PARALLELIZATION OF SEQUENTIAL MONTE CARLO METHODS USING PARTICLE ISLANDS: THE B^{2}ASIL ALGORITHM

C. Vergé[*], *E. Moulines*[†], *J. Olsson*[‡]

[*] ONERA — The French Aerospace Lab, Palaiseau, France

[†] Telecom ParisTech/ CNRS UMR LTCI, Paris, France

[‡] Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden

ABSTRACT

Particle island methods [1, 2] implement particle filters in a parallel architecture by dividing the particle system into islands evolving according to local particle filters. The islands are however allowed to interact on the basis of their mean potential. In this note, focus is set on the double bootstrap algorithm with adaptive selection on the island level (B^{2}ASIL), in which island interaction is applied only sparingly according to a criterion measuring the skewness of the island weights. Besides discussing, using novel results convergence obtained in [3], the theoretical properties of B^{2}ASIL, we also illustrate the performance of the algorithm by simulations.

Index Terms— SMC methods, island models, particle filter, central limit theorem (CLT), parallelization.

1. INTRODUCTION

This paper discusses approaches to parallelization of *sequential Monte Carlo methods* (or *particle filters*) approximating normalized *Feynman-Kac distribution flows*. Interacting particle systems are used increasingly for sampling from complex high-dimensional distributions in a wide range of applications, including nonlinear filtering, data assimilation problems, rare event sampling, hidden Markov chain parameter estimation, stochastic control problems, and financial mathematics; see e.g. [4, 5] and the references therein. These algorithms evolve, recursively and randomly in time, a sample of random draws, *particles*, with associated *importance weights*, and the Feynman-Kac distribution flow is approximated by the weighted empirical measures associated with this evolving sample. The particle cloud is updated through *selection* and *mutation* operations, where the former duplicates or eliminates, through resampling, particles with large or small importance weights, respectively, while the latter disseminates randomly the particles over the state space and updates accordingly the importance weights for further selection.

Particle filtering is a computationally intensive method. Parallel implementation is an appealing solution to tackle this issue but is not straightforward due to the particle interaction caused by the selection operation. In [1] it is proposed to implement the interacting particle system in parallel as follows. Instead of considering a single large batch of $N = N_1 N_2$ particles, the population is divided into N_1 batches of N_2 particles. The batches are in the sequel referred to as *islands*. Each particle belonging to a given island will also be referred to as an *individual*. Each island evolves according to the usual selection and mutation operations. The islands may evolve, on different processors, independently or may interact through multinomial selection according to weights proportional to the weight averages over the different subpopulations. We will consider two kinds of island interaction: systematic

[*] C. Vergé is supported by CNES (Centre National d'études Spatiales) and ONERA, The French Aerospace Lab.

[‡] J. Olsson is supported by the Swedish Research Council, Grant 2011-5577

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

resampling or adaptive resampling based on the *coefficient of variation* criterion, leading to the *double bootstrap algorithm* (B^2) and the *double bootstrap algorithm with adaptive selection on the island level* (B^{2}ASIL) algorithms, respectively, and where B^2 can be viewed as a particular case of the B^{2}ASIL algorithm. A sequence of Monte Carlo estimators (2) is obtained by weighing up, using the island weights, the self-normalized empirical measures associated with the different particle islands.

The theoretical analysis of B^{2}ASIL-type algorithms is challenging due to the intricate dependence structure imposed by the island selection operation and the "double asymptotics" introduced by N_1 and N_2, resp. island number and island size. Nevertheless, even though the islands are allowed to interact through selection, any two individuals of the system should become more and more statistically independent as the number of islands as well as the size of the islands grow (cf. the *propagation of chaos* property of standard SMC methods [6]). Thus, we may expect a law of large numbers as well as a CLT to hold when N_1 and N_2 tend *jointly* to infinity. Moreover, in analogy with similar result for standard, single batch SMC methods [see 7, 8, 9], we may expect the rate of such a CLT to be $\sqrt{N_1 N_2}$. This was recently established in [3], and we survey these results in Section 4.

