KAC EQUATIONS

IVAN GENTIL,\* *Université Paris-Dauphine*

BRUNO RÉMILLARD,\*\* *HEC Montréal*

## Abstract

While the convergence properties of many sampling selection methods can be proven, there is one particular sampling selection method introduced in Baker (1987), closely related to ‘systematic sampling’ in statistics, that has been exclusively treated on an empirical basis. The main motivation of the paper is to start to study formally its convergence properties, since in practice it is by far the fastest selection method available. We will show that convergence results for the systematic sampling selection method are related to properties of peculiar Markov chains.

*Keywords:* Feynman–Kac formulae; sequential Monte Carlo; genetic algorithm; systematic sampling; Markov chain

2000 Mathematics Subject Classification: Primary 65C35; 60G35  
Secondary 60J10

## 1. Introduction

Let  $(X_k)_{k \geq 0}$  be a nonhomogeneous Markov chain on a locally compact metric space  $E$  with transition kernels  $(K_n)_{n \geq 1}$  and initial law  $\eta_0$  defined on the Borel  $\sigma$ -field  $\mathcal{B}(E)$ . Furthermore, let  $\mathcal{B}_b(E)$  be the set of bounded  $\mathcal{B}(E)$ -measurable functions.

Given a sequence  $(g_n)_{n \geq 1}$  of positive functions in  $\mathcal{B}_b(E)$ , suppose that we want to calculate recursively the following Feynman–Kac formulae,  $(\eta_n)_{n \geq 1}$ :

$$\eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)}, \quad f \in \mathcal{B}_b(E),$$

where

$$\gamma_n(f) = \mathbb{E} \left( f(X_n) \prod_{k=1}^n g_k(X_{k-1}) \right).$$

Note that most nonlinear filtering problems are particular cases of Feynman–Kac formulae.

Following [9] and [12], let  $M_1(E)$  denote the set of probability measures on  $(E, \mathcal{B}(E))$ . If  $\mu \in M_1(E)$  and  $n \geq 0$ , let  $\mu K_n$  be the probability measure defined on  $\mathcal{B}_b(E)$  by

$$\mu K_n(f) = \mu(K_n f) = \int_E \int_E f(z) K_n(x, dz) \mu(dx).$$

Received 12 October 2005; revision received 17 February 2008.

\* Postal address: CEREMADE, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, F-75775 Paris cedex 16, France. Email address: gentil@ceremade.dauphine.fr

\*\* Postal address: Service de l’enseignement des méthodes quantitatives de gestion, HEC Montréal, 3000 chemin de la côte-Sainte-Catherine, Montréal, Canada H3T 2A7. Email address: bruno.remillard@hec.ca

In order to understand the relation between the  $\eta_n$ s for any  $n \geq 1$ , let  $\psi_n : M_1(E) \mapsto M_1(E)$  be defined by

$$\psi_n(\eta)f = \frac{\eta(g_n f)}{\eta(g_n)}, \quad \eta \in M(E), f \in \mathcal{B}_b(E),$$

and let  $\Phi_n$  denote the mapping from  $M_1(E)$  to  $M_1(E)$  defined by

$$\Phi_n(\eta) = \psi_n(\eta)K_n.$$

Then it is easy to check that, for any  $n \geq 1$ ,

$$\eta_n = \Phi_n(\eta_{n-1}). \tag{1.1}$$

Note also that, for any  $n \geq 1$ , the mapping  $\Phi_n$  can be decomposed into

$$\hat{\eta}_n = \psi_{n+1}(\eta_n), \quad \eta_{n+1} = \hat{\eta}_n K_{n+1}, \quad n \geq 0, \eta_0 \in M_1(E). \tag{1.2}$$

Furthermore, note that the first transformation,  $\eta_n \mapsto \hat{\eta}_n$ , is nonlinear, while the second transformation,  $\hat{\eta}_n \mapsto \eta_{n+1}$ , is linear.

Even though the forward system of equations (1.1) looks simple, it can rarely be solved analytically, and if an analytic solution did exist, it would require extensive calculations. This is why algorithms for approximating  $(\eta_n)_{n \in \mathbb{N}}$ , starting from  $\eta_0$ , are so important. One such method, presented in [7], [11], and [12], is to build approximations of the measures  $(\eta_n)_{n \in \mathbb{N}}$  using interacting particle systems. The algorithm uses decomposition (1.2) and, by analogy with genetics, the first step, which is related to a sampling selection method, is often referred to as the *selection* step and the second step is termed the *mutation* step, while in reality it is a Markovian evolution of the particles. The speed of any algorithm depends on the two steps. In [19] the authors focused on the mutation step; in this paper we focus on the sampling selection process.

In this paper we discuss the properties of a particular algorithm, called the ‘systematic sampling’ selection in this paper and the ‘stochastic universal sampling’ selection in the genetic algorithms literature. It seems to have first appeared in [2]. It has been reintroduced in the filtering literature in [5]; see also [17].

In what follows, a description of the general algorithm is given in Section 2, with a few examples of sampling selection methods, together with some tools for studying its convergence. In Section 3 we focus on the systematic sampling selection method, giving some properties, and stating some convergence results and a conjecture, based on results for Markov chains which we prove in Appendix A. Finally, in Section 4 numerical comparisons between the sampling selection methods are made through a simple model of nonlinear filtering for noisy black-and-white images. It serves two purposes. Firstly, to illustrate that our conjecture holds. Secondly, to show that, in terms of precision, the systematic sampling selection method performs as well as other methods.

## 2. Algorithm and sampling selection methods

The general algorithm for approximating the solution of (1.1) is given first, following the exposition in [9] and [12], and particular sampling selection methods are presented second. Throughout the rest of the paper, it is assumed that, for any  $n \geq 1$ ,

$$0 < \inf_{x \in E} g_n(x) \leq \sup_{x \in E} g_n(x) < +\infty.$$

**2.1. General algorithm**

Let  $N$  be an integer representing the number of particles and, for any  $n \geq 0$ , let  $\xi_n = \{\xi_n^1, \dots, \xi_n^N\}$  denote the particles at time  $n$  and set

$$\eta_n^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_n^i}.$$

- At time  $n = 0$  the initial particle system  $\xi_0 = \{\xi_0^1, \dots, \xi_0^N\}$  consists of  $N$  independent and identically distributed particles with common law  $\eta_0$ .
- For each  $n \geq 1$ , the particle system  $\xi_n = \{\xi_n^1, \dots, \xi_n^N\}$  consists of  $N$  particles and is obtained in the following way.

*Sampling/selection.* First calculate the weights vector  $W_n \in (0, 1)^N$ , where

$$W_n^i = \frac{g_n(\xi_{n-1}^i)}{\sum_{i=1}^N g_n(\xi_{n-1}^i)}, \quad i = 1, \dots, N. \tag{2.1}$$

Then select, according to a given sampling selection method, a sample  $\hat{\xi}_{n-1} = \{\hat{\xi}_{n-1}^1, \dots, \hat{\xi}_{n-1}^N\}$  of size  $N$  from  $\xi_{n-1}$  with weights  $(W_n^i)$ .

*Evolution/mutation.* Given  $\hat{\xi}_{n-1}$ , the new particle system  $\xi_n$  consists of particles  $\xi_n^i$  chosen independently from the law  $K_n(\hat{\xi}_{n-1}^i, dx)$ ,  $1 \leq i \leq N$ . In other words, for any  $z = (z^1, \dots, z^N) \in E^N$ ,

$$P(\xi_n \in dx \mid \hat{\xi}_{n-1} = z) = \bigotimes_{i=1}^N K_n(z^i, dx^i).$$

Note that in order to describe a sampling selection method, it suffices to define how the numbers  $M_n^1, \dots, M_n^N \in \{0, 1, \dots, N\}$  are randomly selected, with  $M_n^i$  representing the number of times particle  $\hat{\xi}_{n-1}^i$  appears in the new sample. Therefore, we can write

$$\hat{\eta}_{n-1}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\hat{\xi}_{n-1}^i} = \frac{1}{N} \sum_{i=1}^N M_n^i \delta_{\hat{\xi}_{n-1}^i}.$$

A sampling selection method will be said to be conditionally unbiased if, for any  $i \in \{1, \dots, N\}$  and any  $k \geq 1$ ,  $E(M_k^i \mid \xi_{k-1}) = N W_k^i$ .

