

A SEQUENTIAL PARTICLE ALGORITHM THAT KEEPS THE PARTICLE SYSTEM ALIVE

François LeGland and Nadia Oudjane

IRISA / INRIA
Campus de Beaulieu
35042 RENNES Cédex, France

phone : (+33) 02 99 84 73 62
fax : (+33) 02 99 84 71 71
e-mail : leglandirisa.fr
web : www.irisa.fr/aspi/legland/

EDF, Division R&D
1 avenue du Général de Gaulle
92141 CLAMART Cédex, France

ABSTRACT

We consider the problem of approximating a nonlinear (unnormalized) Feynman–Kac flow, in the special case where the selection functions can take the zero value. We begin with a list of several important practical situations where this characteristics is present. We study next a sequential particle algorithm, proposed by Oudjane (2000), which guarantees that the particle system does not die. Among other results, we obtain a central limit theorem which relies on the result of Rényi (1957) for the sum of a random number of independent random variables.

1. INTRODUCTION

We consider a particle approximation of the linear (unnormalized) flow and of the associated nonlinear (normalized) flow, defined by

$$\langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^n g_k(X_k)], \quad \langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle},$$

for any bounded measurable test–function ϕ , where $\{X_k, k = 0, 1, \dots, n\}$ is a Markov chain with initial probability distribution η_0 and transition kernels $\{Q_k, k = 1, \dots, n\}$, and where $\{g_k, k = 0, 1, \dots, n\}$ are given bounded measurable nonnegative functions, known as selection or fitness functions. We make throughout this paper the minimal assumption that $\langle \gamma_n, 1 \rangle > 0$, or equivalently that $\langle \eta_0, g_0 \rangle > 0$ and $\langle \mu_{k-1} Q_k, g_k \rangle > 0$ for any $k = 1, \dots, n$, otherwise the problem is not well defined. Particle approximations, also known as sequential Monte Carlo methods, have numerous practical applications, see [6], including particle filtering, and have been extensively studied in [2]. We focus here on the special case where the selection functions can possibly take the zero value, which occurs in many practical situations

- simulation of rare events using an *importance splitting* approach [4],

- simulation of a Markov chain conditioned or constrained to visit a given sequence of subspaces of the state space (this includes tracking a mobile in the presence of obstacles : when the mobile is hidden behind an obstacle, occlusion occurs and no observation is available at all, however this information can still be used, with a selection function equal to the indicator function of the region hidden by the obstacle),
- simulation of a r.v. in the tail of a given probability distribution,
- nonlinear filtering with bounded observation noise,
- implementation of a robustification approach in nonlinear filtering, using a truncation of the likelihood function [9, 12],
- algorithms of approximate nonlinear filtering, where hidden state and observation are simulated jointly, and where the simulated observation is validated against the actual observation [3, 14], e.g. if there is no explicit expression available for the likelihood function, or if there does not even exist a likelihood function (nonadditive observation noise, noise–free observations, etc.).

2. NONSEQUENTIAL PARTICLE ALGORITHM

Recall that the evolution of the sequence $\{\mu_k, k = 0, 1, \dots, n\}$ is described by the following diagram

$$\mu_{k-1} \longrightarrow \eta_k = \mu_{k-1} Q_k \longrightarrow \mu_k = g_k \cdot \eta_k,$$

with initial condition $\mu_0 = g_0 \cdot \eta_0$, where the notation \cdot denotes the projective product. The idea behind the particle approach is to look for an approximation

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \delta_{\xi_k^i},$$

in the form of the weighted empirical probability distribution associated with the particle system $\Sigma_k^N =$

$(\xi_k^i, w_k^i, i = 1, \dots, N)$, where N denotes the number of particles. In practice, particles

- are selected within the current particle system according to the respective weights $(w_{k-1}^i, i = 1, \dots, N)$ (selection step),
- move according to the Markov kernel Q_k (mutation step),
- are weighted by evaluating the fitness function g_k (weighting step).

As a result, each particle ξ_k^i has the probability distribution $\mu_{k-1}^N Q_k$ and its weight w_k^i is proportional to $g_k(\xi_k^i)$, for any $i = 1, \dots, N$. If the function g_k can possibly take the zero value, and even if $\langle \mu_{k-1} Q_k, g_k \rangle > 0$, it can happen that $g_k(\xi_k^i) = 0$ for any $i = 1, \dots, N$, i.e. it can happen that the evaluation of the function g_k returns the zero value for all the particles generated at the end of the mutation step, in which case the particle systems dies out and the algorithm cannot go on. A reinitialization procedure has been proposed and studied in [3], in which the particle system is generated from an arbitrary restarting probability distribution ν whenever extinction occurs. Alternatively, one could be interested by the behavior of the algorithm until the extinction time of the particle system, defined by

$$\tau^N = \inf\{k \geq 1 : \langle S^N(\mu_{k-1}^N Q_k), g_k \rangle = 0\},$$

where the notation $S^N(\mu)$ denotes the empirical probability distribution associated with an N -sample with common probability distribution μ . Under the assumption $\langle \gamma_n, 1 \rangle > 0$, the extinction probability $\mathbb{P}[\tau^N \leq n]$ that the algorithm can not go on until the time instant n goes to zero with exponential rate [2, Theorem 7.4.1].

