HAL
open science

Unbiasedness of some generalized adaptive multilevel splitting algorithms

Charles-Edouard Bréhier, Maxime Gazeau, Ludovic Goudenège, Tony Lelièvre, Mathias Rousset

To cite this version:

Charles-Edouard Bréhier, Maxime Gazeau, Ludovic Goudenège, Tony Lelièvre, Mathias Rousset. Unbiasedness of some generalized adaptive multilevel splitting algorithms. The Annals of Applied Probability, 2016, 26 (6), pp.3559 - 3601. 10.1214/16-AAP1185 . hal-01142704

HAL Id: hal-01142704

https://hal.science/hal-01142704

Submitted on 15 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unbiasedness of some generalized Adaptive Multilevel Splitting algorithms

Charles-Edouard Bréhier*[1], Maxime Gazeau[2], Ludovic Goudenège[3], Tony Lelièvre[4], and Mathias Rousset[4]

[1]Université de Neuchâtel, Institut de Mathématiques, Rue Emile Argand 11, CH-2000 Neuchâtel, Switzerland.
[2]University of Toronto, Department of Mathematics, 40 St. George St., Toronto M5S 2E4, Canada.
[3]Fédération de Mathématiques de l'École Centrale Paris, CNRS, Grande voie des vignes, 92295 Châtenay-Malabry, France.
[4]Université Paris-Est, CERMICS (ENPC), INRIA, 6-8 Avenue Blaise Pascal, Cité Descartes, F-77455 Marne-la-Vallée, France.

April 16, 2015

Abstract

We introduce a generalization of the Adaptive Multilevel Splitting algorithm in the discrete time dynamic setting, namely when it is applied to sample rare events associated with paths of Markov chains. By interpreting the algorithm as a sequential sampler in path space, we are able to build an estimator of the rare event probability (and of any non-normalized quantity associated with this event) which is unbiased, whatever the choice of the importance function and the number of replicas. This has practical consequences on the use of this algorithm, which are illustrated through various numerical experiments.

1 Introduction

The efficient sampling of rare events is a very important topic in various application fields such as reliability analysis, computational statistics or molecular dynamics. Let us describe the typical problem of interest in the context of molecular dynamics.

1.1 Motivation and mathematical setting

Let us consider the Markov chain $(X_t)_{t\in\mathbb{N}}$ defined as the discretization of the overdamped Langevin dynamics:

$$\forall t \in \mathbb{N},\ X_{t+1} - X_t = -\nabla V(X_t)\, h + \sqrt{2\beta^{-1}}(W_{t+h} - W_t). \tag{1}$$

Typically, $X_t \in \mathbb{R}^{3N}$ is a high-dimensional vector giving the positions of N particles in $\mathbb{R}^3$ at time $t\,h$ ($h > 0$ being the time step size), $V : \mathbb{R}^{3N} \to \mathbb{R}$ is the potential function (for

*charles-edouard.brehier@unine.ch, gazeauma@math.toronto.edu, goudenege@math.cnrs.fr, lelievre@cermics.enpc.fr, roussetm@cermics.enpc.fr

any set of positions $x \in \mathbb{R}^{3N}$, $V(x)$ is the energy of the configuration), $\beta^{-1} = k_B T$ is the inverse temperature and W_t is a standard Brownian motion (so that $W_{t+h} - W_t$ is a vector of $3N$ i.i.d. centered Gaussian random variables with variance h). In many cases of interest, the dynamics (1) is metastable: the vector X_t remains trapped for very long times in some so-called metastable states. This actually corresponds to a physical reality: the timescale at the molecular level (namely h, the timestep being typically chosen at the limit of stability for the stochastic differential equation) is much smaller than the timescales of interest, which correspond to hopping events between metastable states. Let us denote by $A \subset \mathbb{R}^{3N}$ and $B \subset \mathbb{R}^{3N}$ two (disjoint) metastable states. The problem is then the following: for some initial condition outside A and B, how to efficiently sample paths which reach B before A? In the context of molecular dynamics, such paths are called reactive paths. The efficient sampling of reactive paths is a very important subject in many applications since it is a way to understand the mechanism of the transition between metastable states (namely between long-living conformations of the system under study). In mathematical terms, one is interested in computing, for a given test function $\varphi : (\mathbb{R}^{3N})^{\mathbb{N}} \to \mathbb{R}$ depending on the path $(X_t)_{t\in\mathbb{N}}$ of the Markov chain, the expectation

$$\mathbb{E}\Big(\varphi\big((X_t)_{t\in\mathbb{N}}\big)1_{\tau_B<\tau_A}\Big) \tag{2}$$

where $\tau_A = \inf\{t \in \mathbb{N} : X_t \in A\}$, $\tau_B = \inf\{t \in \mathbb{N} : X_t \in B\}$ and $X_0 = x_0 \notin (A \cup B)$ is assumed (for simplicity) to be a deterministic initial position close to A: most trajectories starting from x_0 hit A before B. If $\varphi = 1$, the above expectation is $\mathbb{P}(\tau_B < \tau_A)$, namely the probability that the Markov chain reaches B before A. This is typically a very small probability: since A is metastable and x_0 is close to A, for most of the realizations, τ_A is smaller than τ_B. This is why naive Monte Carlo methods will not give reliable estimates of (2). We refer for example to [3, 10] for some examples in the context of molecular simulation.

1.2 The adaptive multilevel splitting algorithm

Many techniques have been proposed in the literature in order to compute quantities such as (2), in particular control variate techniques, importance sampling methods and splitting algorithms (see for example the monograph [6] on rare event simulations). Here, we focus on the Adaptive Multilevel Splitting (AMS) method which has been proposed in [8]. Let us roughly describe the principle of the method. The crucial ingredient we need is an importance function:

$$\xi : \mathbb{R}^{3N} \to \mathbb{R} \tag{3}$$

which will be used to measure the advance of the paths towards B. This function is known as a reaction coordinate in the molecular dynamics community, and this is the terminology we will use here. In this paper, we will also call $\xi(X_t)$ the level of the process X_t at time t. A useful requirement on ξ is the existence of $z_{\max} \in \mathbb{R}$ such that

$$B \subset \{x \in \mathbb{R}^{3N} : \xi(x) \in]z_{\max}, \infty[\}.$$

For any path of the Markov chain, we call the maximum level of this path the quantity

$$\sup\{\xi(X_{t\wedge\tau_A})_{t\in\mathbb{N}}\}.$$

Then, starting from a system of n_{rep} replicas (namely paths with initial conditions x_0 and stopped at time τ_A), the idea is to remove the *worst fitted* paths and to duplicate the *best*

fitted paths while keeping a fixed number of replicas (we will discuss below generalizations of the AMS algorithm where the number of replicas may vary). The worst fitted paths are those with the smallest maximum levels $\sup\{\xi(X_{t\wedge\tau_A})_{t\in\mathbb{N}}\}$. As soon as one of the worst fitted paths is removed, it is replaced by a resampling of one of the best fitted path: the new path is a copy of the best fitted path up to the maximum level of the removed paths, and the end of the trajectory is then sampled using independent random numbers. The algorithm thus goes through three steps: (i) the level computation step (to determine the level under which paths will be removed) ; (ii) the splitting step (to determine which paths will be removed and which ones of the remaining best fitted paths will be duplicated) ; (iii) the resampling step (to generate new paths from the selected best fitted paths). By iterating these three steps, one obtains successively systems of n_{rep} paths with an increasing minimum of the maximum levels among the replicas. The algorithm is stopped when this minimum is larger than $z_{\max}$, and an estimator of (2) is then built using a weighted empirical average over the replicas. The adaptive feature of the algorithm is in the first step (the level computation step): indeed, at each iteration, paths are removed if their maximum level is below some threshold, and these thresholds are determined iteratively using empirical quantiles, rather than by fixing a priori a deterministic sequence of levels (as it would be the case in standard sequential Monte Carlo algorithms, see [11, 7]). All the details of the algorithm will be given in Section 2.5.