**Remark 2.1.** Conditional unbiasedness yields the property that

$$E(\eta_k^N f \mid \xi_{k-1}) = \Phi_k(\eta_{k-1}^N)(f), \quad f \in \mathcal{B}_b(E), \tag{2.2}$$

since

$$\begin{aligned} E(\eta_n^N f \mid \xi_{n-1}) &= E(E(\eta_n^N f \mid \xi_{n-1}, \hat{\xi}_{n-1}) \mid \xi_{n-1}) \\ &= E\left(\frac{1}{N} \sum_{i=1}^N M_n^i K_n f(\hat{\xi}_{n-1}^i) \mid \xi_{n-1}\right) \\ &= \sum_{i=1}^N W_n^i K_n f(\hat{\xi}_{n-1}^i) \\ &= \Phi_n(\eta_{n-1}^N)(f). \end{aligned}$$

The mean square error of a particular sampling selection method can be obtained using the following useful result from [12]. Before stating the result, define, for any measurable  $\eta$  with values in  $M(E)$ ,

$$\|\eta\|_2^2 = \sup_{f \in \mathcal{B}_b(E), \|f\|_\infty \leq 1} E((\eta f)^2).$$

**Theorem 2.1.** ([12, Theorem 2.36].) *Assume that the sampling selection method is conditionally unbiased and that the following condition is verified for all  $1 \leq k \leq n$ : there exists a constant  $C_k$  such that, for all  $N$ -dimensional vectors  $\{q^1, \dots, q^N\} \in \mathbb{R}^N$ ,*

$$E\left(\left(\frac{1}{N} \sum_{i=1}^N (M_k^i - NW_k^i) q^i\right)^2 \middle| \xi_{k-1}\right) \leq \frac{1}{N} C_k \max_{1 \leq i \leq N} |q^i|^2. \tag{2.3}$$

Then, for all  $1 \leq k \leq n$ , there exists a constant  $C'_k$  such that

$$\|\eta_k^N - \eta_k\|_2^2 \leq \frac{C'_k}{N}.$$

In what follows, only conditionally unbiased sampling selection methods are considered. As shown in Remark 3.2, we can see that, in general, the systematic sampling selection method defined below does not satisfy condition (2.3), while classical sampling selection methods, like the ones listed in Section 2.3, do satisfy condition (2.3). Therefore, weaker conditions must be imposed in order to obtain mean square convergence. In fact, we have the following result.

**Theorem 2.2.** *Let  $(a_N)$  be a sequence such that  $a_N/N \rightarrow 0$  as  $N \rightarrow \infty$ . Assume that the sampling selection method is conditionally unbiased. Then*

$$\lim_{N \rightarrow \infty} a_N \max_{1 \leq k \leq n} \|\eta_k^N - \eta_k\|_2^2 = 0$$

if and only if, for any  $f \in \mathcal{B}_b(E)$ ,

$$\lim_{N \rightarrow \infty} a_N \max_{1 \leq k \leq n} E\left(\left(\frac{1}{N} \sum_{i=1}^N (M_k^i - NW_k^i) f(\xi_{k-1}^i)\right)^2\right) = 0. \tag{2.4}$$

Moreover,  $\sup_{N \geq 1} a_N \max_{1 \leq k \leq n} \|\eta_k^N - \eta_k\|_2^2$  is finite if and only if

$$\sup_{N \geq 1} a_N \max_{1 \leq k \leq n} E\left(\left(\frac{1}{N} \sum_{i=1}^N (M_k^i - NW_k^i) f(\xi_{k-1}^i)\right)^2\right) < \infty.$$

*Proof.* Suppose that  $a_N/N \rightarrow 0$ , and let  $f \in \mathcal{B}_b(E)$  be given. First note that, using the unbiasedness condition together with (2.2), we have, for any  $k \in \{1, \dots, n\}$ ,

$$E((\eta_k^N f - \eta_k f)^2) = E((\eta_k^N f - \Phi_k(\eta_{k-1}^N) f)^2) + E((\Phi_k(\eta_{k-1}^N) f - \eta_k f)^2).$$

Since  $g_k \geq c_k > 0$  for some positive constant  $c_k$ ,  $k \geq 1$ , by hypothesis, it follows that

$$\lim_{N \rightarrow \infty} a_N \max_{1 \leq k \leq n} \|\eta_k^N - \eta_k\|_2^2 = 0$$

if and only if, for any  $k = 1, \dots, n$ ,  $\lim_{N \rightarrow \infty} a_N E((\eta_k^N f - \Phi_k(\eta_{k-1}^N) f)^2) = 0$ . Next, it can be easily shown that, for any  $k = 1, \dots, n$ ,  $E((\eta_k^N f - \Phi_k(\eta_{k-1}^N) f)^2 \mid \xi_{k-1})$  can be written as

$$E\left(\left(\frac{1}{N} \sum_{i=1}^N (M_k^i - NW_k^i) K_k f(\xi_{k-1}^i)\right)^2 \mid \xi_{k-1}\right) + \frac{1}{N} \Psi_k(\eta_{k-1}^N)(K_k f^2 - (K_k f)^2).$$

Since  $K_k f \in \mathcal{B}_b(E)$ ,

$$0 \leq \frac{1}{N} \Phi_k(\eta_{k-1}^N)(K_k f^2 - (K_k f)^2) \leq \frac{1}{N} \|f\|_\infty^2,$$

and  $a_N/N \rightarrow 0$ , it follows from the calculations above that

$$\lim_{N \rightarrow \infty} a_N \max_{1 \leq k \leq n} E((\eta_k^N f - \Phi_k(\eta_{k-1}^N) f)^2) = 0$$

if and only if (2.4) holds. The rest of the proof is similar, so it is omitted.

**2.2. Systematic sampling**

By obvious analogy with systematic sampling in statistics, the first sampling selection method that is described is simply called ‘systematic sampling’. It appears that this method was first proposed in [2] under the strange name ‘stochastic universal sampling’, in the context of unbiased sampling selection for genetic algorithms. However, nobody has formally studied its convergence properties.

As opposed to the definition given in [2], the sampling selection method can simply be defined in the following way. For  $n \geq 1$ , let  $U_n$  be a uniform random variable on  $[0, 1)$  and note that, for  $w \in [0, 1]$ ,  $M(w, U_n) := \lfloor Nw + U_n \rfloor$ , where  $\lfloor x \rfloor$  denotes the integer part of  $x$ . Then

$$M_n^1 := M(W_n^1, U_n),$$

$$M_n^k := M(W_n^1 + \dots + W_n^k, U_n) - M(W_n^1 + \dots + W_n^{k-1}, U_n), \quad k = 2, \dots, N.$$

Since  $M(1, U_n) = N$ , we find that  $\sum_{i=1}^N M_n^i = N$ . Therefore, the number of particles is always equal to  $N$ . Properties of the systematic sampling selection method are examined in Section 3.

**2.3. Other sampling methods**

We can classify the various sampling selection methods into two categories, according to whether the number of particles is constant or random. The following list is by no means exhaustive. We concentrate on ‘classical’ selection methods. For the first two methods,  $N$  is constant, while  $N_n$  fluctuates in the last two methods. For other sampling selection methods, we refer the reader to [9], [11], [12], and the references therein. See also [4] for a recent review of the comparison of various selections. Note that the last two methods are particular cases of what is known as ‘branching selection methods’ in the filtering literature. However, there are many other sampling methods, for example, the tree sampling method proposed in [9]. The latter is more computationally demanding, and since it is not as simple to implement as the other methods, it is not considered here.

2.3.1. *Simple random sampling.* This selection method is based on simple random sampling without rejection. It follows that

$$(M_n^1, \dots, M_n^N) \sim \text{Multinomial}(N, W_n^1, \dots, W_n^N),$$

where  $(W_n^i)_{1 \leq i \leq N}$  is given in (2.1). This sampling selection method is the classical selection method and has many interesting properties that have been studied mainly by Del Moral and coauthors; see, e.g. [11]. In particular, the conditions of Theorem 2.1 are satisfied; we can also prove a central limit theorem and large deviations properties.

2.3.2. *Remainder stochastic sampling.* This algorithm was first introduced in [3] in the context of unbiased sampling selection for genetic algorithms; see also [1] and [2] for comparisons between sampling selection methods in the latter context. It is also defined as ‘residual sampling’ in [18]. It is a much faster algorithm to implement than the simple random sampling selection method, it satisfies the conditions of Theorem 2.1, and Douc and Moulines [13] recently investigated some of its convergence properties. See also [12] and the references therein. To describe the selection method, first define  $\tilde{N} = N - \sum_{i=1}^N \lfloor NW_n^i \rfloor = \sum_{i=1}^N \{NW_n^i\}$ , where  $\{x\}$  stands for the fractional part of  $x$ , i.e.  $\{x\} = x - \lfloor x \rfloor$ . Next, allocate the (possibly) remaining  $\tilde{N}$  particles via simple random sampling, i.e.

$$(M_n^1 - \lfloor W_n^1 \rfloor, \dots, M_n^N - \lfloor W_n^N \rfloor) \sim \text{Multinomial}(\tilde{N}, \tilde{W}_n^1, \dots, \tilde{W}_n^N),$$

where  $\tilde{W}_n^i = \{NW_n^i\} / \sum_{j=1}^N \{NW_n^j\}$ ,  $1 \leq i \leq N$ . In fact,  $\tilde{N}$  and  $N$  are of the same order; thus, the speed of this selection method is almost the same as for the simple random sampling selection method.