2. FEYNMAN-KAC MODELS

For a sequence of unnormalized transition kernels $\{Q_n\}_{n\in\mathbb{N}}$ defined on some common measurable space $(\mathsf{X}, \mathcal{X})$ and some probability distribution $\eta_0 \in \mathsf{M}_1(\mathcal{X})$, a sequence $\{\eta_n\}_{n\in\mathbb{N}}$ of *Feynman-Kac measures* is defined, for all $h \in \mathsf{F}_\mathrm{b}(\mathcal{X})$ (here $\mathsf{M}_1(\mathcal{X})$ and $\mathsf{F}_\mathrm{b}(\mathcal{X})$ denote the spaces of probability measures and bounded measurable functions on $(\mathsf{X}, \mathcal{X})$, respectively) and $n \in \mathbb{N}$, by

$$\eta_n h = \frac{\int \cdots \int h(x_n)\, \eta_0(\mathrm{d}x_0) \prod_{p=0}^{n-1} Q_p(x_p, \mathrm{d}x_{p+1})}{\int \cdots \int \eta_0(\mathrm{d}x_0) \prod_{p=0}^{n-1} Q_p(x_p, \mathrm{d}x_{p+1})}. \tag{1}$$

We note that the Feynman-Kac measures satisfy the nonlinear recursion $\eta_{n+1} = \eta_n Q_n / \eta_n Q_n \mathbb{1}_\mathsf{X}$.

3. THE B^{2}ASIL ALGORITHM

For each $i \in [\![1, N_1]\!]$, we let $\{(\xi_N(i,j), \omega_N(i,j))\}_{j=1}^{N_2}$ be an island of particles or individuals (the ξ_Ns), with associated importance weights (the ω_ns). The particle weights are assumed to be positive and uniformly bounded. Each island is assigned a positive weight $\Omega_N(i)$, and the set of weighted islands, i.e., $\{(\Omega_N(i), \{(\xi_N(i,j), \omega_N(i,j))\}_{j=1}^{N_2})\}_{i=1}^{N_1}$, is called an *archipelago*. Set $\bar{\Omega}_N(i) \triangleq \Omega_N(i) / \sum_{\ell=1}^{N_1} \Omega_N(\ell)$ and $\bar{\omega}_N(i,j) \triangleq \omega_N(i,j) / \sum_{\ell=1}^{N_2} \omega_N(i,\ell)$.

In this section, our aim is to form, online, a sequence of archipelagos targeting the Feynman-Kac flow $\{\eta_n\}_{n\in\mathbb{N}}$ defined in (1) by subjecting the archipelagos to a number of elementary operations such as *selection on the island level*, *selection on the individual level* and *mutation*. For all $p \in \mathbb{N}$, let R_p be a (normalized) transition kernel on X such that $Q_p(x,\cdot) \ll R_p(x,\cdot)$ for all $x \in \mathsf{X}$, and denote the corresponding Radon-Nikodym derivatives by $w_p(x,\tilde{x}) \triangleq \mathrm{d}Q_p(x,\cdot)/\mathrm{d}R_p(x,\cdot)(\tilde{x})$, $(x,\tilde{x}) \in \mathsf{X}^2$. We will focus on the B^{2}ASIL *algorithm* introduced in [1, Algorithm 3], where selection on the island level is executed on the basis of the coefficient of variation (CV; see [10]) given by $\mathrm{CV}^2(\{\Omega_N(i)\}_{i=1}^{N_1}) = N_1 \sum_{i=1}^{N_1} \bar{\Omega}_N^2(i) - 1$. The CV is in one-to-one correspondence with the efficient sample size (ESS, proposed in [11] and used in [1]). The B^{2}ASIL scheme is detailed in Algorithm 1. Denote by, for $n \in \mathbb{N}$ and $h \in \mathsf{F}_\mathrm{b}(\mathcal{X})$,

$$\eta_n^N h = \sum_{i=1}^{N_1} \bar{\Omega}_N^{(n)}(i) \sum_{j=1}^{N_2} \bar{\omega}_N^{(n)}(i,j) h(\xi_N^{(n)}(i,j)) \tag{2}$$

the estimators returned by the B^{2}ASIL algorithm.

4. CONVERGENCE RESULTS FOR THE B^{2}ASIL ALGORITHM

In the following, let $N = N_1 N_2$ be the total number of individuals of each archipelago. The following results, describing the convergence of B^{2}ASIL as N_1 and N_2 tend jointly to infinity, were obtained in [3] through single-step analyses of the different selection and mutation operations acting on the archipelagos.