In this work we focus on the application of AMS to sample Markov chains, namely discrete time stochastic dynamics, and not continuous time stochastic dynamics as in [8] for example. The reason is mainly practical: in most cases of interest, even if the original model is continuous in time, it is discretized in time when numerical approximations are needed. There are actually also many cases where the original model is discrete in time (for example kinetic Monte Carlo or Markov State Models in the context of molecular dynamics).

The discrete time setting, which is thus of practical interest, raises specific questions in the context of the AMS algorithm. First, in the resampling step, a natural question is whether the path should be copied up to the last time before or first time after it reaches the level of the removed paths. Second, in the discrete time context, it may happen that several paths have exactly the same maximum level. This implies some subtleties in the implementation of the splitting step which have a large influence on the quality of the estimators, see Section 5.1.

1.3 Main results and outline

The main results of this work are the following:

- We prove that the AMS algorithm for Markov chains with an appropriate implementation of the level computation and splitting steps yields an unbiased estimator of the rare event probability, and more generally of any non-normalized expectation related to the rare event of the form (2). We actually prove this unbiasedness result for a general class of splitting algorithm which enters into what we call the Generalized Adaptive Multilevel Splitting (GAMS) framework.

- Using this GAMS framework, we propose various generalizations of the classical AMS algorithm which all yield unbiased estimators, in particular to remove extinction and to reduce the computational cost associated with sorting procedures. Moreover, we show how the general setting we introduce allows for the sampling of other random variables than trajectories of Markov chains.

- We illustrate numerically on toy examples the importance of an appropriate implementation of the level computation and splitting steps in the AMS algorithm to get unbiasedness. We also discuss through various numerical experiments the influence of the choice of the reaction coordinate ξ on the variance of the estimators and we end up with some practical recommendations in order to get reliable estimates using the AMS algorithm (see Section 5.4). In particular, using the unbiasedness property proven in this paper, it is possible to compare the results obtained using different parameters (in particular different reaction coordinates) in order to assess the quality of the numerical results.

Compared to the previous results in the literature concerning the AMS algorithm, the main novelty of this work is the proof of the unbiasedness in a general setting and whatever the parameters: the number of replicas, the (minimum) number of resampled replicas at each iteration and the reaction coordinate ξ. The proof of unbiasedness relies on the interpretation of the AMS algorithm as a sequential Monte Carlo algorithm in path space with the reaction coordinate as a time index, in the spirit of [16] (the selection and mutation steps respectively corresponds to the branching step and the resampling step in the AMS algorithm). This analogy is made precise in Section 3.4. In previous results, see for instance [5, 15, 22], unbiasedness is proved in an idealized setting, namely for the specific reaction coordinate (known as the committor function) $\xi(x) = \mathbb{P}_x(\tau_B < \tau_A)$, and for a different resampling step, where new replicas are sampled according to the conditional distribution of paths conditioned to reach the level of the removed replicas. In many cases of practical interest, these two conditions are not met.

In addition, we illustrate through extensive numerical experiments the influence of the choice of ξ on the variance. Indeed, as for any Monte Carlo algorithm, the bias is only one part of the error when using the AMS algorithm: the statistical error (namely the variance) also plays a crucial role in the quality of the estimator as will be shown numerically in Section 5. There are unfortunately very few theoretical results concerning the influence of the choice of ξ on the statistical error. We refer to [4, 9] for an analysis of the statistical error. For discussions of the role of ξ on the statistical error, we also refer to [14, 13, 20]. In particular, in the numerical experiments, we discuss situations for which the confidence intervals of the estimators associated with different reaction coordinates do not overlap if the number of independent realizations of the algorithm is not sufficiently large. We relate this observation to the well-known phenomenon of "apparent bias" for splitting algorithms, see [14].

We would like to stress that our results hold in the setting where a family of resampling kernel indexed by the levels is available (see Section 3.1.2 for a precise definition). This applies in particular well to the sampling of the probability of trajectories of Markov dynamics (see Section 3.5.4 for another possible setting). In the terminology of [16], we have in mind the dynamic setting (considered for example in [8]), and not the static setting (considered for example in [7, 9]).

The paper is organized as follows. The AMS algorithm applied to the sampling of paths of Markov chains is described in Section 2. This algorithm actually enters into a more general framework, the Generalized Adaptive Multilevel Splitting (GAMS) framework which is described in detail in Section 3. The interest of this generalized setting is twofold. First, it is very useful to allow for variants of the classical AMS algorithm which will still yield unbiased estimators (some of them are described in Section 3.5). Second, it highlights the essential mathematical properties that are required to produce unbiased estimators of quantities such

as (2). This is the subject of Section 4 which is devoted to the main theoretical result of this work: the unbiasedness of some estimators, including estimators of (2). Finally, Section 5 is entirely devoted to some numerical experiments which illustrate the unbiasedness result, and discuss the efficiency of the AMS algorithm to sample rare events.

1.4 Notation

Before going into the details, let us provide a few general notations which are useful in the following.

- The underlying probability space is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. We recall standard notations: for m σ-fields $\mathcal{F}^1, \ldots, \mathcal{F}^m \subset \mathcal{F}$, $\mathcal{F}^1 \vee \ldots \vee \mathcal{F}^m$ denotes the smallest σ-field on Ω containing all the σ-fields $\mathcal{F}^1, \ldots, \mathcal{F}^m$. For any $t, s \in \mathbb{N}$, $t \wedge s = \min\{t, s\}$. We use the convention $\inf \emptyset = +\infty$. For two sets A and B which are disjoint, $A \sqcup B$ denotes the disjoint set union.

- We work in the following standard setting: random variables take values in state spaces $\mathcal{E}$ which are Polish (namely metrizable, complete for some distance $d_{\mathcal{E}}$ and separable). The associated Borel σ-field is denoted by $\mathcal{B}(\mathcal{E})$. We will give precise examples below (see for example Section 2.1 for the space of trajectories for Markov chains).

  Then $\mathrm{Proba}(\mathcal{E})$ denotes the set of probability distributions on $\mathcal{E}$. It is endowed with the standard Polish structure associated with the Prohorov-Levy metric which metrizes convergence in distribution, *i.e.* weak convergence of probabilities tested on continuous and bounded test functions (see for example [1]).

  The distribution of a $\mathcal{E}$-valued random variable X will be denoted by $\mathrm{Law}(X)$.

- If $\mathcal{E}_1$ and $\mathcal{E}_2$ are two Polish state spaces, a Markov kernel (or transition probability kernel) $\Pi(x_1, dx_2)$ from $\mathcal{E}_1$ to $\mathcal{E}_2$ is a measurable map from initial states in $x_1 \in \mathcal{E}_1$, to probability measures in $\mathrm{Proba}(\mathcal{E}_2)$.

- We use the following standard notation associated with probability transitions: for $\varphi : \mathcal{E}_2 \to \mathbb{R}$ a bounded and measurable test function,

$$\Pi(\varphi)(x_1) = \int_{x_2 \in \mathcal{E}_2} \varphi(x_2) \Pi(x_1, dx_2). \tag{4}$$

  Similarly, we use the notation $\pi(\varphi) = \int_{x \in \mathcal{E}_2} \varphi(x) \pi(dx)$ for $\pi \in \mathrm{Proba}(\mathcal{E}_2)$.

- Let X_1 and X_2 be random variables respectively with values in $\mathcal{E}_1$ and $\mathcal{E}_2$ and Π a Markov kernel from $\mathcal{E}_1$ to $\mathcal{E}_2$. In the algorithms we describe below, we will use the notion of conditional sampling: X_2 is sampled conditionally on X_1 with law $\Pi(X_1, \,.\,)$ (denoted by $X_2 \sim \Pi(X_1, \,.\,)$) rigorously means that $X_2 = f(X_1, U)$ a.s. where, on the one hand, U is some random variable independent of X_1 and of all the random variables introduced before (namely at previous iterations of the algorithm) and, on the other hand, f is a measurable function which is such that $\Pi(x_1, \,.\,) = \mathrm{Law}(f(x_1, U))$, for $\mathrm{Law}(X_1)$-almost every $x_1 \in \mathcal{E}_1$.