2.3.3. *Binomial sampling.* As stated before, for this sampling selection method and the next one, the number of particles at time  $n$  is random and denoted by  $N_n$ ,  $n \geq 0$ . Of course,  $N_0$  is fixed. For  $n \geq 1$ , and given that  $\xi_{n-1}$  and  $N_{n-1}$ ,  $M_n^1, \dots, M_n^{N_{n-1}}$  are independent and  $M_n^i \sim \text{Bin}(N_{n-1}, W_n^i)$  for  $i = 1, \dots, N_{n-1}$ . It follows that

$$N_n = \sum_{i=1}^{N_{n-1}} M_n^i.$$

This sampling selection method is a little faster to implement than the simple random sampling selection method, but a major drawback is that there is no control on the number of particles. Moreover,  $P(N_n = 0) > 0$ .

2.3.4. *Bernoulli sampling.* The Bernoulli sampling selection method was introduced in [10]. See also [6] and [8] for additional properties of this sampling selection method. It is worth noting that  $M_n^i$  takes the same values as in the systematic sampling selection method, provided that  $N_{n-1} = N$ . For  $n \geq 1$ , and given  $\xi_{n-1}$  and  $N_{n-1}$ ,  $M_n^1, \dots, M_n^{N_{n-1}}$  are independent, where  $M_n^i$  is defined by

$$M_n^i = \lfloor N_{n-1} W_n^i \rfloor + \varepsilon_n^i, \quad \varepsilon_n^i \sim \text{Ber}(\{N_n W_n^i\}), \quad 1 \leq i \leq N_{n-1}.$$

Note that  $N_n \geq 1$  and that the following alternative representation also holds:

$$M_n^i = \lfloor N(W_n^1 + \dots + W_n^i) + U_n^i \rfloor - \lfloor N(W_n^1 + \dots + W_n^{i-1}) + U_n^i \rfloor,$$

where  $U_n^1, \dots, U_n^{N_{n-1}}$  are independent and  $U_n^i \sim \text{Unif}([0, 1])$ , given  $\xi_{n-1}$  and  $N_{n-1}$ .

### 3. Some properties and results for systematic sampling selection

Throughout the rest of the paper, the selection method used is the one defined in Section 2.2. Let us begin with some elementary properties of the systematic sampling selection method.

**Lemma 3.1.** *Suppose that  $U_n$  is uniformly distributed over  $[0, 1)$ . Then, conditionally on  $\xi_{n-1}$ , we have, for any  $i \in \{1, \dots, N\}$ ,*

$$M_n^i - \lfloor NW_n^i \rfloor \sim \text{Ber}(\{NW_n^i\}).$$

*In particular, for any  $i \in \{1, \dots, N\}$ ,  $E(M_n^i \mid \xi_{n-1}) = NW_n^i$ .*

*Proof.* It suffices to show that whenever  $U \sim \text{Unif}([0, 1))$  and  $x, y \geq 0$ ,  $\lfloor U + x + y \rfloor - \lfloor U + x \rfloor - \lfloor y \rfloor$  is a Bernoulli random variable with parameter  $p = \{y\}$ . To this end, first note that  $V = \{U + x\}$  is also uniformly distributed on  $[0, 1)$ . Furthermore,

$$\lfloor U + x + y \rfloor - \lfloor U + x \rfloor - \lfloor y \rfloor = \lfloor \{U + x\} + \{y\} \rfloor = \lfloor V + \{y\} \rfloor \sim \text{Ber}(\{y\});$$

hence the result holds.

**Remark 3.1.** Using the same proof as in Lemma 3.1, we obtain, conditionally on  $\xi_{n-1}$ ,  $M_n^i + \dots + M_n^j - \lfloor N(W_n^i + \dots + W_n^j) \rfloor \sim \text{Ber}(\{N(W_n^i + \dots + W_n^j)\})$  for any  $i \leq j \in \{1, \dots, N\}$ . Also note that, since the sampling selection method is unbiased, for any  $n \geq 1$ , we have (2.2).

To obtain  $L^2$ -convergence of the algorithm based on the systematic sampling selection method, we apply Theorem 2.1 of [12]. All the sampling selections presented in Section 2.3 satisfy property (2.4). If  $N_n$  is random, there is a similar condition to (2.4). But, as we show next, systematic sampling behaves differently.

**Remark 3.2.** Inequality (2.3) is not verified in general for the systematic sampling selection method. Here is an illustration. Suppose that  $N = 2m$  and let, for any  $i \in \{1, \dots, N/2\}$ ,  $W_n^{2i} = 1/2N$  and  $W_n^{2i-1} = 3/2N$ . Then we can check that, for any  $1 \leq i \leq N/2$ ,

$$\begin{aligned} M_n^{2i-1} &= 1 & \text{if } U_n \in [0, \frac{1}{2}), & & M_n^{2i-1} &= 2 & \text{if } U_n \in [\frac{1}{2}, 1), \\ M_n^{2i} &= 1 & \text{if } U_n \in [0, \frac{1}{2}), & & M_n^{2i} &= 0 & \text{if } U_n \in [\frac{1}{2}, 1). \end{aligned}$$

Next, if  $1 \leq i \leq N/2$ , set  $q^{2i} = 1$  and  $q^{2i-1} = -1$ . It follows that

$$E\left(\left(\frac{1}{N} \sum_{i=1}^N (M_k^i - NW_k^i) q^i\right)^2 \mid \xi_{k-1}\right) = \frac{1}{4},$$

which shows that inequality (2.3) is false.

The description of all the sampling selection methods in Sections 2.2 and 2.3 prove that the systematic sampling selection method is the least computationally demanding. In light of Remark 3.2, we can see that the  $L^2$ -convergence of the particle system with the systematic sampling selection is not easy. However, we believe that the following conjecture holds.

**Conjecture 3.1.** *Suppose that  $\eta_0$  and  $(K_n)_{n \geq 1}$  are absolutely continuous laws, and consider that  $M_n^1, \dots, M_n^N$  are obtained using the systematic sampling selection method. Then, for all  $f \in \mathcal{B}_b(E)$  and  $n \geq 1$ , (2.4) holds with  $a_N \equiv 1$ , i.e.*

$$\lim_{N \rightarrow \infty} E\left(\left(\frac{1}{N} \sum_{i=1}^N (M_n^i - NW_n^i) f(\xi_{n-1}^i)\right)^2\right) = 0. \tag{3.1}$$

Note that it follows from Theorem 2.2 that the above conjecture is equivalent to  $\|\eta_n^N - \eta_n\|_2 \rightarrow 0$  as  $N \rightarrow \infty$  for any  $n \geq 0$ . In what follows, we try to determine why Conjecture 3.1 might be true. To this end, first note that, for any  $1 \leq k \leq N$ ,

$$M_n^k - NW_n^k = \{N(W_n^1 + \dots + W_n^{k-1}) + U_n\} - \{N(W_n^1 + \dots + W_n^k) + U_n\}.$$

Now, set  $F_n^0 = 0$  and  $F_n^k = \sum_{j=1}^k g_n(\xi_{n-1}^j)$ ,  $1 \leq k \leq N$ . For any  $\alpha > 0$  and any  $f \in \mathcal{B}_b(E)$ , further define

$$\begin{aligned} Z_n^N(f, \alpha) &= \frac{1}{\sqrt{N}} \sum_{k=1}^N f(\xi_{n-1}^k) \left( \left\{ \frac{F_n^{k-1}}{\alpha} + U_n \right\} - \left\{ \frac{F_n^k}{\alpha} + U_n \right\} \right) \\ &= \frac{1}{\sqrt{N}} \sum_{k=1}^N f(\xi_{n-1}^k) (\{S_n^{k-1}\} - \{S_n^k\}), \end{aligned}$$

where  $S_n^k = F_n^k/\alpha + U_n$  and  $S_n^0 = U_n$ . Then, setting

$$\bar{g}_n = \frac{1}{N} \sum_{k=1}^N g_n(\xi_n^k),$$

and defining  $Y_n^N(f) = Z_n^N(f, \bar{g}_n)$ , we have

$$Y_n^N(f) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (M_n^i - NW_n^i) f(\xi_{n-1}^i).$$

So we can rewrite (3.1) in the form

$$\lim_{N \rightarrow \infty} \frac{1}{N} E((Y_n^N(f))^2) = 0.$$

Unfortunately, working with  $Y_n^N$  appears to be impossibly difficult, so instead we could work with a more tractable quantity, namely  $Z_n^N$ . In the case in which  $n = 1$  we have the following result, which is a first step in proving Conjecture 3.1. Before stating it, recall that  $D([0, 1])$  is the space of càdlàg functions (i.e. those that are continuous from the right with left limits) with Skorokhod’s topology.

