Optimal potential functions for the interacting particle system method

Hassane Chraibi, Anne Dutfoy, Thomas Galtier and Josselin Garnier

Abstract. The assessment of the probability of a rare event with a naive Monte-Carlo method is computationally intensive, so faster estimation or variance reduction methods are needed. We focus on one of these methods which is the interacting particle system (IPS) method. The method is not intrusive in the sense that the random Markov system under consideration is simulated with its original distribution, but selection steps are introduced that favor trajectories (particles) with high potential values. An unbiased estimator with reduced variance can then be proposed. The method requires to specify a set of potential functions. The choice of these functions is crucial, because it determines the magnitude of the variance reduction. So far, little information was available on how to choose the potential functions. This paper provides the expressions of the optimal potential functions minimizing the asymptotic variance of the estimator of the IPS method and it proposes recommendations for the practical design of the potential functions.

Keywords. Rare event simulation; interacting particle system; Sequential Monte-Carlo Samplers; Feymann-Kac particle filters.

AMS classification. MSC 2010 : 65C05, 65C35, 60K35 65C40, 60J22.

1. Introduction

The simulation of rare events in Markov systems has been intensively studied. Standard Monte Carlo simulations are usually prohibitive because too many simulations are needed to achieve the rare events. The general approach to speeding up such simulations is to favor the occurrence of the rare events by using importance sampling (IS). In the IS strategy the system is simulated using a new set of input and/or transition probability distributions, and unbiased estimates are obtained by multiplying the simulation output by a likelihood ratio [24]. The tricky part in IS is to properly choose the biased distributions in order to get a significant variance reduction. Another difficulty is to run simulations with the biased dynamics, which is not always possible with big numerical codes in particular in industrial contexts. Other promising biasing Monte Carlo methods have been studied, such as sampling importance resampling (SIR), sequential Monte Carlo (SMC), and multicanonical Monte Carlo (MMC) [19, 14, 3]. All these methods involve biasing the distributions of some input variables to probe the tail of the pdf of the output variable of interest. The biased distributions are not assumed to be known a priori, but this knowledge is accumulated during the iterations. How-

ever, all these methods are intrusive. They require to modify the numerical codes or the experiments in the sense that the distributions of the input random variables have to be changed.

Non-intrusive techniques have been proposed, that do not require to modify the distributions of the input variables. The principle of the Interacting Particle System (IPS) method is to substitute for the bias of the input random variables selection steps based on the output variable. It was first introduced in [10], and with an alternative formulation in [11]. The algorithm simulates the Markov system with selection and mutation steps and it can give an unbiased estimator of the probability of the rare event [12]. The IPS methodology is closely related to a class of Monte Carlo acceptance/rejection simulation techniques introduced in the fifties and nowadays used in many applications [14].

The IPS algorithm generates a system of particles by a sequence of selection and mutation steps. The selection steps of the IPS algorithm are determined by potential functions. As we will see below, particles with low potential values are killed while those with high potential values are multiplied during the selection steps. When these potential functions are properly tuned, the IPS method can yield an estimator with a smaller variance than the Monte-Carlo estimator. The main purpose of the IPS method is to provide such a variance reduction without modifying the dynamics of the process. Indeed, as we will see below, the mutation steps of the IPS algorithm are carried out with the original Markov dynamics. If we can also modify the dynamics of the process, then it is theoretically possible to reach a zero variance and it is practically possible to design an iterative algorithm that approximates the optimal dynamics [17]. Here we want to consider the case where the dynamics of the process is left untouched, because it is not always possible to modify the dynamics of the system in complex numerical codes or in real experiments, as seen for instance in [15] where the numerical code is imposed. We consider such a framework in this paper (as we will discuss in the conclusion a particularly favorable context is the multifidelity Monte Carlo framework [23]). The tricky part in IPS is the choice of the potential functions that determine the selection steps [12, 15, 25, 18, 22]. In this paper, we consider the IPS method and we determine the optimal potential functions.

Let $(Z_k)_{0\leq k\leq n}$ be a Markov chain with values in the measurable spaces $(E_k, \mathcal{E}_k)$, with initial law ν_0, and with a kernel Q_k such that for $k > 0$ and for any bounded measurable function $h : E_k \to \mathbb{R}$

$$\mathbb{E}[h(Z_k)|Z_{k-1}] = \int_{E_k} h(z_k)Q_k(dz_k|Z_{k-1}). \tag{1.1}$$

For any bounded measurable function $h : E_0 \times \cdots \times E_n \to \mathbb{R}$ we have:

$$\mathbb{E}[h(Z_0, \ldots, Z_n)] = \int_{E_n\times\cdots\times E_0} h(z_0, \ldots, z_n)Q_n(dz_n|z_{n-1}) \cdots Q_1(dz_1|z_0)\nu_0(dz_0). \tag{1.2}$$

Let $\mathbf{Z}_k = (Z_0, Z_1, \ldots, Z_k)$ be a trajectory of size k, and let $\mathbf{E}_k = E_0 \times E_1 \times \cdots \times E_k$ be the set of trajectories of size k that we equip with the product σ-algebra $\boldsymbol{\mathcal{E}}_k = \mathcal{E}_0 \otimes \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_k$. For $i < j$, when it is necessary to differentiate the coordinates of the trajectories $\mathbf{z}_i$ and $\mathbf{z}_j$ we write the coordinates $z_{i,k}$ for $k \leq i$ and $z_{j,k}$ for $k \leq j$ such that $\mathbf{z}_i = (z_{i,0}, z_{i,1}, \ldots, z_{i,i})$ and $\mathbf{z}_j = (z_{j,0}, z_{j,1}, \ldots, z_{j,j})$. We write $\mathbf{z}_k = (z_0, \ldots, z_k)$ when there is no ambiguity. We introduce the Markov chain of the trajectories $(\mathbf{Z}_k)_{0 \leq k \leq n}$ with values in the measurable spaces $(\mathbf{E}_k, \boldsymbol{\mathcal{E}}_k)$, and with the transition kernels M_k such that:

$$M_k(d\mathbf{z}_k | \mathbf{z}_{k-1}) = \delta_{\mathbf{z}_{k-1}}\big(d(z_{k,0}, \ldots, z_{k,k-1})\big) Q_k(dz_{k,k} | z_{k-1,k-1}). \tag{1.3}$$

For any bounded measurable function $h : \mathbf{E}_n \to \mathbb{R}$ we have

$$\begin{aligned} p_h &= \mathbb{E}\left[h(\mathbf{Z}_n)\right] \\ &= \int_{\mathbf{E}_n \times \cdots \times \mathbf{E}_0} h(\mathbf{z}_n) \prod_{k=1}^{n} M_k(d\mathbf{z}_k | \mathbf{z}_{k-1}) \nu_0(d\mathbf{z}_0). \end{aligned} \tag{1.4}$$

Rare event analyses are often carried out for reliability assessment where the system under consideration is modeled by the Markov chain of the trajectories $(\mathbf{Z}_k)_{0 \leq k \leq n}$ and we want to assess the probability of failure of the system. The system failure is charcaterized by a region D of $\mathbf{E}_n$ and we want to assess the probability of failure which happens when $\mathbf{Z}_n$ enters D. Typically the system is reliable, which means that probability of failure $\mathbb{P}(\mathbf{Z}_n \in D)$ is small. In order to assess the probability of failure, we can take $h(\mathbf{z}_n) = \mathbb{1}_D(\mathbf{z}_n)$ so that $p_h = \mathbb{P}(\mathbf{Z}_n \in D)$ and use the IPS method to get an estimation $\hat{p}_h$ of p_h. Although the main application of the IPS method is reliability assessment where it relates to the case $h = \mathbb{1}_D$, the result of this paper will be presented in its most general form, where h is an arbitrary bounded measurable function.

The IPS method provides an estimator $\hat{p}_h$ of p_h. As we have said, the choice of the potential functions $G_k : \mathbf{E}_k \to [0, +\infty)$ for $k \leq n-1$ used in the selection steps is critical because it determines the variance of the IPS estimator $\hat{p}_h$. So far, little information has been provided on the forms of efficient potential functions. The standard approach is to find the best potential functions within a set of parametric potential functions. As a consequence the efficiency of the method strongly depends on the quality of the chosen parametric family. For instance, in [12] (where $E_s = \mathbb{R}$) the authors obtain their best variance reduction by choosing (remember that we denote $\mathbf{z}_k = (z_0, \ldots, z_k)$)

$$G_k(\mathbf{z}_k) = \frac{\exp\left[-\lambda V(z_k)\right]}{\exp\left[-\lambda V(z_{k-1})\right]},$$

where λ is a positive tuning parameter, and the quantity $V(z) = a - z$ roughly measures the proximity of z to the critical region that was of the form $D = [a, +\infty)$. In

[25] is is shown that it seems better to take a time-dependent proximity function V_k instead of V, yielding:

$$G_k(\mathbf{z}_k) = \frac{\exp\left[-\lambda V_k(z_k)\right]}{\exp\left[-\lambda V_{k-1}(z_{k-1})\right]},$$

where the quantities $V_k(z)$ are again measuring the proximity of z to D. Once the set of parametric potential functions is chosen, it is necessary to optimize the tuning parameters of the potentials. Different methods have been proposed. In [18], an empirical heuristic algorithm is provided; in [22] a meta model of the variance is minimized; in [12] a large deviation principle is used as a guide. One other common option for the potential functions is the one done in splitting methods. Indeed the splitting method can also be seen as a version of the IPS method [6]. In this method one wants to assess the probability that a random variable Z belongs to a subset B_n. A succession of nested sets $E = B_0 \supseteq B_1 \supseteq B_2 \supseteq \cdots \supseteq B_n$ is chosen by the practitioner or possibly chosen in an adaptive manner [5]. One considers a sequence of random variables $(Z_i)_{i=1,\dots,n}$ such that the small probability $\mathbb{P}(Z \in B_n)$ can be decomposed into a product of conditional probabilities: $\mathbb{P}(Z \in B_n) = \prod_{i=1}^{n} \mathbb{P}(Z_i \in B_i | Z_{i-1} \in B_{i-1})$ and $\mathbb{P}(Z \in B_n) = \mathbb{E}\left[h(\mathbf{Z}_n)\right]$ by setting $h(\mathbf{z}_n) = \mathbb{1}_{B_n}(z_n)$. In this method the potential functions are chosen of the following form

$$G_k(\mathbf{z}_k) = \mathbb{1}_{B_k}(z_k).$$

One usually optimizes the variance reduction within this family of potential functions by optimizing the choice of the sets $(B_k)_{k \leq n}$.