Example 2.1. [Binary selection] In the special case of binary selection functions (taking only the value 0 or 1), such as indicator functions of Borel subsets for instance, it holds

$$\langle \gamma_n, 1 \rangle = \mathbb{P}[g_0(X_0) = \dots = g_n(X_n) = 1],$$

and on the *good set* $\{\tau^N > n\}$

$$\langle \gamma_n^N, 1 \rangle = \prod_{k=0}^n \langle S^N(\mu_{k-1}^N Q_k), g_k \rangle = \prod_{k=0}^n \frac{|I_k^N|}{N},$$

where

$$I_k^N = \{i = 1, \dots, N : g_k(\xi_k^i) = 1\},$$

denotes the set of successful particles within an N -sample with common probability distribution $\mu_{k-1}^N Q_k$. In other words, the probability of a successful sequence is approximated as the product of the fraction of successful particles at each generation. Notice that the computational effort, i.e. the number N of simulated particles at each generation, is *fixed* in advance, whereas the number $|I_k^N|$ of successful particles at the k -th generation is *random*.

For the nonsequential particle algorithm with a constant number N of particles, the following results have been obtained : a nonasymptotic estimate [2, Theorem 7.4.3]

$$\begin{aligned} & \sup_{\phi : \|\phi\|=1} \mathbb{E}|1(\tau^N > n) \langle \mu_n^N, \phi \rangle - \langle \mu_n, \phi \rangle| \\ & \leq \frac{c_n}{\sqrt{N}} + \mathbb{P}[\tau^N \leq n], \end{aligned}$$

and a central limit theorem (see [2, Section 9.4] for a slightly different algorithm)

$$\sqrt{N}[1(\tau^N > n) \langle \mu_n^N, \phi \rangle - \langle \mu_n, \phi \rangle] \Longrightarrow \mathcal{N}(0, v_n(\phi))$$

as $N \uparrow \infty$, for any bounded measurable test function ϕ , with the following expression

$$v_n(\phi) = \sum_{k=0}^n \frac{\langle \eta_k, |g_k R_{k+1:n}[\phi - \langle \mu_n, \phi \rangle]|^2 \rangle}{\langle \eta_k, g_k R_{k+1:n} 1 \rangle^2},$$

for the asymptotic variance. In the expression above

$$\begin{aligned} R_{k+1:n} \phi(x) &= R_{k+1} \dots R_n \phi(x) \\ &= \mathbb{E}[\phi(X_n) \prod_{p=k+1}^n g_p(X_p) | X_k = x], \end{aligned}$$

where the nonnegative (unnormalized) kernel R_p is defined by $R_p(x, dx') = Q_p(x, dx') g_p(x')$ for any $p = 1, \dots, n$, and in particular for $\phi \equiv 1$

$$g_{k:n}(x) = g_k(x) R_{k+1:n} 1(x) = \mathbb{E}[\prod_{p=k}^n g_p(X_p) | X_k = x].$$

In the simple case where the fitness functions are positive, i.e. cannot take the zero value, these results are well-known and can be found in [5, Proposition 2.9, Corollary 2.20], where the proof relies on a central limit theorem for triangular arrays of martingale increments, or in [8, Theorem 4], where the same central limit theorem is obtained by induction.

3. SEQUENTIAL PARTICLE ALGORITHM

The purpose of this work is to study a sequential particle algorithm, already proposed in [11, 10], which automatically keeps the particle system alive, i.e. which ensures its non-extinction. For any level $H > 0$, and for any $k = 0, 1, \dots, n$, define the random number

$$N_k^H = \inf\{N \geq 1 : \sum_{i=1}^N g_k(\xi_k^i) \geq H \sup_{x \in E} g_k(x)\},$$

of particles, where the r.v.'s $\xi_k^1, \dots, \xi_k^i, \dots$ are i.i.d. with common probability distribution η_0 (for $k = 0$),

and common probability distribution $\mu_{k-1}^H Q_k$ (for $k = 1, \dots, n$). The particle approximation

$$\mu_k \approx \mu_k^H = \sum_{i=1}^{N_k^H} w_k^i \delta_{\xi_k^i},$$

is now parameterized by the level $H > 0$, and is associated with the particle system $\Sigma_k^H = (\xi_k^i, w_k^i, i = 1, \dots, N_k^H)$, where N_k^H denotes the random number of particles. In practice, particles

- are selected within the current particle system according to the respective weights $(w_{k-1}^i, i = 1, \dots, N_{k-1}^H)$ (selection step),
- move according to the Markov kernel Q_k (mutation step),
- are weighted by evaluating the fitness function g_k (weighting step).