**Theorem 3.1.** *Assume that the law of  $\{g_1(\xi_0^1)\}$  is absolutely continuous. Then, for any  $\alpha > 0$  and any  $f \in \mathcal{B}_b(E)$ , the sequence of processes  $B_N \in D([0, 1])$  defined by*

$$B_{f,\alpha}^N(t) = Z_1^{\lfloor Nt \rfloor}(f, \alpha), \quad t \in [0, 1],$$

*converges in  $D([0, 1])$  to  $\sigma B_{f,\alpha}$ , where  $B_{f,\alpha}$  is a Brownian motion and*

$$\lim_{N \rightarrow \infty} E((Z_1^N(f, \alpha))^2) = \sigma^2.$$

The proof of Theorem 3.1 is an easy consequence of Theorem A.2 in Appendix A with  $X_k = f(\xi_0^k)$ ,  $Y_k = \{g_1(\xi_0^k)\}$ , and  $f(x, y, s) = x(y - s)$ . In addition, there is an ‘explicit’ expression for  $\sigma^2$ . More details can be found in Appendix A.



**Remark 3.3.** Theorem 3.1 does not prove Conjecture 3.1 in the case in which  $n = 1$ . However, if we are willing to deal with a random number of particles at step  $n = 1$  then the following interesting result is obtained. Set  $N_1 = \lfloor U_1 + \sum_{j=1}^N g_1(\xi_0^k)/\alpha \rfloor$ , and define

$$\hat{\eta}_0^{N_1} = \frac{1}{N_1} \sum_{k=1}^{N_1} M_1^k \delta_{\xi_0^k}.$$

Then, as  $N \rightarrow \infty$ ,  $N_1/N \rightarrow \eta_0(g_1)/\alpha$  and  $\limsup_{N \rightarrow \infty} N \|\eta_1^{N_1} - \eta_1\|_2^2 < \infty$ .

To prove convergence for higher orders, i.e.  $n > 1$ , we would need results from nonhomogeneous Markov chains. This approach will be examined in the near future using, for example, the results of [20].

**Remark 3.4.** In order to keep  $N_1$  fixed, we could try to control the  $Z_1^N(\bar{g}_1) - Z_1^N(\alpha)$  term. Since

$$\sqrt{N}(\bar{g}_1 - \alpha) \rightsquigarrow \sqrt{\eta_0(g_1^2)} \mathcal{Z},$$

where ‘ $\rightsquigarrow$ ’ represents an intuition for the computation and  $\mathcal{Z} \sim \mathcal{N}(0, 1)$ , it follows that

$$\sqrt{N} \left( \frac{1}{\bar{g}_1} - \frac{1}{\alpha} \right) = -\frac{\sqrt{\eta_0(g_1^2)}}{\eta_0^2(g_1)} \mathcal{Z} + o_P(1).$$

If we could differentiate term by term, we would obtain

$$(Z_1^N(\bar{g}_1) - Z_1^N(\alpha)) \rightsquigarrow \eta_0(K_1 f g_1) \frac{\sqrt{\eta_0(g_1^2)}}{\eta_0(g_1)^2} \mathcal{Z} = \eta_1(f) \frac{\sqrt{\eta_0(g_1^2)}}{\eta_0(g_1)} \mathcal{Z}.$$

So we could guess that

$$Y_N \rightsquigarrow \eta_1(f) \frac{\sqrt{\eta_0(g_1^2)}}{\eta_0(g_1)} \mathcal{Z} + B_{f,\alpha}(1).$$

On the other hand, if the sequence  $Z_1^N(\alpha)$  was tight for  $\alpha$  in a closed interval not containing 0 then we would obtain  $Z_1^N(\bar{g}_1) - Z_1^N(\alpha) \rightarrow 0$  in probability. There is no indication so far in favor of one of these two approaches.

### 4. Numerical comparisons

The numerical comparisons will be made through a simple model of filtering for tracking a moving target using noisy black-and-white images. It serves two purposes. Firstly, it illustrates that our conjecture holds. Secondly, it shows that, in terms of precision, the systematic sampling selection method performs as well as the other selection methods.

The main interest of this model is that we can explicitly compute the exact filter, that is,  $\eta_n$  is known for any  $n \geq 1$  (see, e.g. [15]), enabling us to make a comparison between the approximation  $(\eta_n^N)_{N>0}$  and the optimal measure  $\eta_n$ .

#### 4.1. Description of the model

We will assume that the target moves on  $\mathbb{Z}^2$  according to a Markov chain. Observations consist of black-and-white noisy images in a finite fixed region  $R \subset \mathbb{Z}^2$ . More precisely, let  $(X_n)_{n \geq 0}$

be a homogeneous Markovian chain with values in  $\mathfrak{X} = \{\omega \in \{0, 1\}^{\mathbb{Z}} : \sum_{x \in \mathbb{Z}} \omega(x) = 1\}$ . Of course, the position of the target at step  $n$  is  $x_0$  if and only if  $X_n(x_0) = 1$ . Set

$$M(a, b) = P(X_{n+1}(a) = 1 \mid X_n(b) = 1), \quad a, b \in \mathbb{Z}^2. \tag{4.1}$$

Note that  $M$  describes the movement of the target exactly.

The model for observations  $Y_k \in \{0, 1\}^R, k = 1, \dots, n$ , is the following. Given  $X_0, \dots, X_n$ , assume that the  $\{Y_n(x)\}_{x \in R}$  are independent and, for any  $x \in R$ ,

$$P(Y_n(x) = 0 \mid X_n(x) = 0) = p_0, \quad P(Y_n(x) = 1 \mid X_n(x) = 1) = p_1,$$

where  $0 < p_0, p_1 < 1$ . For a more realistic result, we can assume that  $\frac{1}{2} < p_0, p_1 < 1$ . We want to compute the distribution of  $X_k$  conditionally to  $\mathcal{Y}_n$ , where  $\mathcal{Y}_n$  is the  $\sigma$ -algebra generated by the observations  $Y_1, \dots, Y_n$  and  $\mathcal{Y}_0$  is the trivial  $\sigma$ -algebra. As in Section 2 of [15], note that, for any  $(\omega, \omega') \in \{0, 1\}^R \otimes \mathfrak{X}$ , the conditional probability  $P(Y_k = \omega \mid X_k = \omega') = \Lambda(\omega, \omega')$  satisfies

$$\Lambda(\omega, \omega') = p_0^{|R|-1} (1 - p_1) \left( \frac{1 - p_0}{p_0} \right)^{\langle \omega \rangle} \left( \frac{p_0 p_1}{(1 - p_0)(1 - p_1)} \right)^{\langle \omega \omega' \rangle},$$

where  $\langle \omega \rangle = \sum_{x \in R} \omega(x)$  and  $\langle \omega \omega' \rangle = \sum_{x \in R} \omega(x) \omega'(x)$ .

Let  $P$  be the joint law of the Markovian targets (with initial distribution  $\nu$ ) and the observations. Also, let  $Q$  be the joint law of the Markovian targets (with initial distribution  $\nu$ ) and independent Bernoulli observations with mean  $\frac{1}{2}$ . Furthermore, let  $\mathcal{G}_n$  be the  $\sigma$ -algebra generated by  $Y_1, \dots, Y_n, X_0, \dots, X_n$ . Then it is easy to check that, with respect to  $\mathcal{G}_n$ ,  $P$  is equivalent to  $Q$  and

$$\frac{dP}{dQ} \Big|_{\mathcal{G}_n} = \prod_{j=1}^n 2^{|R|} \Lambda(Y_j, X_j).$$

Furthermore, define  $L_n = \prod_{j=1}^n \Lambda(Y_j, X_j)$ . Denoting by  $E_P$  and  $E_Q$  the expectation with respect to  $P$  and  $Q$ , respectively, observe that, for any  $f \in \mathcal{B}_b(\mathfrak{X})$ , we have

$$\hat{\eta}_n(f) = E_P(f(X_n) \mid \mathcal{Y}_n) = \frac{E_Q(f(X_n)L_n \mid \mathcal{Y}_n)}{E_Q(L_n \mid \mathcal{Y}_n)}. \tag{4.2}$$

This formula is a consequence of the properties of the conditional expectations and, in the context of filtering, (4.2) is known as the Kallianpur–Stribel formula; see, e.g. [16].