/* Initialization */
: **for** $i \leftarrow 1$ **to** N_1 **do**
  for $j \leftarrow 1$ **to** N_2 **do**
    $\xi_N^{(0)}(i,j) \sim \eta_0$;
    $\xi_N^{(1)}(i,j) \sim R_0(\xi_N^{(0)}(i,j), \cdot)$;
    $\omega_N^{(1)}(i,j) \leftarrow w_0(\xi_N^{(0)}(i,j), \xi_N^{(1)}(i,j))$;
  end
  $\Omega_N^{(1)}(i) \leftarrow \sum_{j=1}^{N_2} \omega_N^{(1)}(i,j)/N_2$;
end
for $p \leftarrow 1$ **to** $n-1$ **do**
  /* Island selection */
  if $\mathrm{CV}^2(\{\Omega_N^{(p)}(i)\}_{i=1}^{N_1}) > \tau$ **then**
    for $i \leftarrow 1$ **to** N_1 **do**
      $I_N(i) \sim \mathsf{Mult}(\{\bar{\Omega}_N^{(p)}(i')\}_{i'=1}^{N_1})$
    end
  else
    for $i \leftarrow 1$ **to** N_1 **do**
      $I_N(i) \leftarrow i$;
    end
  end
  for $i \leftarrow 1$ **to** N_1 **do**
    /* Individual selection */
    for $j \leftarrow 1$ **to** N_2 **do**
      $J_N(i,j) \sim \mathsf{Mult}(\{\bar{\omega}_N^{(p)}(I_N(i), j')\}_{j'=1}^{N_1})$
    end
    /* Mutation */
    for $j \leftarrow 1$ **to** N_2 **do**
      $\xi_N^{(p+1)}(i,j) \sim R_p(\xi_N^{(p)}(I_N(i), J_N(i,j)), \cdot)$;
      $\omega_N^{(p+1)}(i,j) \leftarrow w_p(\xi_N^{(p)}(I_N(i), J_N(i,j)), \xi_N^{(p+1)}(i,j))$;
    end
    $\Omega_N^{(p+1)}(i) \leftarrow \sum_{j=1}^{N_2} \omega_N^{(p+1)}(i,j)/N_2$;
  end
end

Algorithm 1: The B^{2}ASIL algorithm

Theorem 1 ([3]). *For all $n \in \mathbb{N}$ and $h \in \mathsf{F}_\mathrm{b}(\mathcal{X})$, $\eta_n^N h \xrightarrow{\mathbb{P}} \eta_n h$ as $N \to \infty$.*

Impose the following additional assumption.

(S) For all $\beta > 0$, $N_1 \exp(-\beta N_2) \to 0$ as $N \to \infty$,

Under **(S)**, which guarantees that N_1 grows only subexponentially fast with respect to N_2, one may derive the following CLT.

Theorem 2 ([3]). *Assume **(S)**. Then for all $n \in \mathbb{N}$, the random variable $\mathbb{1}\{\mathrm{CV}^2(\{\Omega_N^{(n)}(i)\}_{i=1}^{N_1}) > \tau\}$ has a deterministic limit ε_n in probability. Moreover, for all $h \in \mathsf{F}_\mathrm{b}(\mathcal{X})$, as $N \to \infty$,*

$$\sqrt{N}(\eta_n^N h - \eta_n h) \xrightarrow{\mathcal{D}} \mathsf{N}(0, V_n(h) + V_n^\varepsilon(h)),$$

where $V_0(h) = \eta_0\{(h - \eta_0 h)^2\}$, $V_0^\varepsilon = 0$, and

$$V_n h = \sum_{\ell=0}^{n-1} \frac{\eta_\ell R_\ell\{w_\ell^2 Q_{\ell+1} \cdots Q_{n-1}(h - \eta_n h)^2\}}{(\eta_\ell Q_\ell \cdots Q_{n-1} \mathbb{1}_\mathsf{X})^2},$$

$$V_n^\varepsilon h = \sum_{\ell=0}^{n-1} \sum_{p=\ell+1}^{n-1} \varepsilon_p \frac{\eta_\ell R_\ell\{w_\ell^2 Q_{\ell+1} \cdots Q_{n-1}(h - \eta_n h)^2\}}{(\eta_\ell Q_\ell \cdots Q_{n-1} \mathbb{1}_\mathsf{X})^2}.$$

Note that the B^2 algorithm, which selects systematically the islands, is a particular case of B^{2}ASIL algorithm for which $\tau = 0$ (and hence $\varepsilon_n = 1$) for all $n \in \mathbb{N}^*$. We may hence deduce the asymptotic variance $\sigma_n^2(h)$ of the B^2 algorithm: $\sum_{\ell=0}^{n-1}(n-\ell)\frac{\eta_\ell R_\ell\{w_\ell^2 Q_{\ell+1} \cdots Q_{n-1}(h - \eta_n h)^2\}}{(\eta_\ell Q_\ell \cdots Q_{n-1} \mathbb{1}_\mathsf{X})^2}$.