In this paper we tackle the issue of the choice of the potential functions. Our contribution is to provide the expressions of the optimal potential functions that minimize the asymptotic variance of the IPS estimator. By showing there is a positive minimum to the variance reduction we confirm the empirical result reported in the literature: the maximal variance reduction reachable with the IPS method is bounded from below. These expressions will hopefully lead the practitioners to design more efficient potential functions, that are close to the optimal ones. We give a few comments and recommendations based on the expressions of the optimal potential functions to guide the user in the choice of the potential functions.

The rest of the paper is organized as follows. Section 2 introduces the IPS method. Section 3 presents the potential functions that minimize the asymptotic variance of the IPS estimator. Section 4 presents three examples of applications. Finally Section 5 discusses the implications of our results.

In the rest of the paper we use the following notations: We denote by $\mathcal{M}(A)$ the set of bounded measurable functions on a set A. If f is a bounded measurable function, and η is a measure we note $\eta(f) = \int f \, d\eta$. If M is a Markovian kernel, we denote

by $M(f)$ the function such that $M(f)(x) = \int f(y)M(dy|x)$, and for a measure η, we denote by ηM the measure such that

$$\eta M(f) = \int \int f(y)M(dy|x)\eta(dx). \tag{1.5}$$

2. The IPS method

2.1. A Feynman-Kac model

The IPS method relies on a Feynman-Kac model [10] which is defined in this subsection. Let $(G_s)_{0\leq s<n}$ be a set of potential functions $G_s : \mathbf{E}_s \to [0, +\infty)$ such that $\mathbb{E}[\prod_{s=0}^{n-1} G_s(\mathbf{Z}_s)] > 0$. For each $k < n$, we define a target probability measure $\tilde{\eta}_k$ on $(\mathbf{E}_k, \boldsymbol{\mathcal{E}}_k)$, such that $\forall B \in \boldsymbol{\mathcal{E}}_k$:

$$\tilde{\eta}_k(B) = \frac{\mathbb{E}\left[\mathbb{1}_B(\mathbf{Z}_k)\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right]}{\mathbb{E}\left[\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right]}. \tag{2.1}$$

For each $0 \leq k \leq n-1$, we define the propagated target probability measure η_{k+1} on $(\mathbf{E}_{k+1}, \boldsymbol{\mathcal{E}}_{k+1})$ such that $\eta_{k+1} = \tilde{\eta}_k M_k$ and $\eta_0 = \tilde{\eta}_0$. We have $\forall B \in \boldsymbol{\mathcal{E}}_{k+1}$:

$$\eta_{k+1}(B) = \frac{\mathbb{E}\left[\mathbb{1}_B(\mathbf{Z}_{k+1})\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right]}{\mathbb{E}\left[\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right]}. \tag{2.2}$$

Let Ψ_k be the application that transforms a measure η defined on $\mathbf{E}_k$ into a measure $\Psi_k(\eta)$ defined on $\mathbf{E}_k$ and such that

$$\Psi_k(\eta)(f) = \frac{\int G_k(\mathbf{z})f(\mathbf{z})\eta(d\mathbf{z})}{\eta(G_k)}. \tag{2.3}$$

We say that $\Psi_k(\eta)$ gives the selection of η through the potential G_k. Notice that $\tilde{\eta}_k$ is the selection of η_k as $\tilde{\eta}_k = \Psi_k(\eta_k)$. The target distributions can therefore be built according to the following successive selection and propagation steps:

$$\eta_k \xrightarrow{\Psi_k} \tilde{\eta}_k \xrightarrow{.M_k} \eta_{k+1}.$$

We also define the associated unnormalized measures $\tilde{\gamma}_k$ and γ_{k+1}, such that for $f \in \mathcal{M}(\mathbf{E}_k)$:

$$\begin{aligned} \tilde{\gamma}_k(f) &= \mathbb{E}\left[f(\mathbf{Z}_k)\prod_{s=0}^{k} G_s(\mathbf{Z}_s)\right], \\ \tilde{\eta}_k(f) &= \frac{\tilde{\gamma}_k(f)}{\tilde{\gamma}_k(1)}, \end{aligned} \tag{2.4}$$

and for $f \in \mathcal{M}(\mathbf{E}_{k+1})$:

$$\begin{aligned} \gamma_{k+1}(f) &= \mathbb{E}\left[f(\mathbf{Z}_{k+1}) \prod_{s=0}^{k} G_s(\mathbf{Z}_s) \right], \\ \eta_{k+1}(f) &= \frac{\gamma_{k+1}(f)}{\gamma_{k+1}(1)}. \end{aligned} \tag{2.5}$$

Denoting $f_h(\mathbf{z}) = \frac{h(\mathbf{z})}{\prod_{s=0}^{n-1} G_s(\mathbf{z}_s)}$ if $\prod_{s=0}^{n-1} G_s(\mathbf{z}_s) > 0$ and $f_h(\mathbf{z}) = 0$ otherwise, we have:

$$p_h = \gamma_n(f_h) = \eta_n(f_h) \prod_{k=0}^{n-1} \eta_k\big(G_k\big). \tag{2.6}$$

2.2. The IPS algorithm and its estimator

The IPS method provides an algorithm to generate samples whose weighted empirical measures approximate the probability measures η_k and $\tilde{\eta}_k$ for each step k. These approximations are then used to provide an estimator of p_h. For the sample approximating η_k, we denote $\mathbf{Z}_k^j$ the j^{th} trajectory and W_k^j its weight. Similarly, in the sample approximating $\tilde{\eta}_k$, we denote $\tilde{\mathbf{Z}}_k^j$ the j^{th} trajectory and $\tilde{W}_k^j$ its associated weight. For simplicity reasons, in this paper, we consider that the samples all contain N trajectories, but it is possible to modify the sample size at each step, as illustrated in [21]. The empirical measure approximating η_k and $\tilde{\eta}_k$ are denoted by η_k^N and $\tilde{\eta}_k^N$ and are defined by:

$$\tilde{\eta}_k^N = \sum_{i=1}^{N} \tilde{W}_k^i \, \delta_{\tilde{\mathbf{Z}}_k^i} \quad \text{and} \quad \eta_k^N = \sum_{i=1}^{N} W_k^i \, \delta_{\mathbf{Z}_k^i} \, . \tag{2.7}$$

So for all $f \in \mathcal{M}(\mathbf{E}_k)$,

$$\tilde{\eta}_k^N(f) = \sum_{i=1}^{N} \tilde{W}_k^i \, f\big(\tilde{\mathbf{Z}}_k^i\big) \quad \text{and} \quad \eta_k^N(f) = \sum_{i=1}^{N} W_k^i \, f\big(\mathbf{Z}_k^i\big) \, . \tag{2.8}$$

By plugging these estimators into equations (2.4) and (2.5), we get estimators for the unnormalized mesures. Denoting by $\tilde{\gamma}_k^N$ and γ_k^N these estimators, for all $f \in \mathcal{M}(\mathbf{E}_k)$, we have:

$$\tilde{\gamma}_k^N(f) = \tilde{\eta}_k^N(f) \prod_{s=0}^{k-1} \eta_s^N(G_s) \tag{2.9}$$

and

$$\gamma_k^N(f) = \eta_k^N(f) \prod_{s=0}^{k-1} \eta_s^N(G_s). \tag{2.10}$$

Initialization:

$k = 0,\ \forall j = 1, \dots, N,\ \mathbf{Z}_0^j \overset{i.i.d.}{\sim} \eta_0,$

$W_0^j = \frac{1}{N}$, and $\tilde{W}_0^j = \frac{G_0(\mathbf{Z}_0^j)}{\sum_{s=1}^N G_0(\mathbf{Z}_0^s)}$

while $k < n$ **do**

Selection:

$(\tilde{N}_k^j)_{j=1,\dots,N} \sim Mult\big(N, (\tilde{W}_k^j)_{j=1,\dots,N}\big)$

$\forall j = 1, \dots, N,\ \tilde{W}_k^j = \frac{1}{N}$

Propagation:

for $j := 1, \dots, N$ **do**

using the kernel M_k, continue the trajectory $\tilde{\mathbf{Z}}_k^j$ to get $\mathbf{Z}_{k+1}^j$

set $W_{k+1}^j = \tilde{W}_k^j$ and $\tilde{W}_{k+1}^j = \frac{W_{k+1}^j G_{k+1}(\mathbf{Z}_{k+1}^j)}{\sum_s W_{k+1}^s G_{k+1}(\mathbf{Z}_{k+1}^s)}$

if $\forall j,\ \tilde{W}_{k+1}^j = 0$ **then**

$\forall q > k$, set $\eta_q^N = \tilde{\eta}_q^N = 0$ and Stop

else

$k = k + 1$

Table 1. IPS algorithm.