Denote by  $K$  the Markov kernel associated with the Markov chain  $(X_n)_{n \geq 0}$  defined by  $M$ , as in (4.1). We can check that  $\eta_n$  and  $\hat{\eta}_n$  satisfy (1.2) with  $g_n(x) = \Lambda(y_n, x)$  and  $K_n = K$ . Also, note that in this case  $g_n$  takes only two values, which can be assumed to belong to  $\mathbb{Q}$  (the set of all rational numbers) because of rounding errors. It follows from Remark A.2 in Appendix A that

$$\sup_{N_0 \geq 1} E(N_1 \|\eta_1^{N_1} - \eta_1\|^2) < \infty.$$

The results proved in Section 2 of [15] provide an algorithm for recursively computing the exact filter, i.e. the law of  $X_n$  given  $\mathcal{Y}_n$ . In the next subsection we will compare the results from the exact filter with those obtained using the Monte Carlo algorithm described in Section 2 with various sampling methods.

## 4.2. Simulation results

In what follows,  $R$  is chosen to be the window of size  $100 \times 100$  defined by  $R = \{0, \dots, 99\}^2$ . To make things simple, the target starts at  $(50, 50)$  and moves according to a simple symmetric random walk, i.e. it goes up, down, right, or left to the nearest neighbor with probability  $\frac{1}{4}$ . The estimation of the position of the target is taken to be the mean of the various measures. The simulations were performed with  $p_0 = 0.9$  and  $p_1 = 0.9$ , with a 10% error in the pixels.

We use two methods of comparison for the sampling methods described in Section 2, i.e. simple random sampling (SRS), remainder stochastic sampling (RSS), systematic sampling (SyS), binomial sampling (BiS), and Bernoulli sampling (BeS).

The first method of comparison between the selection methods and the optimal filter (OF) is based on the mean square error. We computed  $\|\eta_n^N - \eta_n\|_2$  for different values of  $n$  (1, 2, 3, 4, 5, and 100) and  $N$  (1000, 10 000, 30 000, and 50 000). These results are reported in Table 1. We find that Conjecture 3.1 holds and that there is no significant difference, in terms of the precision as measured by  $\|\eta_n^N - \eta_n\|_2$ , between the selection methods.

The second method of comparison illustrates how efficient the sampling methods are at estimating the exact position of the target. We also compute the efficiency of the OF. More precisely, we compute the mean absolute error between the estimated position and the true position, as calculated over several time intervals, namely  $[2, 100]$ ,  $[10, 100]$ , and  $[30, 100]$ .

TABLE 1: Differences in  $L^2$  between the particle filters and the optimal filter for one target performing a simple symmetric random walk in images of size  $100 \times 100$  with a 10% error in the pixels.

$N$	Sampling method	$n$					
		1	2	3	4	5	100
1000	SRS	0.27	0.78	0.79	0.81	1.0	1.1
	RSS	0.12	0.37	0.64	0.65	0.68	1.3
	SyS	0.07	0.28	0.31	0.85	0.54	1.1
	BiS	0.04	0.058	0.087	0.15	0.17	0.93
	BeS	0.07	0.16	0.21	0.65	0.69	1.3
10 000	SRS	0.031	0.11	0.26	0.40	0.36	1.2
	RSS	0.039	0.12	0.40	0.50	0.91	1.2
	SyS	0.025	0.13	0.22	0.38	0.40	1.3
	BiS	0.028	0.049	0.082	0.15	0.17	0.96
	BeS	0.028	0.07	0.11	0.23	0.25	1.1
30 000	SRS	0.018	0.10	0.23	0.22	0.50	1.4
	RSS	0.17	0.10	0.23	0.21	0.56	1.4
	SyS	0.018	0.098	0.17	0.20	0.25	1.0
	BiS	0.027	0.048	0.8	0.15	0.17	0.95
	BeS	0.018	0.056	0.10	0.17	0.22	0.97
50 000	SRS	0.013	0.094	0.26	0.22	0.37	1.2
	RSS	0.021	0.068	0.18	0.26	0.39	1.3
	SyS	0.012	0.095	0.26	0.24	0.45	1.2
	BiS	0.02	0.04	0.08	0.15	0.17	0.70
	BeS	0.017	0.052	0.086	0.16	0.20	1.2

TABLE 2: Mean absolute error for one target performing a simple symmetric random walk in images of size  $100 \times 100$  with a 10% error in the pixels.

$N$	Sampling method	$n$		
		[2, 100]	[10, 100]	[30, 100]
1000	OF	4.1	2.9	0.8
	SRS	57.4	60.3	56.2
	RSS	51.8	53.6	43.8
	SyS	42.7	43.7	36.9
	BiS	54.0	56.5	45.5
	BeS	13.8	12.1	6.9
10 000	SRS	76.0	81.0	64.1
	RSS	1.9	0.8	0.5
	SyS	2.4	0.7	0.5
	BiS	6.4	6.7	0.5
	BeS	77.5	82.5	85.4
	30 000	SRS	8.7	3.0
RSS		2.4	0.6	0.4
SyS		3.9	1.5	0.4
BiS		4.0	2.1	0.5
BeS		8.1	6.2	0.4
50 000		SRS	3.9	3.4
	RSS	10.2	5.2	0.7
	SyS	5.0	2.5	0.6
	BiS	4.8	2.1	0.8
	BeS	3.6	1.5	0.3

The number of particles  $N$  takes the values 1000, 10 000, 30 000, and 50 000. The results are reported in Table 2.

According to these results, we conclude that the algorithm based on the systematic sampling selection method performs quite well, provided that the number of particles is large enough. Again, we also conclude that there is no significant difference, in terms of the precision as measured by the mean absolute error, between the selection methods. Furthermore, based on the results of Table 2 for the time interval [30, 100], note that when the target is precisely detected, the error seems to stabilize near 0, indicating that the distance between  $\eta_n^N$  and  $\eta_n$  might be uniform on  $n$ .

Surprisingly, the Monte Carlo based approximate filters seem to perform better than the optimal filter, but recall that the optimal filter is optimal when the precision is measured by the mean square error.

Finally, let us recall that in terms of the two measures of precision chosen here, there is no significant difference between the different selection methods. We never anticipated that a simple method (in terms of the computational complexity), like the systematic sampling selection method, would be more precise than other proven selection methods. We are just glad that it performs as well as its competitors and that its main advantage is its simplicity. Furthermore, the numerical comparisons have showed that Conjecture 3.1 seems plausible, at least in the special case considered here.

**Appendix A. Convergence results for a Markov chain**

Suppose that  $(X_i, Y_i)_{i \geq 1}$  are independent observations of  $(X, Y) \in \mathfrak{Z} := \mathbb{R} \times [0, 1)$  of law  $\mathbb{P}$  with marginal distributions  $\mathbb{P}_X$  and  $\mathbb{P}_Y$ , respectively. Furthermore, let  $\lambda$  denote Lebesgue’s measure on  $[0, 1)$ . Given  $Z_0 = (X_0, S_0) \in \mathbb{R} \times [0, 1)$ , set  $Z_i = (X_i, \{S_i\})$ , where  $S_i = S_{i-1} + Y_i, i \geq 1$ .

For  $n \in \mathbb{Z}$ , set  $e_n(s) = e^{2\pi i n s}, s \in [0, 2)$ , and let  $\zeta_n = \mathbb{E}(e_n(Y))$ . Furthermore, set  $\mathcal{N} = \{n \in \mathbb{Z} : \zeta_n = 1\}$ . Recall that  $(e_n)_{n \in \mathbb{Z}}$  is a complete orthonormal basis of the Hilbert space  $H = L^2([0, 1), \lambda)$  with scalar product  $(f, g) = \int_0^1 f(s)\bar{g}(s) ds$  and norm  $\|f\|_2 = \sqrt{(f, f)}$ .

It is easy to check that  $(Z_i)_{i \geq 0}$  is a Markov chain on  $\mathfrak{Z}$  with kernel  $\mathcal{K}$  defined by

$$\mathcal{K}f(x, s) := \int_{\mathfrak{Z}} f(x', \{s + y\}) \mathbb{P}(dx', dy), \quad f \in \mathcal{B}_b(\mathfrak{Z}), \tag{A.1}$$

and stationary distribution  $\mu = \mathbb{P}_X \otimes \lambda$ . Note that, for any  $f \in L^2(\mu)$ , by Tonelli’s theorem,  $\mathcal{K}f$  is well defined, depends only on  $s \in [0, 1)$ , and belongs to  $H$  since

$$\begin{aligned} \|\mathcal{K}f\|_2^2 &\leq \int_0^1 \int_{\mathfrak{Z}} f^2(x, \{s + y\}) \mathbb{P}(dx, dy) ds \\ &= \int_{\mathfrak{Z}} \int_0^1 f^2(x, u) du \mathbb{P}(dx, dy) \\ &= \int_{\mathfrak{Z}} f^2(z) \mu(dz) \\ &= \|f\|_{L^2(\mu)}^2. \end{aligned}$$