As a result, each particle ξ_k^i has the probability distribution $\mu_{k-1}^H Q_k$ and its weight w_k^i is proportional to $g_k(\xi_k^i)$, for $i = 1, \dots, N_k^H$. If $\langle \mu_{k-1}^H Q_k, g_k \rangle > 0$ — a sufficient condition for which is

$$\widehat{g}_k(x) = Q_k g_k(x) = \mathbb{E}[g_k(X_k) | X_{k-1} = x] > 0,$$

for any x in the support of μ_{k-1}^H — then

$$\langle \mu_{k-1}^H Q_k, g_k \rangle = \langle \mu_{k-1}^H, \widehat{g}_k \rangle > 0,$$

and the random number N_k^H of particles is a.s. finite, with $N_k^H \geq H$ obviously. Moreover

$$\begin{aligned} \langle S_k^{N_k^H}(\mu_{k-1}^H Q_k), g_k \rangle &= \frac{1}{N_k^H} \sum_{i=1}^{N_k^H} g_k(\xi_k^i) \\ &\geq \frac{H}{N_k^H} \sup_{x \in E} g_k(x) > 0, \end{aligned}$$

by construction, i.e. the particle system never dies out and the algorithm can always go on. We can show in addition that

$$\frac{N_k^H}{H} \longrightarrow \rho_k = \frac{\sup_{x \in E} g_k(x)}{\langle \mu_{k-1}^H Q_k, g_k \rangle} < \infty,$$

in probability, with rate $1/\sqrt{H}$.

Example 3.1. [Binary selection] In the special case of binary selection functions (taking only the value 0 or 1), such as indicator functions of Borel subsets for instance, it holds

$$\langle \gamma_n^H, 1 \rangle = \prod_{k=0}^n \langle S_k^{N_k^H}(\mu_{k-1}^H Q_k), g_k \rangle = \prod_{k=0}^n \frac{H}{N_k^H},$$

where in this case, for any integer $H \geq 1$

$$N_k^H = \inf\{N \geq 1 : |I_k^N| = H\},$$

and for any integer $N \geq 1$

$$I_k^N = \{i = 1, \dots, N : g_k(\xi_k^i) = 1\},$$

denotes the set of successful particles within an N -sample with common probability distribution $\mu_{k-1}^H Q_k$. Here again, the probability of a successful sequence is approximated as the product of the fraction of successful particles at each generation. In opposition to the nonsequential algorithm, notice that the number H of successful particles at each generation is *fixed* in advance, whereas the computational effort, i.e. the number N_k^H of simulated particles needed to get H successful particles exactly at the k -th generation, is *random*.

For the sequential particle algorithm, with a random number of particles defined by the level $H > 0$, the following nonasymptotic estimate has been obtained in [10, Theorem 5.4]

$$\sup_{\phi: \|\phi\|=1} \mathbb{E}|\langle \mu_n^H - \mu_n, \phi \rangle| \leq \frac{c'_n}{\sqrt{H}},$$

where it is formulated under a mixing assumption which is not needed here but gives a constant c'_n uniformly bounded in time. The main contribution of this paper is the following central limit theorem.

Theorem 3.2. *If $\langle \mu_{k-1}^H, \widehat{g}_k \rangle > 0$ for any $k = 1, \dots, n$, then*

$$\sqrt{H} \langle \mu_n^H - \mu_n, \phi \rangle \Longrightarrow \mathcal{N}(0, v_n^{\text{seq}}(\phi))$$

as $H \uparrow \infty$, for any bounded measurable test function ϕ , with the following expression

$$v_n^{\text{seq}}(\phi) = \sum_{k=0}^n \frac{\langle \eta_k, |g_k R_{k+1:n}[\phi - \langle \mu_n, \phi \rangle]|^2 \rangle}{\langle \eta_k, g_k R_{k+1:n} 1 \rangle^2} \frac{1}{\rho_k},$$

for the asymptotic variance.

Two different proofs can be given for Theorem 3.2. A first proof follows the approach of [5, Proposition 2.9, Corollary 2.20], and relies on an enumeration of all particles across generations with random sizes, and on a central limit theorem for triangular arrays of martingale increments [1, Theorem 2.8.42]. An alternate proof can also be given, which follows the approach of [8, Theorem 4] by induction, and relies on a central limit theorem for the sum of a random number of random variables [13], which is known in sequential analysis since the 1950's, see also [7, Theorem I.3.1] or [15, Theorem 2.40].