In particular if we apply (2.10) to the test function f_h, we get an estimator $\hat{p}_h$ of p_h defined by:

$$\hat{p}_h = \eta_n^N(f_h) \prod_{k=0}^{n-1} \eta_k^N\big(G_k\big). \tag{2.11}$$

The IPS algorithm builds the samples sequentially, alternating between a selection step and a propagation step. The k^{th} selection step transforms the sample $(\mathbf{Z}_k^j, W_k^j)_{j \le N}$, into the sample $(\tilde{\mathbf{Z}}_k^j, \tilde{W}_k^j)_{j \le N}$. This transformation is done with a multinomial resampling scheme. This means that the $\tilde{\mathbf{Z}}_k^j$'s are drawn with replacement from the sample $(\mathbf{Z}_k^j)_{j \le N}$, each trajectory $\mathbf{Z}_k^j$ having a probability $\frac{W_k^j G_k(\mathbf{Z}_k^j)}{\sum_{i=1}^N W_k^i G_k(\mathbf{Z}_k^i)}$ to be drawn each time. We let $\tilde{N}_k^j$ be the number of times the particle $\mathbf{Z}_k^j$ is replicated in the sample $(\tilde{\mathbf{Z}}_k^j, \tilde{W}_k^j)_j$, so $N = \sum_{j=1}^N \tilde{N}_k^j$. After this resampling the weights $\tilde{W}_k^j$ are set to $1/N$. The interest of this selection by resampling is that it discards low potential trajectories and replicates high potential trajectories. Thus, the selected sample focuses on trajectories that will have a greater impact on the estimations of the next distributions once extended.

If one specifies potential functions that are not positive everywhere, there can be a possibility that at a step k we get $\forall j, G_k(\mathbf{Z}_k^j) = 0$. When this is the case, the probability

for resampling cannot be defined, the algorithm stops, and we consider that $\forall s \geq k$ the measures $\tilde{\eta}_s^N$ and η_{s+1}^N are equal to the null measure.

Then the k^{th} propagation step transforms the sample $(\tilde{\mathbf{Z}}_k^j, \tilde{W}_k^j)_{j\leq N}$, into the sample $(\mathbf{Z}_{k+1}^j, W_{k+1}^j)_{j\leq N}$. Each trajectory $\mathbf{Z}_{k+1}^j$ is obtained by extending the trajectory $\tilde{\mathbf{Z}}_k^j$ one step further using the transition kernel M_k. The weights satisfy $W_{k+1}^j = \tilde{W}_k^j$, $\forall j$. Then the procedure is iterated until the step n. The full algorithm to build the samples is displayed in table 1.

For $k < n$, we denote by $\hat{\mathbf{E}}_k = \{\mathbf{z}_k \in \mathbf{E}_k, G_k(\mathbf{z}_k) > 0\}$ the support of G_k, and we denote $\hat{\mathbf{E}}_n = \{\mathbf{z}_n \in \mathbf{E}_n, h(\mathbf{z}_n) > 0\}$ the support of h. We will make the following assumption on the potential functions:

$$\exists\, \varepsilon > 0, \quad \forall k \leq n,\ \forall \mathbf{z}_{k-1} \in \hat{\mathbf{E}}_{k-1},\ M_{k-1}(\hat{\mathbf{E}}_k | \mathbf{z}_{k-1}) > \varepsilon. \tag{G}$$

Theorem 2.1. *When the potential functions satisfy* (G)*, $\hat{p}_h$ is unbiased and strongly consistent.*

The proof of theorem 2.1 can be found in [10] chapter 7.

Theorem 2.2. *When the potential functions are positive-valued, we have*

$$\sqrt{N}(\hat{p}_h - p_h) \xrightarrow[N\to\infty]{d} \mathcal{N}\big(0, \sigma^2_{IPS,G}\big), \tag{2.12}$$

where

$$\sigma^2_{IPS,G} = \sum_{k=0}^{n} \left\{ \mathbb{E}\left[\prod_{i=0}^{k-1} G_i(\mathbf{Z}_i) \right] \mathbb{E}\left[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s) \right] - p_h^2 \right\}, \tag{2.13}$$

with the convention $\prod_{i=0}^{-1} G_i(\mathbf{Z}_i) = \prod_{i=0}^{-1} G_i^{-1}(\mathbf{Z}_i) = 1$.

A proof of this Central Limit Theorem (CLT) can be found in [10] chapter 9 or [12]. This CLT is an important result of the particle filter literature. The non-asymptotic fluctuations of the particle filters have been studied in [4]. Recently two weakly consistent estimators of the asymptotic variance $\sigma^2_{IPS,G}$, that are based on a single run of the method, have been proposed in [21]. One of these estimators is closely related to the one that was proposed in [7].

3. The optimal potential functions

Here we aim at estimating p_h. The choice of the potential functions has an impact on the variance of the estimation, so we would like to find potential functions that

minimize the asymptotic variance (2.13). This choice depends on the function h. Also, note that if potential functions G_k and G'_k are such that $G_k = a_k G'_k$ with $a_k > 0$, then they yield the same variance: $\sigma^2_{IPS,G} = \sigma^2_{IPS,G'}$. If potential functions G_k and G'_k are such that $G_k(\mathbf{Z}_k) = G'_k(\mathbf{Z}_k)$ a.s., then they also yield the same variance. Therefore all potential functions will be defined up to a multiplicative constant ν_k a.s., where ν_k is the distribution of $\mathbf{Z}_k$.

Theorem 3.1. *For $k = 0, \dots, n-1$, let $g_k : \mathbf{E}_k \to [0, +\infty)$ be defined by:*

$$g_k(\mathbf{z}_k) = \mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_{k+1}\big]^2\big|\mathbf{Z}_k = \mathbf{z}_k\Big]. \tag{3.1}$$

For $k = 0, \dots, n-1$, let $G^*_k : \mathbf{E}_k \to [0, +\infty)$ be defined by:

$$G^*_k(\mathbf{z}_k) = \sqrt{\frac{g_k(\mathbf{z}_k)}{g_{k-1}(\mathbf{z}_{k-1})}} \tag{3.2}$$

with $\mathbf{z}_{k-1} = (z_{k,0}, \dots, z_{k,k-1})$, if $g_{k-1}(\mathbf{z}_{k-1}) > 0$, and $G^*_k(\mathbf{z}_k) = 0$ otherwise. We use the convention $g_{-1} = 1$.

The potential functions minimizing $\sigma^2_{IPS,G}$ are the G^*_k's, $k \leq n-1$. They are unique in the sense that they are defined up to a multiplicative constant ν_k a.s..

The optimal variance of the IPS method with n steps is then

$$\begin{aligned}\sigma^2_{IPS,G^*} =& \mathbb{E}\left[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_0\big]^2\right] - p^2_h \\ &+ \sum_{k=1}^{n}\left\{\mathbb{E}\left[\sqrt{\mathbb{E}\Big[\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2\big|\mathbf{Z}_{k-1}\Big]}\right]^2 - p^2_h\right\}.\end{aligned} \tag{3.3}$$

Proof. As we lack mathematical tools to minimize $\sigma^2_{IPS,G}$ over the set of potential functions $(G_k)_{k<n}$, the strategy is the following one: First we guess the expressions (3.2) by a heuristic (and greedy) argument in the spirit of the one-step-ahead minimization method used in [13, 16] for instance. Second we provide the proof that (3.2) indeed minimizes the asymptotic variance of the IPS estimator.

Assuming that we already know the $k-2$ first potential functions, we start by trying to find the $k-1$-th potential function G_{k-1} that minimizes the k-th term of the sum in (2.13). This is equivalent to minimize the quantity

$$\mathbb{E}\Big[\prod_{i=0}^{k-1} G_i(\mathbf{Z}_i)\Big]\mathbb{E}\Big[\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2\big|\mathbf{Z}_{k-1}\big]\prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \tag{3.4}$$

over G_{k-1}. As the G_{k-1} are equivalent up to a multiplicative constant, we simplify the equation by choosing a multiplicative constant so that $\mathbb{E}\Big[\prod_{i=0}^{k-1} G_i(\mathbf{Z}_i)\Big] = 1$.