- A random system of replicas in $\mathcal{E}$ is denoted by

$$\mathcal{X} = \left(X^{(n)}\right)_{n \in I} \in \mathcal{E}^{\mathrm{rep}}, \quad \mathrm{card}\, I < +\infty, \tag{5}$$

where $I \subset \mathbb{N}^*$ is a random finite subset of labels and $(X^{(n)})_{n\in I}$ are elements of $\mathcal{E}$. The space $\mathcal{E}^{\text{rep}}$ is endowed with the following distance: for $\mathcal{X}^1 = \left(x^{(1,n)}\right)_{n\in I^1}$ and $\mathcal{X}^2 = \left(x^{(2,n)}\right)_{n\in I^2}$ in $\mathcal{E}^{\text{rep}}$, we set

$$d(\mathcal{X}^1, \mathcal{X}^2) = \begin{cases} 2 & \text{if } I^1 \neq I^2, \\ \min\left\{\sum_{n\in I^1} d_{\mathcal{E}}(x^{(1,n)}, x^{(2,n)}), 1\right\} & \text{if } I^1 = I^2. \end{cases}$$

Endowed with this distance, the set $\mathcal{E}^{\text{rep}}$ is Polish and we denote by $\mathcal{B}(\mathcal{E}^{\text{rep}})$ the Borel σ-field. This σ-field can also be written as follows:

$$\mathcal{B}(\mathcal{E}^{\text{rep}}) = \bigsqcup_{I\in\mathcal{I}} \mathcal{B}(\mathcal{E})^{\operatorname{card}\mathcal{I}}$$

where $\mathcal{I}$ denotes the ensemble of finite subsets of $\mathbb{N}^*$ (which is a discrete set).

- When we consider systems of weighted replicas, to each replica $X^{(n)}$ of the system $\mathcal{X}$ with label $n \in I$ is attached a weight $G^{(n)} \in \mathbb{R}_+$, and we use the notation $\mathcal{X} = \left(X^{(n)}, G^{(n)}\right)_{n\in I}$. The topological setting is the same as in the previous item, $\mathcal{E}$ being replaced by the augmented state space $\mathcal{E} \times \mathbb{R}$.

2 The AMS algorithm for Markov chains

The aims of this section are to define the AMS algorithm applied to paths of a Markov chain (namely discrete time stochastic processes) and to state the main result on the unbiasedness of properly defined estimators in this setting.

A special care should be taken to treat the situations when many replicas have the same maximum level, or the situations when there is extinction of the population of replicas. These aspects were not treated in details in many previous works where continuous time diffusions were considered.

2.1 The Markov chain setting

Let $\tilde{X} = (\tilde{X}_t)_{t\in\mathbb{N}}$ be a Markov chain defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with probability transition P. We assume that $\tilde{X}_t$ takes values in a Polish state space $\mathcal{S}$. Without loss of generality, we assume that $\tilde{X}_0 = x_0$ where $x_0 \in \mathcal{S}$ is a deterministic initial condition. The generalization to a random initial condition $\tilde{X}_0$ is straightforward.

The path space is denoted by

$$\mathcal{P} = \{x = (x_t)_{t\in\mathbb{N}} : x_t \in \mathcal{S} \text{ for all } t \in \mathbb{N}\}. \tag{6}$$

It is well-known that, by introducing the distance $d_{\mathcal{P}}(x, y) = \sum_{t\in\mathbb{N}} \frac{1}{2^t}\left(1 \wedge \sup_{s\leq t} d_{\mathcal{S}}(x_s, y_s)\right)$ (which is a metric for the product topology), the space $(\mathcal{P}, d_{\mathcal{P}})$ is again complete and separable. We denote by $\mathcal{B}(\mathcal{P})$ the corresponding Borel σ-field. We thus see $\tilde{X}$ as a random variable with values in $\mathcal{P}$.

The set of paths $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ is endowed with the natural filtration in time $(\mathcal{B}_t)_{t\in\mathbb{N}}$: $\mathcal{B}_t \subset \mathcal{B}(\mathcal{P})$ is the smallest σ-field such that $(x_s)_{s\in\mathbb{N}} \in \mathcal{P} \mapsto (x_1, \ldots, x_t) \in \mathcal{S}^t$ is measurable. The natural filtration for a given Markov chain $\tilde{X}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathcal{S}$ is then given by the pullback of the filtration $(\mathcal{B}_t)_{t\in\mathbb{N}}$ by $\tilde{X} : (\Omega, \mathcal{F}) \to (\mathcal{P}, \mathcal{B}(\mathcal{P}))$.

2.2 The rare event of interest

Given two disjoint Borel subsets A and B of $\mathcal{S}$, our main objective is the efficient sampling of events such as $\{\tau_B < \tau_A\}$ where

$$\tau_A = \inf\left\{t \in \mathbb{N} : \tilde{X}_t \in A\right\} \qquad \text{and} \qquad \tau_B = \inf\left\{t \in \mathbb{N} : \tilde{X}_t \in B\right\}$$

are respectively the first entrance times in A and B. Both τ_A and τ_B are stopping times with respect to the natural filtration of the process $\tilde{X}$.

We are mainly interested in the estimation of the probability $\mathbb{P}(\tau_B < \tau_A)$ in the rare event regime, namely when this probability is very small (typically less than 10^{-8}). This occurs for example if the initial condition $x_0 \in A^c \cap B^c$ is such that x_0 is close to A, and A and B are metastable regions for the dynamics. The Markov chain starting from (a neighborhood of) A (resp. B) remains for a very long time near A (resp. B) before exiting, and thus, the Markov chain starting from x_0 reaches A before B with a probability close to one. Specific examples will be given in Section 5.

Let us introduce the Markov chain stopped at time τ_A: $X = (X_t)_{t\in\mathbb{N}}$ where

$$X_t = \tilde{X}_{t\wedge\tau_A} \quad \text{for any } t \in \mathbb{N}. \tag{7}$$

The probability distribution of the stopped Markov chain X (seen as a $\mathcal{P}$-valued random variable) is denoted by

$$\pi = \mathrm{Law}\,(X) \in \mathrm{Proba}(\mathcal{P}). \tag{8}$$

The probability of interest can be rewritten in terms of π:

$$\mathbb{P}(\tau_B < \tau_A) = \mathbb{E}_\pi\left(\mathbb{1}_{\mathrm{T}_B(X)<\mathrm{T}_A(X)}\right)$$

where we denote for any path $x \in \mathcal{P}$

$$\mathrm{T}_A(x) = \inf\{t \in \mathbb{N} : x_t \in A\} \qquad \text{and} \qquad \mathrm{T}_B(x) = \inf\{t \in \mathbb{N} : x_t \in B\}\,.$$

More generally, the algorithm allows us to estimate expectations of the following form:

$$\pi(\varphi) = \mathbb{E}\left(\varphi(X)\right), \tag{9}$$

for any observable $\varphi : \mathcal{P} \to \mathbb{R}$ such that $\pi(|\varphi|)$ is finite.

Remark 2.1 (On the stopping times τ_A and τ_B)**.** We defined above the stopping times as first entrance times in some sets A and B. As will become clear below, the definition of the algorithm and the unbiasedness result only require τ_A and τ_B to be stopping times with respect to the natural filtration of the chain $\tilde{X}$ (*i.e.* T_A and T_B to be stopping times on $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ endowed with the natural filtration).