Finally, let  $\mathcal{L}$  and  $\mathcal{A}$  be the linear bounded operators from  $L^2(\mu)$  to  $H$  defined by

$$\mathcal{L}f(s) = \sum_{n \in \mathcal{N}} (Kf, e_n) e_n(s) \quad \text{and} \quad \mathcal{A}f(s) = \int_{\mathbb{R}} f(x, s) \mathbb{P}_X(dx), \quad s \in [0, 1).$$

**Theorem A.1.** *Let  $f \in L^2(\mu)$  be given, and set  $W_N = (1/N) \sum_{k=1}^N f(Z_k)$ . Then the following assertions hold.*

- (i) *If the initial distribution of  $Z_0 = (X_0, S_0)$  is  $\mu$  then  $W_N$  converges almost surely and in mean square to  $\mathcal{W}$  given by*

$$\mathcal{W} = \mathcal{L}f(S_0) = \sum_{n \in \mathcal{N}} (\mathcal{K}f, e_n) e_n(S_0). \tag{A.2}$$

*If, in addition,*

$$\sum_{n \in \mathbb{Z} \setminus \mathcal{N}} \frac{|(\mathcal{K}f, e_n)| |(\mathcal{A}f, e_n)|}{|1 - \zeta_n|} < \infty \tag{A.3}$$

*then  $N \mathbb{E}((W_N - \mathcal{W})^2)$  converges as  $N \rightarrow \infty$  to*

$$\|f\|_{L^2(\mu)}^2 - \|\mathcal{L}f\|_2^2 + 2 \sum_{n \in \mathbb{Z} \setminus \mathcal{N}} \frac{(\mathcal{K}f, e_n) \overline{(\mathcal{A}f, e_n)}}{1 - \zeta_n}. \tag{A.4}$$

(ii) If the initial distribution of  $Z_0 = (X_0, S_0)$  is  $\mu$ , if  $\mathcal{N} = \{0\}$ , and if

$$\sum_{n \in \mathbb{Z} \setminus \{0\}} \frac{|(\mathcal{K}f, e_n)|^2}{|1 - \zeta_n|^2} < \infty, \tag{A.5}$$

then the sequence of processes  $B_N$ , defined by  $B_N(t) = \sqrt{N}(W_{\lfloor Nt \rfloor} - \mu(f))$ ,  $t \in [0, 1]$ , converges in  $D([0, 1])$  to  $\sigma B$ , where  $B$  is a Brownian motion and  $\sigma^2$  is given by (A.4).

(iii) If  $P_Y$  admits a square integrable density  $h$  then the Markov chain is geometrically ergodic, that is, there exists  $\rho \in (0, 1)$  such that, for any  $f \in L^2(\mu)$ ,

$$|\mathcal{K}^n f(Z_0) - \mu(f)| \leq \|h\|_2 \rho^{n-2} \|f\|_{L^2(\mu)}, \quad n \geq 2.$$

*Proof.* For simplicity, set  $\psi = \mathcal{K}f \in H$ . To prove (i), start the Markov chain from  $\mu$  and denote the law of the chain by  $Q$ . Then the sequence  $(Z_n)_{n \geq 0}$  is stationary and Birkhoff's ergodic theorem (see, e.g. [14, Section 6.2]) can be invoked to claim that  $W_N$  converges almost surely and in mean square to some random variable  $\mathcal{W}$ . To show that  $\mathcal{W}$  is indeed given by (A.2), it suffices to show that  $E((W_N - \mathcal{W})^2) \rightarrow 0$  as  $N \rightarrow \infty$ . First, note that  $E(\mathcal{W}^2) = \|\mathcal{L}f\|_2^2 = \sum_{n \in \mathcal{N}} |(\psi, e_n)|^2$ . Next, set  $\varphi(s) = \mathcal{A}f(s)$ . If  $n \in \mathcal{N}$  then  $e_n(Y) = 1$   $P$ -almost surely and it follows, by Fubini's theorem, that

$$\begin{aligned} (\psi, e_n) &= \int_0^1 \int_{\mathfrak{I}_3} f(x, \{s + y\}) \overline{e_n(s)} P(dx, dy) ds \\ &= \int_0^1 \int_{\mathfrak{I}_3} f(x, u) \overline{e_n(u)} P(dx, dy) du \\ &= (\varphi, e_n). \end{aligned}$$

As a result,  $E(\mathcal{W}^2) = \sum_{n \in \mathcal{N}} |(\varphi, e_n)|^2$ . Next, using the fact that, for any  $k \in \mathbb{Z}$  and any  $s, y \in [0, 1)$ , we have  $e_k(\{s + y\}) = e_k(s + y) = e_k(s)e_k(y)$ , it follows that

$$\mathcal{K}e_k(s) = \int_{\mathfrak{I}_3} e_k(\{s + y\}) P(dx, dy) = \int_{[0,1)} e_k(\{s + y\}) P_Y(dy) = \zeta_k e_k(s), \quad s \in [0, 1).$$

Hence, for any  $k \geq 1$  and any  $n \in \mathbb{Z}$ , we obtain

$$\mathcal{K}^k e_n = \zeta_n^k e_n. \tag{A.6}$$

Now, using the Markov property of the chain together with (A.6), we have

$$\begin{aligned} E(W_N \mathcal{W}) &= E(W_N \overline{\mathcal{W}}) \\ &= \frac{1}{N} \sum_{k=1}^N \sum_{n \in \mathcal{N}} \overline{(\psi, e_n)} E(f(Z_k) \overline{e_n(S_0)}) \\ &= \frac{1}{N} \sum_{k=1}^N \sum_{n \in \mathcal{N}} \overline{(\psi, e_n)} E(\mathcal{K}^k f(S_0) \overline{e_n(S_0)}) \\ &= \frac{1}{N} \sum_{k=1}^N \sum_{n \in \mathcal{N}} \overline{(\psi, e_n)} (\mathcal{K}^{k-1} \psi, e_n) \\ &= \frac{1}{N} \sum_{j \in \mathbb{Z}} \sum_{k=1}^N \sum_{n \in \mathcal{N}} (\psi, e_j) \overline{(\psi, e_n)} (\mathcal{K}^{k-1} e_j, e_n) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{N} \sum_{j \in \mathbb{Z}} \sum_{k=1}^N \sum_{n \in \mathcal{N}} (\psi, e_j) \overline{(\psi, e_n)} \zeta_j^{k-1} (e_j, e_n) \\
 &= \frac{1}{N} \sum_{k=1}^N \sum_{n \in \mathcal{N}} |(\psi, e_n)|^2 \zeta_n^{k-1} \\
 &= \mathbb{E}(\mathcal{W}^2),
 \end{aligned}$$

since, by definition,  $\zeta_n = 1$  for any  $n \in \mathcal{N}$ .

Next, using stationarity, the Markov property, (A.6), and the identity

$$\frac{2}{N^2} \sum_{k=1}^{N-1} \sum_{j=1}^k z^{j-1} = \frac{N-1}{1-z} - \frac{z-z^N}{(1-z)^2}, \quad z \in \mathbb{C}, z \neq 1,$$

it follows that

$$\begin{aligned}
 \mathbb{E}(W_N^2) &= \frac{1}{N} \mathbb{E}(f^2(Z_0)) + \frac{2}{N^2} \sum_{k=1}^{N-1} \sum_{j=1}^k \mathbb{E}(\mathcal{K}^{j-1} \psi(S_0) f(Z_0)) \\
 &= \frac{1}{N} \|f\|_{L^2(\mu)}^2 + \frac{2}{N^2} \sum_{k=1}^{N-1} \sum_{j=1}^k (\mathcal{K}^{j-1} \psi, \varphi) \\
 &= \frac{1}{N} \|f\|_{L^2(\mu)}^2 + \sum_{n \in \mathbb{Z}} (\psi, e_n) \overline{(\varphi, e_n)} \left( \frac{2}{N^2} \sum_{k=1}^{N-1} \sum_{j=1}^k \zeta_n^{j-1} \right) \\
 &= \frac{1}{N} \|f\|_{L^2(\mu)}^2 + \frac{N-1}{N} \sum_{n \in \mathcal{N}} (\psi, e_n) \overline{(\varphi, e_n)} \\
 &\quad + \frac{2}{N^2} \sum_{n \in \mathbb{Z} \setminus \mathcal{N}} (\psi, e_n) \overline{(\varphi, e_n)} \left( \frac{N-1}{1-\zeta_n} - \frac{\zeta_n - \zeta_n^N}{(1-\zeta_n)^2} \right) \\
 &= \frac{1}{N} \|f\|_{L^2(\mu)}^2 + \frac{N-1}{N} \mathbb{E}(\mathcal{W}^2) \\
 &\quad + \frac{2}{N^2} \sum_{n \in \mathbb{Z} \setminus \mathcal{N}} (\psi, e_n) \overline{(\varphi, e_n)} \left( \frac{N-1}{1-\zeta_n} - \frac{\zeta_n - \zeta_n^N}{(1-\zeta_n)^2} \right).
 \end{aligned}$$