Our minimizing problem then becomes the minimization of (3.4) under the constraint $\mathbb{E}\left[\prod_{i=0}^{k-1} G_i(\mathbf{Z}_i)\right] = 1$. In order to be able to use a Lagrangian minimization we temporarily assume that the distribution of $\mathbf{Z}_{k-1}$ is discrete and that $\mathbf{Z}_{k-1}$ takes its values in a finite or numerable set $\mathbf{E}_{k-1}$. For $\mathbf{z} \in \mathbf{E}_{k-1}$, we denote $a_{\mathbf{z}} = \mathbb{P}(\mathbf{Z}_{k-1} = \mathbf{z})$, $g_{\mathbf{z}} = \mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2\big|\mathbf{Z}_{k-1} = \mathbf{z}\big]$ and $d_{\mathbf{z}} = \prod_{i=0}^{k-2} G_i(\mathbf{z}_i)G_{k-1}(\mathbf{z})$ (with $\mathbf{z}_i = (z_0, \ldots, z_i)$). The minimization problem becomes the minimization of

$$\mathcal{L} = \left(\sum_{\mathbf{z}\in\mathbf{E}_{k-1}} \frac{p_{\mathbf{z}} g_{\mathbf{z}}}{d_{\mathbf{z}}}\right) - \lambda\left(1 - \sum_{\mathbf{z}\in\mathbf{E}_{k-1}} p_{\mathbf{z}} d_{\mathbf{z}}\right). \tag{3.5}$$

Finding the minimum of this Lagrangian we get that $d_{\mathbf{z}} = \frac{\sqrt{g_{\mathbf{z}}}}{\sum_{\mathbf{z}'\in\mathbf{E}_{k-1}} p_{\mathbf{z}'}\sqrt{g_{\mathbf{z}'}}}$. Now relaxing the constraint of the multiplicative constant, we get that

$$\prod_{i=0}^{k-1} G_i(\mathbf{z}_i) \propto \sqrt{\mathbb{E}\big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2\big|\mathbf{Z}_{k-1} = \mathbf{z}\big]} = \sqrt{g_{k-1}(\mathbf{z})},$$

which gives the desired expressions. After these heuristic arguments we can now rigorously check that these expressions, obtained by minimizing each of the term of sum in (2.13) one by one, also minimize the whole sum for any distribution of the $\mathbf{Z}_{k-1}$'s.

The proof now consists in showing that, for any set of positive-valued potential functions (G_s), we have $\sigma^2_{IPS,G} \geq \sigma^2_{IPS,G^*}$. This is done by bounding from below each term of the sum in (2.13).

We denote by $supp\prod_{s=0}^{k-1} G_s^*$, i.e., the support of the function $\mathbf{z}_{k-1} \mapsto \prod_{s=0}^{k-1} G_s^*(\mathbf{z}_s)$, with the convention $\mathbf{z}_s = (z_{k-1,0}, \ldots, z_{k-1,s})$ in the product. We start by decomposing a product of potential functions as follows: $\forall k \in \{1, \ldots, n\}$,

$$\prod_{s=0}^{k-1} G_s(\mathbf{z}_s) = \epsilon_{k-1}(\mathbf{z}_{k-1}) \prod_{s=0}^{k-1} G_s^*(\mathbf{z}_s) + \bar{\epsilon}_{k-1}(\mathbf{z}_{k-1}) \tag{3.6}$$

where

$$\epsilon_{k-1}(\mathbf{z}_{k-1}) = \frac{\prod_{s=0}^{k-1} G_s(\mathbf{z}_s)}{\prod_{s=0}^{k-1} G_s^*(\mathbf{z}_s)} \text{ and } \bar{\epsilon}_{k-1}(\mathbf{z}_{k-1}) = 0$$

if $\mathbf{z}_{k-1} \in supp\prod_{s=0}^{k-1} G_s^*$, and

$$\epsilon_{k-1}(\mathbf{z}_{k-1}) = 0 \text{ and } \bar{\epsilon}_{k-1}(\mathbf{z}_{k-1}) = \prod_{s=0}^{k-1} G_s(\mathbf{z}_s),$$

if $\mathbf{z}_{k-1} \notin supp \prod_{s=0}^{k-1} G_s^*$. Using (3.6) we get that

$$\begin{aligned}
&\mathbb{E}\Big[\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)\Big]\mathbb{E}\Big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\
&= \mathbb{E}\Big[\epsilon_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\Big]\mathbb{E}\Big[g_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\
&\quad + \mathbb{E}\Big[\bar{\epsilon}_{k-1}(\mathbf{Z}_{k-1})\Big]\mathbb{E}\Big[g_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\
&\geq \mathbb{E}\Big[\epsilon_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\Big]\mathbb{E}\left[\frac{g_{k-1}(\mathbf{Z}_{k-1})}{\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)}\right] + 0. \qquad (3.7)
\end{aligned}$$

For $\mathbf{z}_{k-1} \in supp \prod_{s=0}^{k-1} G_s^*$ we have:

$$\prod_{s=0}^{k-1} G_s^*(\mathbf{z}_s) \propto \sqrt{g_{k-1}(\mathbf{z}_{k-1})}.$$

Consequently $supp\, g_{k-1} = supp \prod_{s=0}^{k-1} G_s^*$ and we get

$$\begin{aligned}
\mathbb{E}\left[\frac{g_{k-1}(\mathbf{Z}_{k-1})}{\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)}\right] &= \mathbb{E}\left[\frac{g_{k-1}(\mathbf{Z}_{k-1})}{\epsilon_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)}\right] \\
&= \mathbb{E}\left[\frac{1}{\epsilon_{k-1}(\mathbf{Z}_{k-1})} \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\right]. \qquad (3.8)
\end{aligned}$$

Combining (3.8) with inequality (3.7) we obtain that

$$\begin{aligned}
&\mathbb{E}\Big[\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)\Big]\mathbb{E}\Big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\
&\geq \mathbb{E}\Big[\epsilon_{k-1}(\mathbf{Z}_{k-1}) \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\Big]\mathbb{E}\left[\frac{1}{\epsilon_{k-1}(\mathbf{Z}_{k-1})} \prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\right]
\end{aligned}$$

and using the Cauchy-Schwarz inequality on the right term, we get that

$$\begin{aligned}
&\mathbb{E}\Big[\prod_{s=0}^{k-1} G_s(\mathbf{Z}_s)\Big]\mathbb{E}\Big[\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k]^2 \prod_{s=0}^{k-1} G_s^{-1}(\mathbf{Z}_s)\Big] \\
&\geq \mathbb{E}\Big[\prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\Big]^2 = \mathbb{E}\Big[\prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)\Big]\mathbb{E}\left[\frac{g_{k-1}(\mathbf{Z}_{k-1})}{\prod_{s=0}^{k-1} G_s^*(\mathbf{Z}_s)}\right]. \qquad (3.9)
\end{aligned}$$

By summing the inequalities (3.9) with respect to k, we easily see that

$$\sigma^2_{IPS,G} \geq \sigma^2_{IPS,G^*}.$$

The equality is achieved when $\epsilon_{k-1}(\mathbf{Z}_{k-1})$ is constant ν_k-a.s.. This completes the proof of the theorem. □

If the optimal potential functions G_k^* are positive-valued (i.e., if the functions g_k defined by (3.1) are positive valued), then they satisfy the hypothesis under which the TCL was proven and they can be used as potential functions for the IPS estimation method and they give an estimator with the minimal asymptotic variance (3.3). This is the case for all the examples addressed in the next section. If the optimal potential functions G_k^* are not positive everywhere, then they violate the hypothesis under which the TCL was proven in [10]. One could revisit the proof to extend its validity, in order to get a central limit theorem with weaker conditions [8, 20]. Here we simply state that, if the optimal potential functions G_k^* are not positive everywhere, then Theorem 3.1 can be understood in the following way: Eq. (3.3) gives the infimum value of the asymptotic variance of the IPS estimator amongst all acceptable potential functions. This value can be reached by a minimizing sequence of potential functions that are of the form $G_k = \max(G_k^*, \varepsilon)$, $k \leq n-1$, for $\varepsilon \searrow 0$.

4. Numerical illustrations

In this section we apply the IPS method to toy models. The first model under consideration in section 4.1 is a Gaussian random walk, that has been studied several times in the literature [12, 18]. In section 4.2 we consider an Ornstein-Uhlenbeck process. In sections 4.3-4.4 we consider a more complex diffusion process.

4.1. First example

We consider the Gaussian random walk $Z_{k+1} = Z_k + \varepsilon_{k+1}$, $Z_0 = 0$, where the ε_k are i.i.d. Gaussian random variables with mean zero and variance one. The goal is to compute the probability that Z_n exceeds a large positive value a. Therefore we take $h(\mathbf{z}_n) = \mathbb{1}_{[a,+\infty)}(z_n)$ so that

$$p_h = \mathbb{P}(Z_n \geq a). \tag{4.1}$$

In this case $\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k = \mathbf{z}_k]$ can be expressed in terms of the cdf Φ of the standard normal distribution:

$$\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k = \mathbf{z}_k] = \Phi\Big(\frac{z_k - a}{\sqrt{n-k}}\Big).$$

The optimal potential functions satisfy

$$\prod_{i=0}^{k-1} G_i^*(\mathbf{z}_i)^2 = \mathbb{E}\Big[\,\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2\big|\mathbf{Z}_{k-1}=\mathbf{z}_{k-1}\Big] = \Psi\big(z_{k-1}-a;n-k\big),$$

where

$$\Psi(z;\rho)=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\Phi\Big(\frac{z-z'}{\sqrt{\rho}}\Big)^2\exp\big(-\frac{z'^2}{2}\big)dz'. \tag{4.2}$$

Therefore, the optimal potential functions are given by

$$G_0^*(\mathbf{z}_0)=\Psi\big(z_0-a;n-1\big)^{1/2}, \tag{4.3}$$

and for $1\leq k\leq n-1$

$$G_k^*(\mathbf{z}_k)=\frac{\Psi\big(z_k-a;n-k-1\big)^{1/2}}{\Psi\big(z_{k-1}-a;n-k\big)^{1/2}}. \tag{4.4}$$

The Chernov-Bernstein's inequality gives the rather sharp exponential bound:

$$\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k=\mathbf{z}_k]\leq\exp\left[-\frac{(a-z_k)_+^2}{2(n-k)}\right],$$

from which we can deduce that:

$$\prod_{i=0}^{k-1} G_i^*(\mathbf{z}_i)^2 \leq \exp\left[-\frac{(a-z_{k-1})_+^2}{2}\right]+\exp\left[-\frac{(a-z_{k-1})^2}{n-k+2}\right].$$