2.3 Reaction coordinate

The crucial ingredient we need to introduce the AMS algorithm is an importance function, also known as a reaction coordinate or an order parameter in the context of molecular dynamics. This is a measurable $\mathbb{R}$-valued mapping defined on the state space $\mathcal{S}$:

$$\xi : \mathcal{S} \to \mathbb{R}.$$

The choice of a good function ξ for given sets A and B is a difficult problem in general. One of the main aims of this paper is to show that whatever the choice of ξ, it is possible to define an unbiased estimator of (9). The only requirement we impose on ξ is that there exists a constant $z_{\max} \in \mathbb{R}$ such that

$$B \subset \xi^{-1}\left(]z_{\max}, +\infty[\right). \tag{10}$$

In what follows, the values of ξ are called levels and we will very often refer to the maximum level of a path, defined as follows:

Definition 2.2. *For any path $x \in \mathcal{P}$, the* maximum level *of x is defined as the supremum of ξ along the path x stopped at $\mathrm{T}_A(x)$:*

$$\Xi(x) = \sup\{\xi(x_{t\wedge \mathrm{T}_A(x)})_{t\in\mathbb{N}}\} \in \mathbb{R} \cup \{+\infty\}. \tag{11}$$

The function Ξ can be seen as a reaction coordinate on the path space $\mathcal{P}$.

We also introduce for any level $z \in \mathbb{R}$ and any path $x \in \mathcal{P}$

$$\mathrm{T}_z(x) = \inf\{t \in \{0, \dots, \mathrm{T}_A(x)\} : \xi(x_t) > z\}, \tag{12}$$

which is the first entrance time of the path x stopped at $\mathrm{T}_A(x)$ in the set $\xi^{-1}(]z, +\infty[)$. We emphasize on the strict inequality in the above definition of the entrance times $\mathrm{T}_z(x)$: it is one of the important ingredients of the proof of the unbiasedness of the estimator of (9). Notice that the above assumption (10) on B is equivalent to the inequality

$$\forall x \in \mathcal{P},\ \mathrm{T}_{z_{\max}}(x) \leq \inf\{t \in \{0, \dots, \mathrm{T}_A(x)\} : x_t \in B\}.$$

We denote by

$$\tau_z = \mathrm{T}_z(X) = \inf\{t \in \{0, \dots, \tau_A\} : \xi(X_t) > z\}. \tag{13}$$

the entrance time associated with the (stopped) Markov chain X. It is a stopping time for the natural filtration of the Markov chain.

Remark 2.3 (On Assumption (10))**.** Assumption (10) is extremely useful in practice when computing approximations of averages of the form $\mathbb{E}_\pi\left(\varphi(X)\mathbb{1}_{\mathrm{T}_B(X)<\mathrm{T}_A(X)}\right)$: it is possible to remove from memory the replicas which are "declared retired" in the splitting step at each iteration in the AMS algorithm described below, since by construction we know in advance that they will not contribute to the computation of the associated estimator. The algorithm thus only requires to retain a fixed number of replicas, denoted by n_{rep} below.

2.4 Resampling kernel

The AMS algorithm is based on an interacting system of weighted replicas. At each iteration, copies of the Markov chain are simulated (in parallel) and are ranked according to their maximum level. The less fitted trajectories, according to their maximum level, are resampled according to a resampling kernel. For any $z \in \mathbb{R}$ the resampling kernel π_z (which is in fact a transition probability kernel from $\mathbb{R} \times \mathcal{P}$ to $\mathcal{P}$) is denoted by

$$\pi_z : \begin{cases} \mathcal{P} \to \mathrm{Proba}(\mathcal{P}) \\ x \mapsto \pi_z(x, dx') \end{cases} \tag{14}$$

and is defined as follows: for any $x \in \mathcal{P}$, $\pi_z(x, dx')$ is the law of the $\mathcal{P}$-valued random variable Y such that

$$\begin{cases} Y_t = x_t & \text{if } t \leq \mathrm{T}_z(x) \\ \mathrm{Law}(Y_t | Y_s,\, 0 \leq s \leq t-1) = P(Y_{t-1},\, .\,) & \text{if } t > \mathrm{T}_z(x) \end{cases} \tag{15}$$

and is stopped at $\mathrm{T}_A(Y)$ when Y hits A. We recall that P is the transition kernel of the Markov chain X. In other words, for $t \leq \mathrm{T}_z(x)$, Y is identically x, while for $t > \mathrm{T}_z(x)$, Y_t is generated according to the Markov dynamics on $\mathcal{S}$, with probability transition P, and stopped when reaching A. We thus perform a branching of the path x at time $\mathrm{T}_z(x)$ and position $x_{\mathrm{T}_z(x)}$.

Notice that the definition of the resampling kernel implies that if $\Xi(x) \leq z$, then the resampling kernel does not modify x: $\mathrm{T}_z(x) = +\infty$ and $\pi_z(x, dx')$ is a Dirac mass: $Y_t = x_{t \wedge \mathrm{T}_A(x)}$ for any $t \in \mathbb{N}$.

One of the important ingredients to prove the unbiasedness of the estimator of (9) is the following right-continuity property: for all $x \in \mathcal{P}$, and for any continuous bounded test function $\varphi : \mathcal{P} \to \mathbb{R}$, the mapping

$$z \mapsto \int_{\mathcal{P}} \varphi(x') \pi_z(x, dx')$$

is right-continuous, see Section 3.3 for a proof of this statement.

2.5 The AMS algorithm

In this section, we introduce the AMS algorithm in the specific context of sampling of Markov chain trajectories. The associated unbiased estimator of (9) is given in the next section. A generalized adaptive multilevel splitting framework which encompasses the AMS algorithm described here will be provided in Section 3.

Short description of the AMS algorithm In addition to the choice of the reaction coordinate ξ, the two parameters of the algorithm are n_{rep}, the number of replicas, and $k \in \{1, \ldots, n_{\mathrm{rep}} - 1\}$ the (minimum) number of replicas resampled at each step of the algorithm. At the q-th iteration, the replicas of the Markov chain X are denoted by $X^{(n,q)}$ where n is the label of the replica. The algorithm defines a non-decreasing sequence of random levels $Z^{(q)}$. At the beginning of iteration q, the level $Z^{(q)}$ is the k-th order statistics of the maximum levels of the "working" replicas $X^{(n,q)}$ (the notion of "working" replicas is defined below). Then, all replicas with maximum levels lower or equal to $Z^{(q)}$ are declared retired, and resampled in order to keep a fixed number n_{rep} of replicas with maximum level strictly larger than $Z^{(q)}$. As explained above, the resampling procedure consists in duplicating one of the replica such that its maximum level is larger than $Z^{(q)}$ up to the time $\tau_{Z^{(q)}}$, and then in using the resampling kernel, which amounts in completing the trajectory up to time τ_A with the Markov transition kernel P. The set of labels of the n_{rep} replicas obtained at the end of the q-th iteration and which, by construction, all have a maximum level larger than $Z^{(q)}$ is denoted by $I_{\mathrm{on}}^{(q+1)}$. The subscript "on" indicates that these are the replicas which have to be retained to pursue the algorithm (the so-called "working" replicas in the terminology introduced below). Consistently, the subscript "off" refers to the "retired" replicas at a given iteration. The algorithm stops either if the level $z_{\max}$ is reached (namely $Z^{(q)} > z_{\max}$) or if all the replicas at the end of iteration q have maximum levels lower than or equal to the k-th order statistics. The (random) number of iterations is used in the definition of the unbiased estimator of (9).

Full description of the AMS algorithm The AMS algorithm generates iteratively a system of weighted replicas in the state space $\mathcal{P}^{\text{rep}}$, using selection and resampling steps.