In this part we discuss conditions under which interaction on the island level is desirable. The case of independently evolving islands is a particular case of the B^{2}ASIL algorithm with $\tau = \infty$ (and hence $\varepsilon_n = 0$ for all $n \in \mathbb{N}$). We hence deduce, from Theorem 2, that the asymptotic variance of the estimator non-interactive bootstrap filters is, when N is large, V_n/N, i.e., inversely proportional to N. In addition, it was shown in [1] that the asymptotic bias of this estimator is B_n/N_2, i.e., inversely proportional to N_2 and where B_n is some number depending on the model, which is presumably defined in [1]. Consequently, island interaction implies an additional, positive variance term while decreasing

the bias, and a trade-off between bias and variance has to be made to decide when island interaction is beneficial. For this purpose, we compare the mean squared errors (MSE) when the islands interact and when they are kept independent. The MSE for independent islands is given by $V_n/N + B_n^2/N_2^2$, whereas the MSE of the B^{2}ASIL algorithm is given by $(V_n + V_n^\varepsilon)/N$. Therefore,

$$\frac{V_n+V_n^\varepsilon}{N} < \frac{V_n}{N} + \frac{B_n^2}{N_2^2} \quad \Leftrightarrow \quad N_2 < \frac{B_n^2}{V_n^\varepsilon} N_1.$$

Consequently, the B^{2}ASIL algorithm outperforms a bank of independently evolving bootstrap filters when the number of particles N_2 within each island is small compared to the number of islands N_1; the interaction reduces the bias (which is independent of N_1 when the islands are kept independent). On the contrary, when N_2 is larger than N_1, the variance increase introduced by the interaction on the island level may be larger than the bias reduction.

5. APPLICATION

We consider the stochastic volatility model $X_{p+1} = \alpha X_p + \sigma U_{p+1}, \quad Y_p = \beta \exp(X_p/2) V_p$, where $X_0 \sim \mathcal{N}(0, \sigma^2/(1-\alpha^2))$, $\{U_p\}_{p\in\mathbb{N}}$ and $\{V_p\}_{p\in\mathbb{N}}$ are independent sequences of mutually independent standard Gaussian random variables (independent of X_0). We based our simulations on $n = 100$ observations generated under the model dynamics $(\alpha, \sigma, \beta) = (0.98, 0.5, 1)$. We estimate the means of the predictive distributions $X_p \mid (Y_0 = y_0, \ldots, Y_{p-1} = y_{p-1})$, for $p \in [\![0, 100]\!]$, a problem that can be cast into the Feynman-Kac framework by setting, for all $p \in \mathbb{N}$,

$$Q_p(x_p, \mathrm{d}x_{p+1}) = \frac{1}{2\pi\sigma\beta} \exp\left(\frac{y_p^2 \exp(-x_p)}{2\beta^2} - \right.$$
$$\left. - \frac{x_p}{2} - \frac{(x_{p+1}-\alpha x_p)^2}{2\sigma^2} \right) \mathrm{d}x_{p+1}.$$

As a reference, we computed such predictive expectations using a single run of the bootstrap filter with 10^6 particles. Figure 1 displays box plots based on 250 replicates of the different algorithms for different values of N_1 and N_2. Table 1 reports the average numbers of island interactions for B^{2}ASIL over the 250 simulations. These numbers should be compared to $n = 100$ for B^2, since island interaction is systematic in this case. For a given number of islands, the number of island interactions decreases for B^{2}ASIL when the island size grows. Moreover, as soon as the number of particles N_2 of each island is large enough, B^{2}ASIL is no longer resampling the islands. Figure 2 displays CPU times for a $N_2 = 100$ and different numbers of islands. CPU time for the B^{2}ASIL is bigger than for independent bootstraps (due to the additional resampling step), and increases proportionally to the island number. Also notice that the parallel architecture of B^{2}ASIL reduces significantly CPU times compared to a classic bootstrap with $N_1 N_2$ particles.

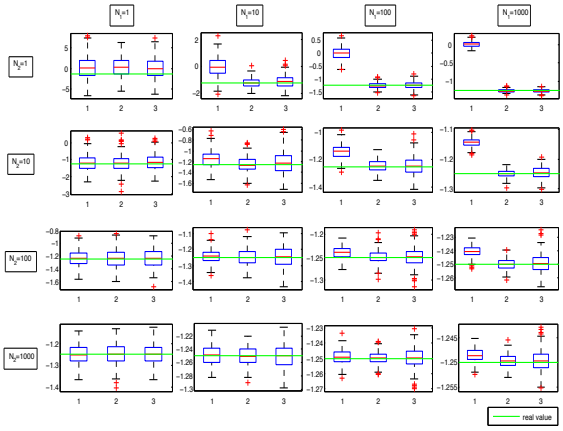

Fig. 1. Comparison of different interactions across the islands for the Stochastic volatility model (1) Independent bootstraps; (2) B^{2}ASIL; (3) B^2.