By taking inspiration of this upper bound, we can propose to use the following explicit potential functions that turn out to be good approximations of the exact optimal potential functions G_k^*: for $1\leq k\leq n-1$,

$$G_k(\mathbf{z}_k)=\exp\left[-\frac{(z_k-a)^2}{2(n-k+1)}+\frac{(z_{k-1}-a)^2}{2(n-k+2)}\right] \tag{4.5}$$

and

$$G_0(\mathbf{z}_0)=\exp\left[-\frac{(z_0-a)^2}{2(n+1)}\right]. \tag{4.6}$$

In order to confirm the theoretical predictions, we have carried out a simulation study. We have run the method 1000 times with $N=2000$, $n=10$, different values

of a, and different potential functions. For each case we have computed the empirical mean and variance of the estimations. We have first used the potentials defined in equations (4.5-4.6). The results are displayed in table 2. We compare these estimations with the actual values of p_h and the asymptotic variance of the Monte-Carlo method which is simply p_h, showing that the potentials (4.5-4.6) built with the Chernov-Bernstein large deviation inequality yield a significant variance reduction. Indeed the variance reduction compared to the Monte-Carlo method is at least by a factor 500 and at best by a factor $5\,10^8$.

a	p_h	σ^2_{MC}	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$
$4\sqrt{n}$	$3.2\,10^{-5}$	$3.2\,10^{-5}$	$3.2\,10^{-5}$	$6.5\,10^{-8}$
$5\sqrt{n}$	$2.9\,10^{-7}$	$2.9\,10^{-7}$	$2.9\,10^{-7}$	$1.3\,10^{-11}$
$6\sqrt{n}$	$9.9\,10^{-10}$	$9.9\,10^{-10}$	$1.0\,10^{-9}$	$7.4\,10^{-16}$
$7\sqrt{n}$	$1.3\,10^{-12}$	$1.3\,10^{-12}$	$1.2\,10^{-12}$	$2.6\,10^{-21}$

Table 2. Theoretical and empirical comparisons (example 1) with the potentials (4.5-4.6). The results are obtained with $N = 2000$ and $n = 10$. The IPS algorithm has been run 1000 times to get the empirical values $mean(\hat{p}_h)$ and $\hat{\sigma}^2_{IPS,G}$.

We have also compared the efficiencies of different potentials. We have run the method 1000 times with $N = 2000$, $n = 10$, $a = 15$, and different potential functions:
1) the potential that selects the best values $G_k(\mathbf{z}_k) = \exp[\alpha z_k]$ (we have determined that the optimal parameter α is $\alpha = 0.22$);
2) the potential used on a Gaussian random walk in [18] that selects the particles with the best increments: $G_k(\mathbf{z}_k) = \exp[\alpha(z_k - z_{k-1})]$ (we have determined that the optimal parameter α is $\alpha = 1.4$);
3) the potential (4.5-4.6) built with the Chernov-Bernstein large deviation inequality;
4) the optimal potential $(G^*_k)_{k<n}$ defined by (4.3-4.4) that we can compute using Gauss-Hermite quadrature formulas.
The results displayed in table 3 show that
1) By comparing $mean(\hat{p}_h)$ with p_h, we can see that the estimator is unbiased whatever the choice of the potential.
2) By comparing the empirical variance multiplied by N of the IPS estimator $\hat{\sigma}^2_{IPS,G}$, which is close to the asymptotic variance of the IPS estimator $\sigma^2_{IPS,G}$, to the asymptotic variance of the standard Monte Carlo estimator σ^2_{MC}, which is simply p_h, we can see that the IPS method provides a significant variance reduction, as the variance can be reduced by a factor 10^4 with the optimal potential.
3) By comparing the two potentials parameterized with α, we conclude that it is important to take into account the increments in the selection process and not only the

values, because the minimal variance obtained with a potential of the form $G_k(\mathbf{z}_k) = \exp[\alpha z_k]$ (the minimal variance is obtained with $\alpha = 0.22$) is 17 times larger than the minimal variance obtained with a potential of the form $G_k(\mathbf{z}_k) = \exp[\alpha(z_k - z_{k-1})]$ (the minimal variance is obtained with $\alpha = 1.4$).
4) The optimal potential functions $(G_k^*)_{k<n}$ yield the best variance but the potential (4.5-4.6) has almost the same performance, and the potential $G_k(\mathbf{z}_k) = \exp[\alpha(z_k - z_{k-1})]$ with the tuned parameter $\alpha = 1.4$ has also a similar performance.

$G_k(\mathbf{z}_k)$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$
$\exp[\alpha z_k]$, $\alpha = 0.14$	$1.05\,10^{-6}$	$5.6\,10^{-9}$
$\exp[\alpha z_k]$, $\alpha = 0.16$	$1.03\,10^{-6}$	$3.7\,10^{-9}$
$\exp[\alpha z_k]$, $\alpha = 0.18$	$1.03\,10^{-6}$	$3.5\,10^{-9}$
$\exp[\alpha z_k]$, $\alpha = 0.20$	$1.08\,10^{-6}$	$2.9\,10^{-9}$
$\exp[\alpha z_k]$, $\boldsymbol{\alpha = 0.22}$	$1.03\,10^{-6}$	$\mathbf{2.8\,10^{-9}}$
$\exp[\alpha z_k]$, $\alpha = 0.24$	$1.08\,10^{-6}$	$3.3\,10^{-9}$
$\exp[\alpha z_k]$, $\alpha = 0.26$	$1.09\,10^{-6}$	$3.6\,10^{-9}$
$\exp[\alpha z_k]$, $\alpha = 0.28$	$1.05\,10^{-6}$	$4.0\,10^{-9}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.0$	$1.05\,10^{-6}$	$3.2\,10^{-10}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.1$	$1.06\,10^{-6}$	$2.5\,10^{-10}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.2$	$1.04\,10^{-6}$	$1.9\,10^{-10}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.3$	$1.05\,10^{-6}$	$1.7\,10^{-10}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\boldsymbol{\alpha = 1.4}$	$1.05\,10^{-6}$	$\mathbf{1.7\,10^{-10}}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.5$	$1.05\,10^{-6}$	$1.8\,10^{-10}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.6$	$1.05\,10^{-6}$	$2.0\,10^{-10}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.7$	$1.05\,10^{-6}$	$2.4\,10^{-10}$
$G_k(\mathbf{z}_k)$ Eqs. (4.5-4.6)	$1.05\,10^{-6}$	$\mathbf{1.5\,10^{-10}}$
$G_k^*(\mathbf{z}_k)$ Eqs. (4.3-4.4)	$1.05\,10^{-6}$	$\mathbf{1.3\,10^{-10}}$

Table 3. Comparisons of the efficiencies of different potentials (example 1). The results are obtained for $N = 2000$, $n = 10$, $a = 15$ (so $p_h = 1.05\,10^{-6}$ and $\sigma^2_{MC} = 1.05\,10^{-6}$). The IPS algorithm has been run 1000 times to get the empirical values $mean(\hat{p}_h)$ and $\hat{\sigma}^2_{IPS,G}$.

4.2. Second example

Here we consider the Ornstein-Uhlenbeck process, that is the solution of the stochastic differential equation:

$$dX_t = \sqrt{2}dW_t - X_t dt, \qquad X_0 = 0. \tag{4.7}$$

It is a Markov, Gaussian, mean-reverting, and ergodic process. The stationary distribution is $\mathcal{N}(0,1)$ and the mean-reversion time is 1. If the process is observed at times $t_k = k\tau$, $k = 0, \ldots, n$, with $\tau > 0$, then the observed process $Z_k = X_{t_k}$ is a homogeneous Markov chain with kernel

$$Q(dz_k|z_{k-1}) = \frac{1}{\sqrt{2\pi(1-e^{-2\tau})}} \exp\Big[-\frac{(z_k - z_{k-1}e^{-\tau})^2}{2(1-e^{-2\tau})}\Big] dz_k,$$

in other words the law of Z_k knowing $(Z_j)_{j\leq k-1}$ is Gaussian with mean $Z_{k-1}e^{-\tau}$ and variance $1-e^{-2\tau}$. The goal is to compute the probability that X_{t_n} exceeds a large positive value a. Therefore we take $h(\mathbf{z}_n) = \mathbb{1}_{[a,+\infty)}(z_n)$ so that $p_h = \mathbb{P}(Z_n \geq a)$. In this case we have for $0 \leq k \leq n-1$:

$$\begin{aligned}\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k = \mathbf{z}_k] &= \frac{1}{\sqrt{2\pi(1-e^{-2(n\text{-}k)\tau})}} \int_a^\infty \exp\left[\text{-}\frac{(z_n' \text{-} z_k e^{\text{-}(n\text{-}k)\tau})^2}{2(1-e^{-2(n\text{-}k)\tau})}\right] dz_n' \\ &= \Phi\Big(\frac{z_k e^{\text{-}(n\text{-}k)\tau} \text{-} a}{\sqrt{1-e^{-2(n\text{-}k)\tau}}}\Big),\end{aligned}$$