In order to define an estimator of $\pi(\varphi)$ for any observable φ, we will need to consider the set $I_{\text{off}}^{(q)}$ of all the labels of the replicas which have been declared retired before iteration q (namely those with a maximum level smaller or equal than $Z^{(q-1)}$), and we will denote by $I^{(q)} = I_{\text{on}}^{(q)} \sqcup I_{\text{off}}^{(q)}$ the set of all the labels of the replicas generated by the algorithm. We recall that $\sqcup$ denotes the disjoint set union. Notice that the cardinal of $I^{(q)}$ is increasing, while $\operatorname{card} I_{\text{on}}^{(q)} = n_{\text{rep}}$ for any q. At step q, the replicas with labels in $I_{\text{on}}^{(q)}$ are referred to as the working replicas, and the replicas with labels in $I_{\text{off}}^{(q)}$ as the retired replicas. The unbiased estimator of (9) associated with the AMS algorithm will be defined in the next section.

We are now in position to introduce the AMS algorithm in full detail (see Figure 1 for a schematic representation).

The initialization step ($q = 0$)

(i) Let $\left(X^{(n,0)}\right)_{1\leq n\leq n_{\text{rep}}}$ be i.i.d. replicas in $\mathcal{P}$ distributed according to π, defined by (8). At this initial stage, all replicas are working replicas *i.e.* $I^{(0)} = I_{\text{on}}^{(0)} = \{1, \ldots, n_{\text{rep}}\}$ and $I_{\text{off}}^{(0)} = \emptyset$.

(ii) Initialize uniformly the weights: $G^{(n,0)} = 1/n_{\text{rep}}$ for $n \in \{1, \ldots, n_{\text{rep}}\}$.

(iii) Compute the order statistics of $\left(\Xi(X^{(n,0)})\right)_{n\in I_{\text{on}}^{(0)}}$: namely a permutation $\Sigma^{(0)}$ of the set of labels $I_{\text{on}}^{(0)} = \{1, \ldots, n_{\text{rep}}\}$ such that

$$\Xi(X^{(\Sigma^{(0)}(1),0)}) \leq \ldots \leq \Xi(X^{(\Sigma^{(0)}(n_{\text{rep}}),0)})$$

and set the initial level as the k-th order statistics[1]:

$$Z^{(0)} = \Xi(X^{(\Sigma^{(0)}(k),0)}).$$

(iv) If $\operatorname{card}\left\{n \in I_{\text{on}}^{(0)} : \Xi(X^{(n,0)}) \leq Z^{(0)}\right\} = n_{\text{rep}}$, then set $Z^{(0)} = +\infty$.

Iterations Iterate on $q \geq 0$, while the following stopping criterion is not satisfied.

The stopping criterion If $Z^{(q)} > z_{\max}$, then the algorithm stops. When it is the case, set $Q_{\text{iter}} = q$. Else perform the following four steps.

The splitting (branching) step

(i) Consider the following partition of the working replicas' labels in $I_{\text{on}}^{(q)}$:

$$I_{\text{on}}^{(q)} = I_{\text{on},\leq Z^{(q)}}^{(q)} \sqcup I_{\text{on},>Z^{(q)}}^{(q)}$$

[1]Notice that $\Sigma^{(0)}$ is not necessarily unique since several replicas may have the same maximum level. Nevertheless, the level $Z^{(0)}$ does not depend on the choice of $\Sigma^{(0)}$. The same remark applies to the definition of the level $Z^{(q+1)}$ at iteration $q \geq 0$, see Remark 2.4.

where replicas with maximum level smaller or equal than $Z^{(q)}$ have labels in

$$I^{(q)}_{\mathrm{on},\leq Z^{(q)}} = \left\{ n \in I^{(q)}_{\mathrm{on}},\, \Xi(X^{(n,q)}) \leq Z^{(q)} \right\},$$

while the set of replicas' labels with maximum level strictly larger than $Z^{(q)}$ is

$$I^{(q)}_{\mathrm{on},> Z^{(q)}} = \left\{ n \in I^{(q)}_{\mathrm{on}},\, \Xi(X^{(n,q)}) > Z^{(q)} \right\}.$$

We set $K^{(q+1)} = \operatorname{card} I^{(q)}_{\mathrm{on},\leq Z^{(q)}} = n_{\mathrm{rep}} - \operatorname{card} I^{(q)}_{\mathrm{on},> Z^{(q)}}$. The set $I^{(q)}_{\mathrm{on}}$ denotes the working replicas at the beginning of iteration q. Among them, the replicas with labels in $I^{(q)}_{\mathrm{on},\leq Z^{(q)}}$ will be declared retired and replaced by branching replicas with labels in $I^{(q)}_{\mathrm{on},> Z^{(q)}}$ using the resampling kernel $\pi_{Z^{(q)}}$, as explained below. Notice that necessarily, $\operatorname{card} I^{(q)}_{\mathrm{on},> Z^{(q)}} \geq 1$ (otherwise the stopping criterion has been fulfilled before entering the splitting step of iteration q, namely $Z^{(q)} = +\infty$).

(ii) Introduce a new set $I^{(q+1)}_{\mathrm{new}} = \left\{ \operatorname{card} I^{(q)} + 1, \ldots, \operatorname{card} I^{(q)} + K^{(q+1)} \right\} \in \mathbb{N}^* \setminus I^{(q)}$ of labels for the new replicas sampled at iteration q.

(iii) Define the children-parent mapping $P^{(q+1)} : I^{(q+1)}_{\mathrm{new}} \to I^{(q)}_{\mathrm{on},> Z^{(q)}}$ as follows: $\left(P^{(q+1)} \left(\operatorname{card} I^{(q)} + \ell \right) \right)_{1 \leq \ell \leq K^{(}}$ are $K^{(q+1)}$ random labels independently and uniformly distributed in $I^{(q)}_{\mathrm{on},> Z^{(q)}}$.

This mapping associates to the label of a new replica the label of its parent. The parent replica (with label in $I^{(q)}_{\mathrm{on},> Z^{(q)}}$) is used in the resampling procedure to create the new replica (with label in $I^{(q+1)}_{\mathrm{new}}$).

(iv) For any $n \in I^{(q)}_{\mathrm{on},> Z^{(q)}}$, the branching number

$$B^{(n,q+1)} = 1 + \operatorname{card} \left\{ n' \in I^{(q+1)}_{\mathrm{new}} \,:\, P^{(q+1)}(n') = n \right\} \tag{16}$$

represents the number of offsprings of $X^{(n,q)}$. The replica $X^{(n,q)}$ will be split into $B^{(n,q+1)}$ replicas: the old one $X^{(n,q)}$ with label $n \in I^{(q)}_{\mathrm{on},> Z^{(q)}}$ and, if $B^{(n,q+1)} > 1$, $B^{(n,q+1)} - 1$ new ones with labels $n' \in I^{(q+1)}_{\mathrm{new}}$ such that $P^{(q+1)}(n') = n$.

(v) The sets of new labels are then updated as follows:

$$I^{(q+1)}_{\mathrm{on}} = I^{(q)}_{\mathrm{on},> Z^{(q)}} \sqcup I^{(q+1)}_{\mathrm{new}}, \quad I^{(q+1)}_{\mathrm{off}} = I^{(q)}_{\mathrm{off}} \sqcup I^{(q)}_{\mathrm{on},\leq Z^{(q)}}, \quad I^{(q+1)} = I^{(q+1)}_{\mathrm{on}} \sqcup I^{(q+1)}_{\mathrm{off}}.$$

Notice that by construction $\operatorname{card} I^{(q+1)}_{\mathrm{on}} = n_{\mathrm{rep}}$.

The weights are updated with the following rule:

$$\begin{cases} G^{(n,q+1)} = G^{(n,q)} & n \in I^{(q+1)}_{\mathrm{off}} \\ G^{(n,q+1)} = \frac{n_{\mathrm{rep}} - K^{(q+1)}}{n_{\mathrm{rep}}} G^{(n,q)} & n \in I^{(q)}_{\mathrm{on},> Z^{(q)}} \\ G^{(n,q+1)} = G^{(P^{(q+1)}(n),q+1)} & n \in I^{(q+1)}_{\mathrm{new}}. \end{cases} \tag{17}$$

Observe that for $n \in I_{\text{on}}^{(q+1)}$,

$$G^{(n,q+1)} = \frac{n_{\text{rep}} - K^{(q+1)}}{n_{\text{rep}}} \frac{n_{\text{rep}} - K^{(q)}}{n_{\text{rep}}} \dots \frac{n_{\text{rep}} - K^{(1)}}{n_{\text{rep}}} \frac{1}{n_{\text{rep}}}.$$

Moreover, the weight of a replica remains constant as soon as it is retired (namely from the first iteration q such that its label is in $I_{\text{off}}^{(q+1)}$).