Collecting the expressions obtained for  $\mathbb{E}(W_N^2)$  and  $\mathbb{E}(W_N \mathcal{W})$ , we obtain

$$\begin{aligned}
 \mathbb{E}((W_N - \mathcal{W})^2) &= \frac{1}{N} \|f\|_{L^2(\mu)}^2 - \frac{1}{N} \mathbb{E}(\mathcal{W}^2) \\
 &\quad + \frac{2}{N^2} \sum_{n \in \mathbb{Z} \setminus \mathcal{N}} (\psi, e_n) \overline{(\varphi, e_n)} \left( \frac{N-1}{1-\zeta_n} - \frac{\zeta_n - \zeta_n^N}{(1-\zeta_n)^2} \right). \tag{A.7}
 \end{aligned}$$

Since  $\sum_{n \in \mathbb{Z} \setminus \mathcal{N}} |(\psi, e_n)| |(\varphi, e_n)|$  is finite,

$$\sup_{n \in \mathbb{Z} \setminus \mathcal{N}} \left| \frac{N-1}{1-\zeta_n} - \frac{\zeta_n - \zeta_n^N}{(1-\zeta_n)^2} \right| = \left| \sum_{k=1}^{N-1} \sum_{j=1}^k \zeta_n^{j-1} \right| \leq \frac{N^2}{2},$$

it follows, from (A.7) and the dominated convergence theorem, that

$$\lim_{N \rightarrow \infty} \mathbb{E}((W_N - \mathcal{W})^2) = 0,$$

and under the additional condition (A.3), we also obtain

$$\lim_{N \rightarrow \infty} N \mathbb{E}((W_N - \mathcal{W})^2) = \|f\|_{L^2(\mu)}^2 - \mathbb{E}(\mathcal{W}^2) + 2 \sum_{n \in \mathbb{Z} \setminus \mathcal{N}} \frac{(\psi, e_n) \overline{(\varphi, e_n)}}{1 - \zeta_n}.$$

This completes the proof of (i).

The proof of (ii) is inspired by Durrett [14]. First, note that, since  $\mathcal{N} = \{0\}$ ,  $\mathcal{L}f = \mu(f)$  for any  $f \in L^2(\mu)$  and it follows from (i) that  $(1/N) \sum_{k=1}^N f(Z_k)$  converges almost surely and in  $L^p$  to  $\mu(f)$  for any  $1 \leq p \leq 2$ . Moreover, given any  $f \in L^1(\mu)$ , we can find  $f_n \in L^2(\mu)$  such that  $\|f - f_n\|_{L^1(\mu)} < 1/n$ . It follows that, for any  $n \geq 1$ ,

$$\limsup_{N \rightarrow \infty} \mathbb{E} \left( \left| \frac{1}{N} \sum_{k=1}^N f(Z_k) - \mu(f) \right| \right) \leq \frac{2}{n} + \limsup_{N \rightarrow \infty} \mathbb{E} \left( \left| \frac{1}{N} \sum_{k=1}^N f_n(Z_k) - \mu(f_n) \right| \right) = \frac{2}{n}.$$

Since the latter is true for any  $n \geq 1$ , we conclude that  $(1/N) \sum_{k=1}^N f(Z_k)$  converges in  $L^1$  to  $\mu(f)$ . By Birkhoff's ergodic theorem,  $(1/N) \sum_{k=1}^N f(Z_k)$  converges almost surely to  $\mu(f)$ . Next, let  $D$  be the subset of  $H$  defined by

$$D = \left\{ h \in H : \sum_{n \in \mathbb{Z} \setminus \{0\}} \frac{|(h, e_n)|^2}{|1 - \zeta_n|^2} < \infty \right\},$$

and let  $\Xi$  be the operator from  $D$  to  $H$  that satisfies

$$\Xi h = \sum_{n \in \mathbb{Z} \setminus \{0\}} \frac{(h, e_n)}{1 - \zeta_n} e_n.$$

Note that, since  $(I - \mathcal{K})\Xi h = (I - \mathcal{L})h$ ,  $\Xi = (I - \mathcal{K})^{-1}(I - \mathcal{L})$  on  $D$ . Let  $\mathcal{D}$  be the set of all  $f \in L^2(\mu)$  such that  $f$  satisfies (A.5), i.e.  $\mathcal{K}f \in D$ . Then  $\Xi$  can be extended to a mapping from  $\mathcal{D}$  to  $L^2(\mu)$  viz.  $\Xi f = (I - \mathcal{L})f + \Xi \mathcal{K}f$ . Using  $\mathcal{K}\mathcal{L} = \mathcal{L}\mathcal{K} = \mathcal{L}$ , we obtain  $\Xi = (I - \mathcal{K})^{-1}(I - \mathcal{L})$  on  $\mathcal{D}$ . Next, if  $f \in \mathcal{D}$ , set  $g = \Xi f$ . Since  $\mathcal{L}f = \mu(f)$ , it follows that

$$\sqrt{N}(W_N - \mu(f)) = \frac{1}{\sqrt{N}} \sum_{k=1}^N (g(Z_k) - \mathcal{K}g(\{S_{k-1}\})) + \frac{1}{\sqrt{N}} \mathcal{K}g(S_0) - \frac{1}{\sqrt{N}} \mathcal{K}g(\{S_N\}).$$

Now, setting  $\mathcal{F}_k = \sigma\{Z_j : j \leq k\}$ , the terms  $\xi_k = g(Z_k) - \mathcal{K}g(\{S_{k-1}\})$  are square-integrable martingale differences with respect to  $(\mathcal{F}_j)_{j \geq 0}$ , i.e.  $\mathbb{E}(\xi_k | \mathcal{F}_{k-1}) = 0$ , and because  $g^2$  and  $(\mathcal{K}g)^2$  both belong to  $L^1(\mu)$  it follows from (i), as shown above, that

$$\frac{1}{N} \sum_{k=1}^N \mathbb{E}(\xi_k^2 | \mathcal{F}_{k-1}) = \frac{1}{N} \sum_{k=1}^N (\mathcal{K}g^2(\{S_{k-1}\}) - (\mathcal{K}g)^2(\{S_{k-1}\}))$$



converges almost surely to  $\mu(g^2) - \mu((\mathcal{K}g)^2)$ . Note that, since  $\mathcal{K}g = \Xi\mathcal{K}f$ , we have  $(\mathcal{K}g, \mathcal{L}f) = 0$  and expression (A.4) can be written as

$$\begin{aligned} \sigma^2 &= \|(I - \mathcal{L})f\|_{L^2(\mu)}^2 + 2(\Xi\mathcal{K}f, \mathcal{A}f) \\ &= \|(I - \mathcal{K})g\|_{L^2(\mu)}^2 + 2(\mathcal{K}g, \mathcal{A}f) \\ &= \|(I - \mathcal{K})g\|_{L^2(\mu)}^2 + 2(\mathcal{K}g, \mathcal{A}(I - \mathcal{L})f) \\ &= \|(I - \mathcal{K})g\|_{L^2(\mu)}^2 + 2(\mathcal{K}g, \mathcal{A}(I - \mathcal{K})g) \\ &= \mu(g^2 - 2g\mathcal{K}g + (\mathcal{K}g)^2) + 2\mu(g\mathcal{K}g) - 2\mu((\mathcal{K}g)^2) \\ &= \mu(g^2) - \mu((\mathcal{K}g)^2). \end{aligned}$$

Finally, because of the stationarity of  $(\xi_k)_{k \geq 1}$ , it follows that, for any  $\varepsilon > 0$ ,

$$\frac{1}{N} \sum_{k=1}^N \mathbb{E}(\xi_k^2 \mathbf{1}(|\xi_k| > \varepsilon\sqrt{N})) = \mathbb{E}(\xi_1^2 \mathbf{1}(|\xi_1| > \varepsilon\sqrt{N})) \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

The conditions of Theorem 7.4 of [14] are all met, so we may safely conclude that by defining the process  $B_N(t) = \sqrt{N}(W_{\lfloor Nt \rfloor} - \mu(f))$ ,  $t \in [0, 1]$ ,  $B_N$  converges in  $D[0, 1]$  to  $\sigma B$ , where  $B$  is a Brownian motion.