and the optimal potential is: for $1 \leq k \leq n-2$,

$$G_k^*(\mathbf{z}_k) = \sqrt{\frac{\int_{\mathbb{R}} \Phi\Big(\frac{z_{k+1}' e^{\text{-}(n\text{-}k\text{-}1)\tau} \text{-} a}{\sqrt{1-e^{-2(n\text{-}k\text{-}1)\tau}}}\Big)^2 \exp\left[\text{-}\frac{(z_{k+1}' \text{-} z_k e^{-\tau})^2}{2(1-e^{-2\tau})}\right] dz_{k+1}'}{\int_{\mathbb{R}} \Phi\Big(\frac{z_k' e^{\text{-}(n\text{-}k)\tau} \text{-} a}{\sqrt{1-e^{-2(n\text{-}k)\tau}}}\Big)^2 \exp\left[\text{-}\frac{(z_k' \text{-} z_{k\text{-}1} e^{-\tau})^2}{2(1-e^{-2\tau})}\right] dz_k'}}, \tag{4.8}$$

and

$$G_{n-1}^*(\mathbf{z}_{n-1}) = \sqrt{\frac{\int_{\mathbb{R}} \mathbf{1}_{[a,+\infty)}(z_n') \exp\left[\text{-}\frac{(z_n' \text{-} z_{n-1} e^{-\tau})^2}{2(1-e^{-2\tau})}\right] dz_n'}{\int_{\mathbb{R}} \Phi\Big(\frac{z_{n-1}' e^{\text{-}\tau} \text{-} a}{\sqrt{1-e^{-2\tau}}}\Big)^2 \exp\left[\text{-}\frac{(z_{n-1}' \text{-} z_{n\text{-}2} e^{-\tau})^2}{2(1-e^{-2\tau})}\right] dz_{n-1}'}}, \tag{4.9}$$

$$G_0^*(\mathbf{z}_0) = \sqrt{\int_{\mathbb{R}} \Phi\Big(\frac{z_1' e^{\text{-}(n\text{-}1)\tau} \text{-} a}{\sqrt{1-e^{-2(n\text{-}1)\tau}}}\Big)^2 \exp\left[\text{-}\frac{(z_1' \text{-} z_0 e^{-\tau})^2}{2(1-e^{-2\tau})}\right] dz_1'}. \tag{4.10}$$

The Chernov-Bernstein's inequality gives the rather sharp exponential bound:

$$\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k = \mathbf{z}_k] \leq \exp\left[-\frac{(a - z_k e^{-(n-k)\tau})_+^2}{2(1-e^{-2(n-k)\tau})}\right].$$

By taking inspiration of this upper bound, we can propose to use the following explicit potential functions that turn out to be good approximations of the exact optimal potential functions G_k^*: for $1 \leq k \leq n-1$,

$$G_k(\mathbf{z}_k) = \exp\left[-\frac{\left(z_k e^{-(n-k)\tau} - a\right)^2}{2(1+e^{-2(n-k-1)\tau} - 2e^{-2(n-k)\tau})} + \frac{\left(z_{k-1} e^{-(n-k+1)\tau} - a\right)^2}{2(1+e^{-2(n-k)\tau} - 2e^{-2(n-k+1)\tau})} \right] \tag{4.11}$$

and

$$G_0(\mathbf{z}_0) = \exp\left[-\frac{\left(z_0 e^{-n\tau} - a\right)^2}{2(1+e^{-2(n-1)\tau} - 2e^{-2n\tau})} \right]. \tag{4.12}$$

The form of the quasi-optimal potential G_k is instructive: we can observe that, as long as $(n-k)\tau \gg 1$, the potential is constant, which means that there is no selection in the early steps of the dynamics. Selection is effective when the remaining time $(n-k)\tau$ is of order one, that is the mean-reversion time of the process X_t. This result is actually general and can be extended to all ergodic, mean-reverting processes. Indeed, since $\mathbb{E}[h(\mathbf{Z}_n)|\mathbf{Z}_k = \mathbf{z}_k]$ is constant (and approximately equal to $\mathbb{E}[h(\mathbf{Z}_n)]$) when the time difference between k and n is larger than the mean-reversion time, the selection pressure of the optimal selection scheme is low in the early steps. It is only high when the remaining time is of the order of the mean-reversion time. This also explains why the potential $G_k(\mathbf{z}_k) = \exp[\alpha(z_k - z_{k-1})]$ is reasonably efficient (i.e. much better than standard Monte Carlo) but not as efficient as the optimal potential for these cases (see below).

We compare the efficiencies of different potentials in tables 4-6 for different values of t_n. It appears that the quasi-optimal potential (4.11-4.12) has always a performance that is close to the optimal potential G_k^* defined by (4.8-4.10). When t_n is smaller than one, then the mean-reversion force is low and the situation is close to the random Gaussian walk. The potential $G_k(\mathbf{z}_k) = \exp[\alpha(z_k - z_{k-1})]$ with the optimal α ($\alpha = 3.5$ for $t_n = 1$) has a performance that is not far from the one of the optimal potential G_k^*. When t_n is larger than one, then the mean reversion is strong and an efficient selection is concentrated on the last time steps. This is what the optimal potential G_k^* and the quasi-optimal potential (4.11-4.12) do and this explains why the variance reduction is then very strong. The potential $G_k(\mathbf{z}_k) = \exp[\alpha(z_k - z_{k-1})]$ with the optimal α ($\alpha = 2.5$ for $t_n = 2$ and $\alpha = 2$ for $t_n = 3$) imposes a constant selection pressure throughout the history of the Markov process, it does much better than Monte Carlo (by a factor 50–100 in terms of variance), but not as good as the optimal or quasi-optimal potentials (by a factor 2–3 in terms of variance).

$G_k(\mathbf{z}_k)$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$
$\exp[\alpha z_k]$, $\alpha = 0.1$	$8.6\,10^{-6}$	$2.3\,10^{-6}$
$\exp[\alpha z_k]$, $\alpha = 0.2$	$8.4\,10^{-6}$	$7.7\,10^{-7}$
$\exp[\alpha z_k]$, $\alpha = 0.3$	$8.6\,10^{-6}$	$4.7\,10^{-7}$
$\exp[\alpha z_k]$, $\boldsymbol{\alpha = 0.4}$	$8.6\,10^{-6}$	$\mathbf{3.2\,10^{-7}}$
$\exp[\alpha z_k]$, $\alpha = 0.5$	$8.2\,10^{-6}$	$3.5\,10^{-7}$
$\exp[\alpha z_k]$, $\alpha = 0.6$	$8.1\,10^{-6}$	$5.4\,10^{-7}$
$\exp[\alpha z_k]$, $\alpha = 0.7$	$8.1\,10^{-6}$	$8.7\,10^{-7}$
$\exp[\alpha z_k]$, $\alpha = 0.8$	$7.8\,10^{-6}$	$1.5\,10^{-6}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 2.0$	$8.6\,10^{-6}$	$5.2\,10^{-8}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 2.5$	$8.5\,10^{-6}$	$2.8\,10^{-8}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 3.0$	$8.4\,10^{-6}$	$2.0\,10^{-8}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\boldsymbol{\alpha = 3.5}$	$8.5\,10^{-6}$	$\mathbf{1.8\,10^{-8}}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 4.0$	$8.5\,10^{-6}$	$2.5\,10^{-8}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 4.5$	$8.5\,10^{-6}$	$3.8\,10^{-8}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 5.0$	$8.5\,10^{-6}$	$1.0\,10^{-7}$
$G_k(\mathbf{z}_k)$ Eqs. (4.11-4.12)	$8.5\,10^{-6}$	$\mathbf{1.5\,10^{-8}}$
$G^*_k(\mathbf{z}_k)$ Eqs. (4.8-4.10)	$8.5\,10^{-6}$	$\mathbf{1.4\,10^{-8}}$

Table 4. Comparisons of the efficiencies of potentials (example 2). The results are obtained for $N = 2000$, $a = 4$, $\tau = 0.1$, $n = 10$. Here $t_n = 1$ and $p_h = 8.5\,10^{-6}$ (so $\sigma^2_{MC} = 8.5\,10^{-6}$). The IPS algorithm has been run 1000 times to get the empirical values $mean(\hat{p}_h)$ and $\hat{\sigma}^2_{IPS,G}$.

4.3. Third example

We finally consider a diffusion process, that is the solution of the stochastic differential equation

$$dX_t = \sqrt[4]{1 + X_t^2}\, dW_t - (X_t - 0.5\sin(X_t))dt, \qquad X_0 = 0. \tag{4.13}$$

The discretized process $Z_k = X_{k\tau}$ is Markov. The goal is to compute the probability that $X_{n\tau}$ exceeds a large positive value a. Therefore we take $h(\mathbf{z}_n) = \mathbb{1}_{[a,+\infty)}(z_n)$ so that $p_h = \mathbb{P}(Z_n \geq a)$. There is no way to compute analytically p_h. Table 7 shows that the IPS method can easily reduce the variance of the estimation of p_h compared to the standard Monte Carlo method by a factor up to 7000 when $a = 6$ and $p_h \simeq 1.3\,10^{-7}$.