The resampling step

(i) Replicas in $I^{(q)}$ are not resampled: for $n \in I^{(q)}$, $X^{(n,q+1)} = X^{(n,q)}$.

(ii) For $n' \in I_{\text{new}}^{(q+1)}$, $X^{(n',q+1)}$ is sampled according to the resampling kernel $\pi_{Z^{(q)}}(X^{(P^{(q+1)}(n'),q)}, dx')$ defined in Section 2.4. The new replica $X^{(n',q+1)}$ is thus obtained by branching its parent replica $X^{(P^{(q+1)}(n'),q)}$.

The level computation step Compute the order statistics of $\left(\Xi(X^{(n,q+1)})\right)_{n \in I_{\text{on}}^{(q+1)}}$, namely a bijective mapping $\Sigma^{(q+1)} : \{1, \dots, n_{\text{rep}}\} \to I_{\text{on}}^{(q+1)}$ (we recall that $\operatorname{card} I_{\text{on}}^{(q+1)} = n_{\text{rep}}$) such that

$$\Xi(X^{(\Sigma^{(q+1)}(1),q+1)}) \leq \dots \leq \Xi(X^{(\Sigma^{(q+1)}(n_{\text{rep}}),q+1)})$$

and set the new level as the k-th order statistics:

$$Z^{(q+1)} = \Xi(X^{(\Sigma^{(q+1)}(k),q+1)}). \tag{18}$$

If $\operatorname{card}\left\{n \in I_{\text{on}}^{(q+1)} : \Xi(X^{(n,q+1)}) \leq Z^{(q+1)}\right\} = n_{\text{rep}}$ then set $Z^{(q+1)} = +\infty$.

Increment Increment $q \leftarrow q+1$, and go back to the stopping criterion step.

Notice that Q_{iter} is such that

$$Q_{\text{iter}} = \inf\{q \geq 0 : Z^{(q)} > z_{\max}\}.$$

The number of times the loop consisting of the three steps (splitting / resampling / level computation) is performed is exactly Q_{iter}.

If $Z^{(Q_{\text{iter}})} = +\infty$, none of the working replicas at the iteration $Q_{\text{iter}} - 1$ is above the new level $\Xi(X^{(\Sigma^{(Q_{\text{iter}})}(k),Q_{\text{iter}})})$ and thus, all of them would have been declared retired at the iteration Q_{iter}: this situation is referred to as extinction.

Remark 2.4 (On the number of resampled replicas)**.** It is very important to notice that the number of resampled replicas is at least k, but may be larger than k. In other words, at iteration q, with the above notation, $K^{(q+1)}$ may be larger than k. This requires at least two replicas to have $Z^{(q)}$ as the maximum level at the beginning of iteration q. Actually, it may even happen that, in the level computation step, all the replicas in $I_{\text{on}}^{(q+1)}$ have $Z^{(q+1)}$ as the maximum level, which implies extinction: $\operatorname{card}\left\{n \in I_{\text{on}}^{(q+1)} : \Xi(X^{n,q+1}) \leq Z^{(q+1)}\right\} = n_{\text{rep}}$, $Z^{(q+1)} = +\infty$ and the algorithm stops.

As an example, let us explain a three step procedure which leads two replicas to have the same maximum level, which in addition is the minimum of the maximum levels over all the working replicas, in the case $k = 1$.

Assume that in the splitting and resampling steps, the three following events occur (see Figure 2 for a schematic representation):

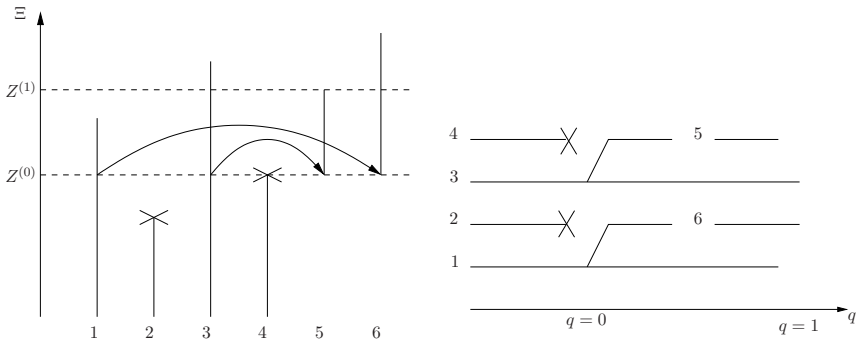

Figure 1: Schematic representation of the first iteration of the AMS algorithm, with $n_{\text{rep}} = 4$ and $k = 2$. The replicas numbered 2 and 4 are declared retired at the first iteration, and are replaced by the replicas with label 5 and 6, which are respectively resampled from the replicas with labels 3 and 1.

1. One of the selected replica (referred to as X) has the smallest maximum level among all the others in $I^{(q)}_{\text{on},>Z^{(q)}}$.

2. The first time $\mathrm{T}_{Z^{(q)}}(X)$ where this replica goes beyond the current level $Z^{(q)}$ also corresponds to the time at which this replica reaches its maximum level:

$$\xi\left(X_{\mathrm{T}_{Z^{(q)}}(X)}\right) = \Xi(X).$$

3. By the resampling procedure, the new replica (referred to as Y) which is generated (starting from $X_{\mathrm{T}_{Z^{(q)}}(X)}$ at time $\mathrm{T}_{Z^{(q)}}(X)$, see the resampling kernel (14)–(15)) is such that $\xi(Y_k) \leq \xi(X_{\mathrm{T}_{Z^{(q)}}(X)}) = \Xi(X)$ for all $k > \mathrm{T}_{Z^{(q)}}(X)$. Thus the new replica Y has the same maximum level as the selected replica X: $\Xi(Y) = \Xi(X)$.

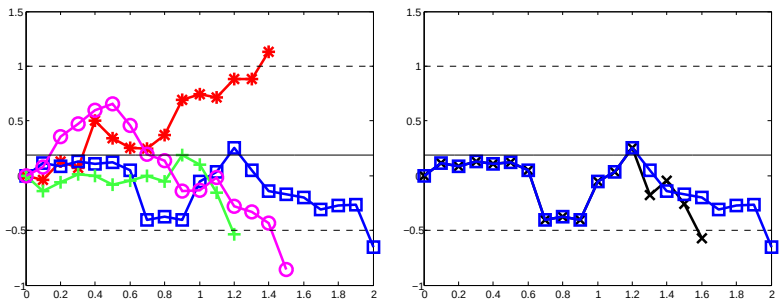

Figure 2: Schematic representation of a procedure leading to the equality of the maximum levels (y-axis) of two replicas. We represent the evolution of the level of each replica at discrete times (x-axis). Left: the current level $Z^{(q)}$ is the maximum level of the green-crosses replica. The blue-squares replica has been selected to be resampled. Right: action of the resampling, where the new black-crosses replica has the same maximum level as the selected replica. Here $A = \{x : \xi(x) < -0.5\}$ and $B = \{x : \xi(x) > 1\}$, and $z_{\max} = 1$.

Then, at the next iteration, one obtains two replicas which have the same maximum level, which is the minimum of the maximum levels over all working replicas. As a consequence, both replicas will be resampled at the next iteration of the algorithm, even if $k = 1$.