To prove (iii), first note that, since the density  $h$  of  $Y$  is square integrable, then  $\mathcal{N} = \{0\}$ ,  $\sup_{n \geq 1} |\zeta_n| = \rho < 1$ ,  $\zeta_n = (e_n, h)$ , and  $\|h\|_2^2 = \sum_{n \in \mathbb{Z}} |\zeta_n|^2$ . Therefore, for any  $g \in H$ ,

$$\sum_{n \in \mathbb{Z}} |(g, e_n)| |\zeta_n| \leq \|g\|_2 \|h\|_2 < \infty.$$

It follows that, for any  $k \geq 2$ ,

$$\mathcal{K}^k f = \mathcal{K}^{k-1} \psi = \sum_{n \in \mathbb{Z}} (\psi, e_n) \zeta_n^{k-1} e_n,$$

with the series converging absolutely. Thus,

$$\begin{aligned} \sup_{s_0 \in \mathfrak{I}} |\mathcal{K}^k f(s) - \mu(f)| &= \sup_{s \in [0, 1]} |\mathcal{K}^{k-1} \psi(s) - \lambda(\psi)| \\ &\leq \sum_{n \in \mathbb{Z} \setminus \{0\}} |(\psi, e_n)| |\zeta_n| \rho^{k-2} \\ &\leq \|h\|_2 \|f\|_{L^2(\mu)} \rho^{k-2}. \end{aligned}$$

This completes the proof of the theorem.

**Remark A.1.** Note that if  $\zeta_n = 1$  for some  $n > 0$  then  $k \mapsto \zeta_k$  is  $n$ -periodic, so  $\{\zeta_k : n \in \mathbb{Z} \setminus \mathcal{N}\}$  is finite. Therefore,  $\sup_{k \in \mathbb{Z} \setminus \mathcal{N}} |\zeta_k| = \rho < 1$  and condition (A.3) is satisfied. Also, if  $P_Y$  has a nondegenerate absolutely continuous part then  $\mathcal{N} = \{0\}$  and  $\sup_{n \geq 1} |\zeta_n| = \rho < 1$ , so condition (A.3) holds.

The next result is a straightforward extension of the previous theorem. Before stating it, denote by  $\nu$  the joint law of  $(Z_1, S_0)$ , where  $S_0 \sim \text{Unif}([0, 1])$ .

**Theorem A.2.** Suppose that  $f \in L^2(\nu)$ , and set  $W_N = (1/N) \sum_{k=1}^N f(Z_k, \{S_{k-1}\})$ . Then the following assertions hold.

- (i) If the initial distribution of  $Z_0 = (X_0, S_0)$  is  $\mu$  then  $W_N$  converges almost surely and in mean square to  $\mathcal{W}$  given by (A.2), where

$$\mathcal{K}f(s) = \int_{\mathfrak{Z}} f(x, \{s + y\}, s) P(dx, dy).$$

If  $\mathcal{A}f(s) = \int_{\mathfrak{Z}} f(x, s, \{s - y\}) P(dx, dy)$ ,  $s \in [0, 1)$ , and if, in addition,

$$\sum_{n \in \mathbb{Z} \setminus \mathcal{N}} \frac{|(\mathcal{K}f, e_n)| |(\mathcal{A}f, e_n)|}{|1 - \zeta_n|} < \infty, \tag{A.8}$$

then  $NE((W_N - \mathcal{W})^2)$  converges as  $N \rightarrow \infty$  to

$$\|f\|_{L^2(\nu)}^2 - \|\mathcal{L}f\|_2^2 + 2 \sum_{n \in \mathbb{Z} \setminus \mathcal{N}} \frac{(\mathcal{K}f, e_n) \overline{(\mathcal{A}f, e_n)}}{1 - \zeta_n}. \tag{A.9}$$

- (ii) If  $\mathcal{N} = \{0\}$ , if the initial distribution of  $Z_0$  is  $\mu$ , and if

$$\sum_{n \in \mathbb{Z} \setminus \{0\}} \frac{|(\mathcal{K}f, e_n)|^2}{|1 - \zeta_n|^2} < \infty, \tag{A.10}$$

then the sequence of processes  $B_N$ , defined by  $B_N(t) = \sqrt{N}(W_{[Nt]} - \mu(f))$ ,  $t \in [0, 1]$ , converges in  $D([0, 1])$  to  $\sigma B$ , where  $B$  is a Brownian motion and  $\sigma^2$  is given by (A.9).

- (iii) If  $P_Y$  admits a square-integrable density  $h$  then the Markov chain is geometrically ergodic, that is, there exists  $\rho \in (0, 1)$  such that, for any  $f \in L^2(\mu)$ ,

$$|\mathcal{K}^n f(Z_1, S_0) - \mu(f)| \leq \|h\|_2 \rho^{n-2} \|f\|_{L^2(\mu)}, \quad n \geq 2.$$

**Remark A.2.** Suppose that  $X_k$  is bounded, and set  $f(x, y, s) = x(y - s)$ . Then it is easy to check that, for any  $n \in \mathcal{N}$ ,

$$\begin{aligned} (\mathcal{K}f, e_n) &= \int_{\mathfrak{Z} \times [0, 1)} x(\{y + s\} - s) P(dx, dy) \overline{e_n(s)} ds \\ &= \int_{\mathfrak{Z} \times [0, 1)} xu(e_n(y) - 1) \overline{e_n(u)} P(dx, dy) du \\ &= 0, \end{aligned}$$

since  $P(e_n(Y) = 1) = 1$ . It follows from Theorem A.2 that

$$W_N = \frac{1}{N} \sum_{k=1}^N X_k(\{S_k\} - \{S_{k-1}\})$$

converges to 0 almost surely and in mean square.

Furthermore, if  $\text{card}(\mathcal{N}) > 1$  then condition (A.8) holds and  $\sup_{N \geq 1} NE(W_N^2) < \infty$ , while if  $P_Y$  is absolutely continuous then condition (A.10) holds and  $\sqrt{N}T_N$  converges in law to a centered Gaussian random variable with variance  $\sigma^2$  given by (A.9).

### Acknowledgement

The authors thank the anonymous referees for helpful remarks and suggestions that led to an improved version of the paper.

### References

- [1] BAKER, J. E. (1985). Adaptive selection methods for genetic algorithms. In *Proc. Internat. Conf. Genetic Algorithms App.*, Erlbaum, pp. 101–111.
- [2] BAKER, J. E. (1987). Reducing bias and inefficiency in the selection algorithm. In *Proc. Second Internat. Conf. Genetic Algorithms App.*, Erlbaum, pp. 14–21.
- [3] BRINDLE, A. (1980). Genetic algorithms for function optimization. Doctoral Thesis, Department of Computer and Communication Sciences, University of Michigan.
- [4] CAPPÉ, O., DOUC, R. AND MOULINES, E. (2005). Comparison of resampling schemes for particle filtering. In *4th Internat. Symp. Image Signal Process. Anal.* (Zagreb, Croatia).
- [5] CARPENTER, J., CLIFFORD, P. AND FEARNHEAD, P. (1999). Building robust simulation-based filters for evolving data sets. Tech. Rep., Department of Statistics, University of Oxford.
- [6] CRISAN, D. (2003). Exact rates of convergence for a branching particle approximation to the solution of the Zakai equation. *Ann. Prob.* **31**, 693–718.
- [7] CRISAN, D. AND DOUCET, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.* **50**, 736–746.
- [8] CRISAN, D. AND LYONS, T. (1999). A particle approximation of the solution of the Kushner–Stratonovitch equation. *Prob. Theory Relat. Fields* **115**, 549–578.
- [9] CRISAN, D., DEL MORAL, P. AND LYONS, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Process. Relat. Fields* **5**, 293–318.
- [10] CRISAN, D., GAINES, J. AND LYONS, T. (1998). Convergence of a branching particle method to the solution of the Zakai equation. *SIAM J. Appl. Math.* **58**, 1568–1590.
- [11] DEL MORAL, P. (2004). *Feynman–Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, New York.
- [12] DEL MORAL, P. AND MICLO, L. (2000). Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV* (Lecture Notes Math. **1729**), Springer, Berlin, pp. 1–145.
- [13] DOUC, R. AND MOULINES, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. To appear in *Ann. Statist.*
- [14] DURRETT, R. (1996). *Probability: Theory and Examples*, 2nd edn. Duxbury Press, Belmont, CA.
- [15] GENTIL, I., RÉMILLARD, B. AND DEL MORAL, P. (2005). Filtering of images for detecting multiple targets trajectories. In *Statistical Modeling and Analysis for Complex Data Problem*, Springer, New York, pp. 267–280.
- [16] KALLIANPUR, G. AND STRIEBEL, C. (1968). Estimation of stochastic systems: arbitrary system process with additive white noise observation errors. *Ann. Math. Statist.* **39**, 785–801.
- [17] KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: algorithms and theoretical analysis. *Ann. Statist.* **33**, 1983–2021.
- [18] LIU, J. S. AND CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93**, 1032–1044.
- [19] PITT, M. K. AND SHEPHARD, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Amer. Statist. Assoc.* **94**, 590–599.
- [20] SETHURAMAN, S. AND VARADHAN, S. R. S. (2005). A martingale proof of Dobrushin’s theorem for non-homogeneous Markov chains. *Electron. J. Prob.* **10**, 1221–1235.