4. CONCLUSION

To get a fair comparison of the nonsequential and sequential particle approximations, we can use the time-average

$$\frac{1}{n+1} \sum_{k=0}^n N_k^H,$$

of the number of particles, which is an indication of how much computing power has been used, as a normalizing factor instead of the level $H > 0$. Since

$$\frac{1}{H} \left[\frac{1}{n+1} \sum_{k=0}^n N_k^H \right] \longrightarrow \frac{1}{n+1} \sum_{k=0}^n \rho_k,$$

in probability, and using the Slutsky lemma, we obtain

$$\left[\frac{1}{n+1} \sum_{k=0}^n N_k^H \right]^{1/2} \langle \mu_n^H - \mu_n, \phi \rangle \implies \mathcal{N}(0, v_n^*(\phi)),$$

as $H \uparrow \infty$, for any bounded measurable test function ϕ , with the following expression for the asymptotic variance

$$v_n^*(\phi) = \left[\frac{1}{n+1} \sum_{k=0}^n \rho_k \right] v_n^{\text{seq}}(\phi).$$

Notice that the asymptotic variance $v_n(\phi)$ for the non-sequential particle algorithm and the asymptotic variance $v_n^*(\phi)$ for the renormalized sequential particle algorithm coincide, in the special case where $\rho_0 = \rho_1 = \dots = \rho_n$.

5. ACKNOWLEDGMENTS

This work was partially supported by CNRS, under the *MathSTIC* project *Chaînes de Markov Cachées et Filtrage Particulaire*, and under the *AS-STIC* project *Méthodes Particulaires* (AS 67).

REFERENCES

- [1] Didier Dacunha-Castelle and Marie Duflo. *Probability and Statistics II*. Springer-Verlag, Berlin, 1986.
- [2] Pierre Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Probability and its Applications. Springer-Verlag, New York, 2004.
- [3] Pierre Del Moral, Jean Jacod, and Philip Protter. The Monte Carlo method for filtering with discrete-time observations. *Probability Theory and Related Fields*, 120(3):346–368, 2001.
- [4] Pierre Del Moral and Pascal Lezaud. Branching and interacting particle interpretation of rare event probabilities. In Henk Blom and John Lygeros, editors, *Stochastic Hybrid Systems: Theory and Applications*. Springer-Verlag, Berlin. To appear.
- [5] Pierre Del Moral and Laurent Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to nonlinear filtering. In Jacques Azéma, Michel Émery, Michel Ledoux, and Marc Yor, editors, *Séminaire de Probabilités XXXIV*, volume 1729 of *Lecture Notes in Mathematics*, pages 1–145. Springer-Verlag, Berlin, 2000.
- [6] Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer-Verlag, New York, 2001.
- [7] Allan Gut. *Stopped Random Walks. Limit Theorems and Applications*. Applied Probability (A Series of the Applied Probability Trust). Springer-Verlag, New York, 1988.
- [8] Hans R. Künsch. Recursive Monte Carlo filters : Algorithms and theoretical analysis. Research Report 112, Seminar für Statistik, ETH, Zürich, January 2003. <ftp://ftp.stat.math.ethz.ch/Research-Reports/112.ps.gz>.
- [9] François Le Gland and Nadia Oudjane. A robustification approach to stability and to uniform particle approximation of nonlinear filters : the example of pseudo-mixing signals. *Stochastic Processes and their Applications*, 106(2):279–316, August 2003.
- [10] François Le Gland and Nadia Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters. *The Annals of Applied Probability*, 14(1):144–187, February 2004.
- [11] Nadia Oudjane. *Stabilité et Approximations Particulaires en Filtrage Non-Linéaire — Application au Pistage*. Thèse de Doctorat, Université de Rennes 1, Rennes, December 2000. <ftp://ftp.irisa.fr/techreports/theses/2000/oudjane.pdf>.
- [12] Nadia Oudjane and Sylvain Rubenthaler. Stability and uniform particle approximation of nonlinear filters in case of non ergodic signal. *Stochastic Analysis and Applications*. To appear.
- [13] Alfréd Rényi. On the asymptotic distribution of the sum of a random number of independent random variables. *Acta Mathematica Academiae Scientiarum Hungaricae*, 8:193–199, 1957.
- [14] Vivien Rossi. *Filtrage Non-Linéaire par Noyaux de Convolution — Application à un Procédé de Dépollution Biologique*. Thèse de Doctorat, École Nationale Supérieure Agronomique de Montpellier, Montpellier, December 2004.
- [15] David Siegmund. *Sequential Analysis. Tests and Confidence Intervals*. Springer Series in Statistics. Springer-Verlag, New York, 1985.