By iterating the procedure, one can thus obtain many replicas having this same maximum level, or even all replicas having the same maximum level (which leads to extinction). Other similar procedures can lead to the equality of the maximum levels of two (or more) different replicas. For instance, even if the selected replica in the first step of the procedure above does not have the smallest maximum level among others (the first step is thus skipped), the two next steps will still create two replicas with the same maximum level. This implies that in some next iteration of the algorithm more than k replicas will be declared retired and resampled.

The three events above have small probabilities especially if n_{rep} is large, or if the time-step size is small if one thinks of the Markov Chain as the time discretization of a continuous time diffusion process (such as (1)). But in practice, over many iterations and many independent runs, such situations are actually observed and must be taken into account carefully in the definition of the algorithm. In Section 5.1, we investigate in a simple test case the phenomenon described in this remark, and we illustrate the importance of a proper implementation of the splitting and resampling steps in such situations, in order to obtain unbiased estimators.

2.6 The AMS estimator

For any bounded observable $\varphi : \mathcal{P} \to \mathbb{R}$, a realization of the above algorithm gives an estimator of the average $\pi(\varphi)$ (see (9)) defined by:

$$\hat{\varphi} = \sum_{n \in I^{(Q_{\text{iter}})}} G^{(n,Q_{\text{iter}})} \varphi(X^{(n,Q_{\text{iter}})}). \tag{19}$$

One of the main aim of this paper is to prove that this estimator is unbiased (see Theorem 4.1 below):

$$\mathbb{E}(\hat{\varphi}) = \pi(\varphi).$$

In order to highlight the main features making the estimator unbiased, we will actually prove this result for a larger class of models and algorithms introduced in Section 3.

A particular choice of interest for some applications is $\varphi(x) = \mathbb{1}_{\mathrm{T}_B(x)<\mathrm{T}_A(x)}$, in which case one obtains an unbiased estimator of the probability $p = \mathbb{P}(\tau_B < \tau_A)$. In this case, due to the assumption (10) on B, only replicas with labels in $I_{\text{on}}^{(Q_{\text{iter}})}$ contribute to the estimation, and thus, from one iteration to the other, only replicas with labels in $I_{\text{on}}^{(q)}$ have to be retained, namely a system of n_{rep} replicas. For this specific observable $\varphi(x) = \mathbb{1}_{\mathrm{T}_B(x)<\mathrm{T}_A(x)}$, the estimator of $p = \mathbb{P}(\tau_B < \tau_A)$ is denoted by $\hat{p}$ and is defined from (19) as:

$$\begin{aligned} \hat{p} &= \sum_{n \in I_{\text{on}}^{(Q_{\text{iter}})}} G^{(n,Q_{\text{iter}})} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\text{iter}})})<\mathrm{T}_A(X^{(n,Q_{\text{iter}})})} \\ &= \frac{n_{\text{rep}} - K^{(Q_{\text{iter}})}}{n_{\text{rep}}} \cdots \frac{n_{\text{rep}} - K^{(1)}}{n_{\text{rep}}} P_{\text{corr}} \end{aligned} \tag{20}$$

where in the above the so-called "corrector term" is given by

$$P_{\text{corr}} = \frac{1}{n_{\text{rep}}} \sum_{n \in I_{\text{on}}^{(Q_{\text{iter}})}} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\text{iter}})})<\mathrm{T}_A(X^{(n,Q_{\text{iter}})})} \tag{21}$$

namely the proportion of working replicas that have reached B before A at the final iteration. The properties of this estimator will be numerically investigated in Section 5.

3 Generalized Adaptative Multilevel Splitting

In this section, we introduce a general framework for adaptive multilevel splitting algorithms, which contains in particular the AMS algorithm of Section 2. We refer to this framework as the Generalized Adaptive Multilevel Splitting (GAMS) framework. In particular, we prove in Section 3.3 that the AMS algorithm of Section 2 fits in the GAMS framework. The interest of this abstract presentation is twofold. First, it highlights the essential mathematical properties that are required to produce unbiased estimators of quantities such as (9). As will be proven in Section 4, any algorithm which enters into the GAMS framework yields unbiased estimators of quantities such as (9). Second, it is very useful to propose variants of the classical AMS algorithm which still yield unbiased estimators: we propose some of them in Section 3.5.

The section is organized as follows. In Section 3.1, we introduce in a general setting the quantities we are interested in computing, and the main ingredients we need to state the GAMS framework. In Section 3.2, the GAMS framework is presented. In Section 3.3, we prove that the AMS algorithm introduced in Section 2 for the sampling of paths of Markov chains enters into the GAMS framework. There is a strong analogy between GAMS and Sequential Monte Carlo algorithms (the branching step and the resampling step below correspond respectively to the so-called selection and mutation steps): this is made precise in Section 3.4. Finally, we propose in Section 3.5 some variants of the classical AMS algorithm to illustrate the flexibility of the GAMS framework.

3.1 The general setting

In this section, we introduce the ingredients and the main assumptions we need in order to introduce the GAMS framework. Throughout this section, the notations are consistent with those used in the context of the AMS algorithm.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let us introduce the state space $\mathcal{P}$, which is assumed to be a Polish space and let us denote $\mathcal{B}(\mathcal{P})$ its Borel σ-field. For example, in Section 2, the state space is the path space of Markov chains (see (6)). Let X be a random variable with values in $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ and probability distribution

$$\pi = \mathbb{P} \circ X^{-1} \in \mathrm{Proba}(\mathcal{P}).$$

The aim of the algorithms we present is to estimate

$$\pi(\varphi) = \int_{\mathcal{P}} \varphi(x)\pi(dx) \tag{22}$$

for a given bounded measurable observable $\varphi : \mathcal{P} \to \mathbb{R}$.

The two main ingredients we need in addition to $(\mathcal{P}, \mathcal{B}(\mathcal{P}), \pi)$ is a filtration on $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ (from which we build filtrations on $(\mathcal{P}^{\mathrm{rep}}, \mathcal{B}(\mathcal{P}^{\mathrm{rep}}))$ and on $(\Omega, \mathcal{F}, \mathbb{P})$) and some probability kernels $\pi_z(x, \cdot)$ from $\mathbb{R} \times \mathcal{P}$ to $\mathcal{P}$. Let us introduce them in the next two sections.

3.1.1 The filtrations

In order to define the GAMS framework, we need an additional structure on $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$, namely a filtration indexed by real numbers that we call in the following "levels". Therefore, we assume in the following that the space $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ is endowed with a filtration indexed by levels $z \in \mathbb{R}$

$$(\mathrm{filt}_z)_{z\in\mathbb{R}} \subset \mathcal{B}(\mathcal{P}), \tag{23}$$

namely a non-decreasing family of σ-fields: for any $z < z'$, $\mathrm{filt}_z \subset \mathrm{filt}_{z'} \subset \mathcal{B}(\mathcal{P})$. For example, in the context of Section 2, the filtration is defined as follows: for any $z \in \mathbb{R}$, filt_z is the smallest σ-field on $\mathcal{P}$ which makes the application $x \in \mathcal{P} \mapsto (x_{t\wedge \mathrm{T}_z(x)})_{t\geq 0} \in (\mathcal{P}, \mathcal{B}(\mathcal{P}))$ measurable.

From the filtration given by (23), we construct a filtration $(\mathrm{filt}_z^{\mathrm{rep}})_{z\in\mathbb{R}}$ on the space of replicas $(\mathcal{P}^{\mathrm{rep}}, \mathcal{B}(\mathcal{P}^{\mathrm{rep}}))$ (defined by (5)) by considering the disjoint union of the filtration filt_z on $\mathcal{P}$:

$$\mathrm{filt}_z^{\mathrm{rep}} = \bigsqcup_{I\in\mathcal{I}} \big(\mathrm{filt}_z\big)^{\mathrm{card}\, I}$$

where, we recall, $\mathcal{I}$ denotes the ensemble of finite subsets of $\mathbb{N}^*$.