$G_k(\mathbf{z}_k)$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$
$\exp[\alpha z_k]$, $\alpha = 0.05$	$2.6\,10^{-5}$	$1.0\,10^{-5}$
$\exp[\alpha z_k]$, $\alpha = 0.10$	$2.6\,10^{-5}$	$5.6\,10^{-6}$
$\exp[\alpha z_k]$, $\boldsymbol{\alpha = 0.15}$	$2.7\,10^{-5}$	$\mathbf{4.3\,10^{-6}}$
$\exp[\alpha z_k]$, $\alpha = 0.20$	$2.7\,10^{-5}$	$5.1\,10^{-6}$
$\exp[\alpha z_k]$, $\alpha = 0.25$	$2.7\,10^{-5}$	$1.9\,10^{-5}$
$\exp[\alpha z_k]$, $\alpha = 0.30$	$2.8\,10^{-5}$	$5.2\,10^{-5}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.0$	$2.8\,10^{-5}$	$1.4\,10^{-6}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.5$	$2.8\,10^{-5}$	$4.5\,10^{-7}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 2.0$	$2.6\,10^{-5}$	$2.1\,10^{-7}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\boldsymbol{\alpha = 2.5}$	$2.7\,10^{-5}$	$\mathbf{1.4\,10^{-7}}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 3.0$	$2.7\,10^{-5}$	$1.9\,10^{-7}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 3.5$	$2.7\,10^{-5}$	$4.6\,10^{-7}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 4.0$	$2.8\,10^{-5}$	$2.4\,10^{-6}$
$G_k(\mathbf{z}_k)$ Eqs. (4.11-4.12)	$2.7\,10^{-5}$	$\mathbf{7.7\,10^{-8}}$
$G^*_k(\mathbf{z}_k)$ Eqs. (4.8-4.10)	$2.7\,10^{-5}$	$\mathbf{7.5\,10^{-8}}$

Table 5. Comparisons of the efficiencies of potentials (example 2). The results are obtained for $N = 2000$, $a = 4$, $\tau = 0.1$, $n = 20$. Here $t_n = 2$ and $p_h = 2.7\,10^{-5}$ (so $\sigma^2_{MC} = 2.7\,10^{-5}$). The IPS algorithm has been run 1000 times to get the empirical values $mean(\hat{p}_h)$ and $\hat{\sigma}^2_{IPS,G}$.

4.4. Fourth example

We consider again the diffusion process (4.13) and its discretized version $Z_k = X_{k\tau}$. The goal is now to compute the probability that $V(X_{n\tau})$ exceeds a large positive value a, where $V(x) = |x|$. Therefore we take $h(\mathbf{z}_n) = \mathbb{1}_{[a,+\infty)}(V(z_n))$ so that $p_h = \mathbb{P}(V(Z_n) \geq a)$ (of course, by obvious symmetry, p_h is here equal to twice the value of p_h in the previous section). This situation could seem more tricky than the one addressed in the previous section, because there are apparently two modes of large deviations, either the process X_t takes large values or it takes small values. Under such circumstances, an importance sampling strategy, if it could be implemented, would have to be calibrated carefully with a bimodal biased distribution. The IPS method is very robust to this kind of difficulties and automatically allocates particles in both important regions. Table 8 shows that the IPS method can easily reduce the variance of the estimation of p_h compared to the standard Monte Carlo method by a factor up to 4000 when $a = 6$ and $p_h \simeq 2.5\,10^{-7}$.

$G_k(\mathbf{z}_k)$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$
$\exp[\alpha z_k]$, $\alpha = 0.05$	$3.2\,10^{-5}$	$1.4\,10^{-5}$
$\exp[\alpha z_k]$, $\boldsymbol{\alpha = 0.10}$	$3.0\,10^{-5}$	$\mathbf{1.3\,10^{-5}}$
$\exp[\alpha z_k]$, $\alpha = 0.15$	$3.1\,10^{-5}$	$2.2\,10^{-5}$
$\exp[\alpha z_k]$, $\alpha = 0.20$	$2.7\,10^{-5}$	$5.8\,10^{-5}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.0$	$3.1\,10^{-5}$	$1.6\,10^{-6}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 1.5$	$3.1\,10^{-5}$	$5.5\,10^{-7}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\boldsymbol{\alpha = 2.0}$	$3.1\,10^{-5}$	$\mathbf{3.1\,10^{-7}}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 2.5$	$3.1\,10^{-5}$	$4.6\,10^{-7}$
$\exp[\alpha(z_k\text{-}z_{k\text{-}1})]$, $\alpha = 3.0$	$3.0\,10^{-5}$	$1.7\,10^{-6}$
$G_k(\mathbf{z}_k)$ Eqs. (4.11-4.12)	$3.1\,10^{-5}$	$\mathbf{9.4\,10^{-8}}$
$G^*_k(\mathbf{z}_k)$ Eqs. (4.8-4.10)	$3.1\,10^{-5}$	$\mathbf{9.0\,10^{-8}}$

Table 6. Comparisons of the efficiencies of potentials (example 2). The results are obtained for $N = 2000$, $a = 4$, $\tau = 0.1$, $n = 30$. Here $t_n = 3$ and $p_h = 3.1\,10^{-5}$ (so $\sigma^2_{MC} = 3.1\,10^{-5}$). The IPS algorithm has been run 1000 times to get the empirical values $mean(\hat{p}_h)$ and $\hat{\sigma}^2_{IPS,G}$.

5. Discussion and conclusion

In this paper we give the expressions of the potential functions minimizing the asymptotic variance of the IPS method with multinomial resampling. The existence of optimal potential functions proves that the possible variance reduction of an IPS method is lower-bounded. The expressions have been validated analytically, and the expression for the minimal variance has been empirically confirmed in numerical simulations.

The analysis of the expressions of the optimal potential functions confirm empirical observations reported in the literature. First, in [12] the authors make the observation that it seems better to build a potential that depends on the increments of an energy function, this observation is confirmed as the optimal potential function is the multiplicative increment of the quantity $\sqrt{\mathbb{E}\Big[\,\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k\big]^2\big|\mathbf{Z}_{k-1}=\mathbf{z}_{k-1}\Big]}$, which is then the optimal energy function. Second, the fact that in [25] the authors find better results with time-dependent potentials is explained by the fact the expressions of the optimal potential functions show an explicit dependence on k. Third, one can notice that the optimal potential function $G^*_k(\mathbf{z}_k)$ only depends on z_k, z_{k-1} and on k. This shows that there is no advantage in looking for potential functions that depend on other variables. Fourth, if the underlying Markov process is ergodic and mean-reverting, then an efficient potential does not impose any pressure selection in the early steps of the random dynamics but only in the final steps, during the time interval that preceeds the target

	$a=3$		$a=4$		$a=5$		$a=6$	
α	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$
0.5	$7.9\,10^{-4}$	$3.1\,10^{-4}$	$4.5\,10^{-5}$	$1.0\,10^{-5}$	$2.7\,10^{-6}$	$4.0\,10^{-7}$	$1.4\,10^{-7}$	$1.3\,10^{-8}$
1.0	$7.8\,10^{-4}$	$1.2\,10^{-4}$	$4.4\,10^{-5}$	$2.6\,10^{-6}$	$2.3\,10^{-6}$	$5.6\,10^{-8}$	$1.6\,10^{-7}$	$1.6\,10^{-9}$
1.5	$7.9\,10^{-4}$	$5.8\,10^{-5}$	$4.4\,10^{-5}$	$8.4\,10^{-7}$	$2.3\,10^{-6}$	$1.3\,10^{-8}$	$1.3\,10^{-7}$	$2.3\,10^{-10}$
2.0	$7.9\,10^{-4}$	$3.8\,10^{-5}$	$4.4\,10^{-5}$	$3.9\,10^{-7}$	$2.3\,10^{-6}$	$4.4\,10^{-9}$	$1.3\,10^{-7}$	$5.6\,10^{-11}$
2.5	$7.9\,10^{-4}$	$\mathbf{3.7\,10^{-5}}$	$4.4\,10^{-5}$	$\mathbf{2.9\,10^{-7}}$	$2.3\,10^{-6}$	$\mathbf{2.4\,10^{-9}}$	$1.3\,10^{-7}$	$\mathbf{1.9\,10^{-11}}$
3.0	$7.9\,10^{-4}$	$1.3\,10^{-4}$	$4.4\,10^{-5}$	$6.1\,10^{-7}$	$2.3\,10^{-6}$	$4.0\,10^{-9}$	$1.3\,10^{-7}$	$3.5\,10^{-11}$
3.5	$8.4\,10^{-4}$	$6.2\,10^{-3}$	$4.9\,10^{-5}$	$4.6\,10^{-4}$	$2.1\,10^{-6}$	$1.8\,10^{-8}$	$1.2\,10^{-7}$	$1.6\,10^{-10}$
4.0	$5.4\,10^{-4}$	$1.9\,10^{-2}$	$4.2\,10^{-5}$	$4.9\,10^{-4}$	$2.1\,10^{-6}$	$8.6\,10^{-7}$	$1.2\,10^{-7}$	$1.3\,10^{-9}$

Table 7. Comparisons of the efficiencies of the potentials $\exp\left[\alpha(z_k\text{-}z_{k\text{-}1})\right]$ for different α in order to estimate $p_h = \mathbb{P}(Z_n \geq a)$ (example 3). The results are obtained for $N = 2000$, $a \in \{3, 4, 5, 6\}$, $\tau = 0.1$, $n = 10$ ($t_n = 1$). The SDE (4.13) is solved by the Euler-Maruyama method with a time step 10^{-3}. The IPS algorithm has been run 1000 times to get the empirical values $mean(\hat{p}_h)$ and $\hat{\sigma}^2_{IPS,G}$.

time and that has a width of the order of the mean-reversion time. Finally, as splitting methods can be viewed as a version of the IPS method with indicator potential functions, our results show that the selections of splitting algorithms are not optimal, and could be improved by using information on the expectations $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}_k\big]$.

The optimal potential functions may be hard to find in practice. Indeed, the expectations $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}_k\big]$ play a big role in the expression of the optimal potentials, but if we are trying to assess $p_h = \mathbb{E}\big[h(\mathbf{Z}_n)\big]$, we typically lack information about the conditional expectations $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}_k\big]$. If no information on the expectation $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}_k\big]$ is available, it might be preferable to use more naive variance reduction method, where no input functions are needed. In such context, the Weighted Ensemble (WE) method [1, 2] seems to be a good candidate, as it does not take in input potential functions but only a partition of the state space. Conversely if the practitioner has qualitative information about the expectation $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}_k\big]$, this information could be used to derive very efficient potentials. This information could be acquired through an adaptive version of the IPS algorithm.