N_2 \ N_1	10	100	1000
1	30.1	35.14	36.108
10	10.9	12.29	12.096
100	1.5	1.86	1.956
1000	0	0	0

Table 1. Number of island interactions for the B^{2}ASIL algorithm.

6. STABILITY OF THE B^2 ALGORITHM

When studying the numerical stability of the B^2 algorithm we will work under the following *strong*

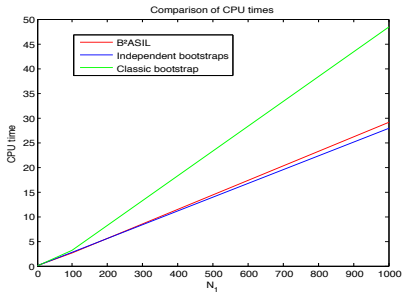

Fig. 2. Comparison of CPU times

mixing condition; see [3, Section 4.5] for considerably weaker assumptions.

(M) (i) There exist constants $0 < \sigma_- < \sigma_+ < \infty$ and $\varphi \in \mathsf{M}_1(\mathcal{X})$ such that for all $p \in \mathbb{N}$, $x \in \mathsf{X}$, and $\mathsf{A} \in \mathcal{X}$,

$$\sigma_- \varphi(\mathsf{A}) \leq M_p(x, \mathsf{A}) \leq \sigma_+ \varphi(\mathsf{A}).$$

(ii) $w_+ \triangleq \sup_{p \in \mathbb{N}} \|w_p\|_\infty < \infty$.

(iii) $c_- \triangleq \inf_{(p,x) \in \mathbb{N} \times \mathsf{X}} Q_p \mathbb{1}_{\mathsf{X}}(x) > 0$.

The assumption **(M)**(i) is rather restrictive and requires typically the state space X to be a compact set. Still, it plays a vital role in the literature of SMC analysis [6, 12]. On the other hand, the weaker assumption **(M)**(ii) is satisfied for most applications and **(M)**(iii) does not require the potential functions to be uniformly bounded from below; the latter is a condition that appears frequently in the literature. Under **(M)**, denote $\rho \triangleq 1 - \sigma_-/\sigma_+$; then we can derive the following explicit time uniform bound on the asymptotic variance. This shows that the B^2 algorithm stays numerically stable in the long-run.

Corollary 3. *Suppose (**M**). Then for all $n \in \mathbb{N}$ and $h \in \mathsf{F}_{\mathrm{b}}(\mathcal{X})$,*

$$\sigma_n^2(h) \leq w_+ \frac{\operatorname{osc}^2(h)}{(1-\rho)^2(1-\rho^2)^2 c_-}.$$

References

[1] C. Vergé, C. Dubarry, P. Del Moral, and E. Moulines, "On parallel implementation of sequential monte carlo methods: the island particle model," *Statistics and Computing*, vol. 23, 2013.

[2] Nick Whiteley, Anthony Lee, and Kari Heine, "On the role of interaction in sequential monte carlo algorithms," 2013, Preprint.

[3] C. Vergé, P. Del Moral, E. Moulines, and J. Olsson, "Convergence properties of weighted particle islands with application to the double bootstrap algorithm," Preprint, 2014.

[4] A. Doucet, N. De Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.

[5] O. Cappé and E. Moulines, "On the use of particle filtering for maximum likelihood parameter estimation," in *European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, September 2005.

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

[7] P. Del Moral and A. Guionnet, "Central limit theorem for nonlinear filtering and interacting particle systems," *Ann. Appl. Probab.*, vol. 9, no. 2, pp. 275–297, 1999.

[8] N. Chopin, "Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference," *Ann. Statist.*, vol. 32, no. 6, pp. 2385–2411, 2004.

[9] R. Douc and E. Moulines, "Limit theorems for weighted samples with applications to sequential Monte Carlo methods," *Ann. Statist.*, vol. 36, no. 5, pp. 2344–2376, 2008.

[10] A. Kong, J. S. Liu, and W. Wong, "Sequential imputation and Bayesian missing data problems," *J. Am. Statist. Assoc.*, vol. 89, no. 278-288, pp. 590–599, 1994.

[11] J. S. Liu, "Metropolized independent sampling with comparisons to rejection sampling and importance sampling," *Stat. Comput.*, vol. 6, pp. 113–119, 1996.

[12] R. Douc, E. Moulines, and D. Stoffer, *Nonlinear Time Series: Theory, Methods and Applications with R Examples*, Chapman.