We believe that a favorable context for the optimal IPS method is the multifidelity Monte Carlo framework [23]. Here we typically consider a complex system whose simulation cost by a high-fidelity code is very large, but it may happen that a low-fidelity code is available that is faster to run but approximate. Under such circumstances, the low-fidelity code could be used to estimate the optimal potential functions, and then the IPS method could be applied with the high-fidelity code.

	$a=3$		$a=4$		$a=5$		$a=6$	
α	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$	$mean(\hat{p}_h)$	$\hat{\sigma}^2_{IPS,G}$
0.5	$1.6\,10^{-3}$	$8.6\,10^{-4}$	$9.0\,10^{-5}$	$2.7\,10^{-5}$	$5.2\,10^{-6}$	$1.0\,10^{-6}$	$2.8\,10^{-7}$	$3.6\,10^{-8}$
1.0	$1.6\,10^{-3}$	$3.8\,10^{-4}$	$8.9\,10^{-5}$	$7.8\,10^{-6}$	$4.5\,10^{-6}$	$1.7\,10^{-7}$	$2.4\,10^{-7}$	$3.9\,10^{-9}$
1.5	$1.6\,10^{-3}$	$1.9\,10^{-4}$	$8.9\,10^{-5}$	$3.0\,10^{-6}$	$4.5\,10^{-6}$	$4.4\,10^{-8}$	$2.6\,10^{-7}$	$9.7\,10^{-10}$
2.0	$1.6\,10^{-3}$	$\mathbf{1.3\,10^{-4}}$	$8.8\,10^{-5}$	$1.5\,10^{-6}$	$4.5\,10^{-6}$	$1.6\,10^{-8}$	$2.5\,10^{-7}$	$1.6\,10^{-10}$
2.5	$1.6\,10^{-3}$	$1.4\,10^{-4}$	$8.8\,10^{-5}$	$\mathbf{1.2\,10^{-6}}$	$4.6\,10^{-6}$	$\mathbf{1.0\,10^{-8}}$	$2.5\,10^{-7}$	$\mathbf{7.0\,10^{-11}}$
3.0	$1.6\,10^{-3}$	$4.2\,10^{-4}$	$8.8\,10^{-5}$	$2.9\,10^{-6}$	$4.5\,10^{-6}$	$1.8\,10^{-8}$	$2.6\,10^{-7}$	$1.0\,10^{-10}$
3.5	$1.5\,10^{-3}$	$8.7\,10^{-3}$	$8.7\,10^{-5}$	$3.7\,10^{-5}$	$4.3\,10^{-6}$	$1.2\,10^{-7}$	$2.5\,10^{-7}$	$7.7\,10^{-10}$
4.0	$1.3\,10^{-3}$	$1.4\,10^{-1}$	$8.4\,10^{-5}$	$7.5\,10^{-4}$	$3.9\,10^{-6}$	$3.7\,10^{-6}$	$2.4\,10^{-7}$	$2.4\,10^{-8}$

Table 8. Comparisons of the efficiencies of the potentials $\exp\left[\alpha(V(z_k)\text{-}V(z_{k\text{-}1}))\right]$ in order to estimate $p_h = \mathbb{P}(V(Z_n) \geq a)$ for $V(z) = |z|$ and for different α (example 4). The results are obtained for $N = 2000$, $a \in \{3, 4, 5, 6\}$, $\tau = 0.1$, $n = 10$ ($t_n = 1$). The SDE (4.13) is solved by the Euler-Maruyama method with a time step 10^{-3}. The IPS algorithm has been run 1000 times to get the empirical values $mean(\hat{p}_h)$ and $\hat{\sigma}^2_{IPS,G}$.

The (partial) knowledge of the expectations $\mathbb{E}\big[h(\mathbf{Z}_n)\big|\mathbf{Z}_k = \mathbf{z}_k\big]$ is therefore important for a well optimized use of the IPS method, but it is interesting to remark that the same knowledge seems to be important for a well optimized importance sampling method [9]. This confirms the well known fact that, with a good knowledge of the dynamics of the process $(\mathbf{Z}_k)_{k\geq 0}$, the importance sampling method may be preferable to the IPS if it can be implemented. Nonetheless the IPS method may still be preferred to the importance sampling method because 1) it is not intrusive and 2) it should be preferred when one fears to be in an over-biasing situation, because the IPS method does not alter the propagation and the over-biasing phenomenon should then be less important than with importance sampling methods.

References

1. D. Aristoff, *Analysis and optimization of weighted ensemble sampling*, ESAIM: Mathematical Modelling and Numerical Analysis 52 (2018), pp. 1219–1238.
2. D. Aristoff, *Optimizing weighted ensemble sampling for steady states and rare events*, SIAM Multiscale Model. Simul. 18 (2020), pp. 646–673.
3. B. A. Berg, *Introduction to Multicanonical Monte Carlo simulations*, Fields Institute Communications 26 (2000), pp. 1–24.
4. F. Cérou, P. Del Moral and A. Guyader, *A nonasymptotic theorem for unnormalized Feynman–Kac particle models*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 47 (2011), pp. 629–649.

5. F. Cérou and A. Guyader, *Adaptive multilevel splitting for rare event analysis*, Stochastic Analysis and Applications 25 (2007), pp. 417–443.

6. F. Cérou, F. LeGland, P. Del Moral and P. Lezaud, *Limit theorems for the multilevel splitting algorithm in the simulation of rare events*. Simulation Conference, Orlando, United States. Proceedings of the Winter, pp. 682–691, IEEE, dec 2005.

7. H.P. Chan and T.L. Lai, *A general theory of particle filters in hidden Markov models and some applications*, Annals of Statistics 41 (2013), pp. 2877–2904.

8. N. Chopin, *Central limit theorem for sequential Monte Carlo methods and its applications to Bayesian inference*, Annals of Statistics 32 (2004), pp. 2385–2411.

9. H. Chraibi, A. Dutfoy, T. Galtier and J. Garnier, *On the optimal importance process for piecewise deterministic Markov process*, ESAIM: Probability and Statistics 23 (2019), pp. 893–921.

10. P. Del Moral, *Feynman-Kac Formulae, Genealogical and Interacting Particle Systems With Applications*. Springer, New York, 2004.

11. P. Del Moral, A. Doucet and A. Jasra, *Sequential Monte-Carlo samplers*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (2006), pp. 411–436.

12. P. Del Moral and J. Garnier, *Genealogical particle analysis of rare events*, Annals of Applied Probability 15 (2005), pp. 2496–2534.

13. R. Douc, E. Moulines and J. Olsson, *Optimality of the auxiliary particle filter*, Probability and Mathematical Statistics 29 (2009), pp. 1–28.

14. A. Doucet, N. de Freitas and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.

15. J. Garnier and P. Del Moral, *Simulations of rare events in fiber optics by interacting particle systems*, Opt. Commun. 267 (2006), pp. 205–214.

16. M. Gerber, N. Chopin and N. Whiteley, *Negative association, ordering and convergence of resampling methods*, Annals of Statistics 47 (2019), pp. 2236–2260.

17. P. Guarniero, A. M. Johansen and A. Lee, *The iterated auxiliary particle filter*, Journal of the American Statistical Association 112 (2017), pp. 1636–1647.

18. D. Jacquemart and J. Morio, *Tuning of adaptive interacting particle system for rare event probability estimation*, Simulation Modelling Practice and Theory 66 (2016), pp. 36–49.

19. H.-R. Künsch, *State space and hidden Markov models*, Complex Stochastic Systems (D. R. Cox O. E. Barndorff-Nielsen and C. Klüppelberg, eds.), Chapman and Hall, London, 2001, pp. 109–173.

20. _____, *Recursive Monte Carlo filters: Algorithms and theoretical analysis*, Annals of Statistics 33 (2005), pp. 1983–2021.

21. A. Lee and N. Whiteley, *Variance estimation in the particle filter*, Biometrika 105 (2018), pp. 609–625.

22. J. Morio, D. Jacquemart, M. Balesdent and J. Marzat, *Optimisation of interacting particle systems for rare event estimation*, Computational Statistics & Data Analysis 66 (2013), pp. 117–128.

23. B. Peherstorfer, K. Willcox and M. Gunzburger, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, SIAM Review 60 (2018), pp. 550–591.

24. R. Y. Rubinstein, *Simulation and the Monte Carlo method*. Wiley, New York, 1981.

25. J. Wouters and F. Bouchet, *Rare event computation in deterministic chaotic systems using genealogical particle analysis*, Journal of Physics A: Mathematical and Theoretical 49 (2016), p. 374002.

Author information

Hassane Chraibi, Électricité de France (EDF), PERICLES department, 7 Boulevard Gaspard Monge, 91120 Palaiseau, France.
Email: hassane.chraibi@edf.fr

Anne Dutfoy, Électricité de France (EDF), PERICLES department, 7 Boulevard Gaspard Monge, 91120 Palaiseau, France.
Email: anne.dutfoy@edf.fr

Thomas Galtier, CMAP, École polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau Cedex, France.
Email: thomas.galtier@polytechnique.edu

Josselin Garnier, CMAP, École polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau Cedex; INRIA, France.
Email: josselin.garnier@polytechnique.edu