For any random variable $X : (\Omega, \mathcal{F}) \to (\mathcal{P}, \mathcal{B}(\mathcal{P}))$, we define a filtration $\big(\mathrm{filt}_z^X\big)_{z\in\mathbb{R}}$ on the probability space by pulling-back the filtration $(\mathrm{filt}_z)_{z\in\mathbb{R}}$:

$$\mathrm{filt}_z^X = X^{-1}\big(\mathrm{filt}_z\big). \tag{24}$$

If $\mathcal{X} = \big(X^{(n)}\big)_{n\in I} \in \mathcal{P}^{\mathrm{rep}}$ denotes a random system of replicas – *i.e.* a random variable $\mathcal{X} : (\Omega, \mathcal{F}) \to (\mathcal{P}^{\mathrm{rep}}, \mathcal{B}(\mathcal{P}^{\mathrm{rep}}))$ – we also define the filtration $\big(\mathrm{filt}_z^{\mathcal{X}}\big)_{z\in\mathbb{R}}$ by the same pulling-back procedure:

$$\mathrm{filt}_z^{\mathcal{X}} = \mathcal{X}^{-1}\big(\mathrm{filt}_z^{\mathrm{rep}}\big).$$

By convention, we set $\mathrm{filt}_{-\infty}^{\mathcal{X}} = \sigma(I)$ the σ-field generated by the random set of labels I, and $\mathrm{filt}_{+\infty}^{\mathcal{X}} = \mathcal{F}$. As a consequence for any $z \in \mathbb{R}$ we have $\mathrm{filt}_{-\infty}^{\mathcal{X}} \subset \mathrm{filt}_z^{\mathcal{X}} \subset \mathrm{filt}_{+\infty}^{\mathcal{X}}$.

We finally introduce the notion of stopping level, which is simply a reformulation of the notion of stopping time in our context where the filtrations are indexed by levels instead of times.

Definition 3.1 (Stopping level, Stopped σ-field)**.** *Let $(\mathcal{F}_z)_{z\in\mathbb{R}}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. A stopping level Z with respect to $(\mathcal{F}_z)_{z\in\mathbb{R}}$ is a random variable with values in $\mathbb{R}$ such that $\{Z \leq z\} \in \mathcal{F}_z$ for any $z \in \mathbb{R}$. The stopped σ-field, denoted by $\mathcal{F}_Z$, is characterized as follows:*

$$A \in \mathcal{F}_Z \quad \textit{if and only if} \quad \forall z \in \mathbb{R},\ A \cap \{Z \leq z\} \in \mathcal{F}_z.$$

In particular, Z is a $\mathcal{F}_Z$-measurable random variable.

Remark 3.2 (On the definition of the filtrations)**.** In many cases of practical interest, for any $z \in \mathbb{R}$, filt_z is defined as the smallest filtration which makes an application $F_z : \mathcal{P} \to (\mathcal{E}, \mathcal{B}(\mathcal{E}))$ measurable, for some Polish space $\mathcal{E}$. For example, in the setting of Section 2, $F_z : \mathcal{P} \to \mathcal{P}$ and $F_z(x) = (x_{t\wedge \mathrm{T}_z(x)})_{t\geq 0}$. Then, $\mathrm{filt}_z^{\mathrm{rep}}$ is the smallest filtration which makes the application $G_z : \mathcal{P}^{\mathrm{rep}} \to \mathcal{E}^{\mathrm{rep}}$ measurable with $G_z((X^{(n)})_{n\in I}) = (F_z(X^{(n)}))_{n\in I}$.

3.1.2 The resampling kernels $\pi_z(x,\cdot)$

The second ingredient we need in addition to the filtrations introduced above is a transition probability kernel from $\mathbb{R}\times\mathcal{P}$ to $\mathcal{P}$ ($\mathbb{R}\times\mathcal{P}$ being endowed with the Borel σ-field $\mathcal{B}(\mathbb{R}\times\mathcal{P})$): $(z,x)\in\mathbb{R}\times\mathcal{P}\mapsto\pi_z(x,\cdot)\in\mathrm{Proba}(\mathcal{P})$. By convention, for any $x\in\mathcal{P}$, we set $\pi_{-\infty}(x,\cdot)=\pi$ (which is consistent with Assumption 1 below) and $\pi_{+\infty}(x,\cdot)=\delta_x$. For an explicit example of a resampling kernel in the Markov chain example of Section 2, we refer to Section 2.4.

This kernel is used in the resampling step as a family of transition probabilities from $\mathcal{P}$ to $\mathcal{P}$, indexed by the level z. For a given level $z\in\mathbb{R}$ and a given state $x\in\mathcal{P}$, $\pi_z(x,dx')$ is the probability distribution of the resampling of the state x from level z. In the following, we will refer to this transition probability kernel as a resampling kernel, since it is used in the resampling step.

3.1.3 Assumptions on $(\mathrm{filt}_z^X)_{z\in\mathbb{R}}$ and $(\pi_z)_{z\in\mathbb{R}}$.

We will need two assumptions on $(\mathrm{filt}_z^X)_{z\in\mathbb{R}}$ and $(\pi_z)_{z\in\mathbb{R}}$. First, in the proof of a key auxiliary result below (Lemma 4.5), we use Doob's optional stopping theorem. The required right-continuity property is ensured by the following assumption.

Assumption 1. *For any $x\in\mathcal{P}$, and any continuous bounded test function $\varphi:\mathcal{P}\to\mathbb{R}$,*

$$\begin{cases}\mathbb{R}\to\mathbb{R}\\ z\mapsto\pi_z(\varphi)(x)\end{cases}$$

is right-continuous. Moreover, $\lim_{z\to-\infty}\pi_z=\pi_{-\infty}=\pi$ (the convergence being in distribution).

We use here the notation introduced in (4): $\forall z\in\mathbb{R},\forall x\in\mathcal{P},\pi_z(\varphi)(x)=\int_{y\in\mathcal{P}}\varphi(y)\pi_z(x,dy)$.

Second, we require a consistency relation between the filtration $(\mathrm{filt}_z^X)_{z\in\mathbb{R}}$ and the transition probability kernel $(\pi_z)_{z\in\mathbb{R}}$.

Assumption 2. *Let us consider X, $(\mathrm{filt}_z^X)_{z\in\mathbb{R}}$ and $(\pi_z)_{z\in\mathbb{R}}$ as introduced above. We assume the following consistency relation: if X is distributed according to $\pi_z(x,\cdot)$ for some $(z,x)\in\mathbb{R}\times\mathcal{P}$, then for any $z'\geq z$ and for any bounded measurable test function $\varphi:\mathcal{P}\to\mathbb{R}$,*

$$\mathbb{E}\left(\varphi(X)|\mathrm{filt}_{z'}^X\right)=\pi_{z'}(\varphi)(X)\ a.s.$$

As a consequence (by letting $z\to-\infty$ in the previous assumption), for X distributed according to π, $\pi_{z'}(X,\cdot)$ is a version of the law of X conditionally on $\mathrm{filt}_{z'}^X$ for any $z'\in\mathbb{R}$. Therefore, the σ-field $\mathrm{filt}_{z'}^X$ can be interpreted as containing all the information on a replica X necessary to perform the resampling from X at a given level $z'\in\mathbb{R}$.

Let us finally mention that in addition to these two assumptions and from a more practical point of view, it is also implicitly assumed that it is possible to sample according to the probability measure π (step (ii) of the initialization step below) and according to the probability distribution $\pi_z(x,\cdot)$, for any $x\in\mathcal{P}$ and $z\in\mathbb{R}$ (step (ii) of the resampling step below).

We will check in Section 3.3 that the Markov chain example of Section 2 enters into the general setting introduced in this section.

We are now in position to introduce the GAMS framework in the following section.

3.2 The Generalized Adaptive Multilevel Splitting framework

The aim of this section is to introduce a general framework for splitting algorithms (which we refer to as the Generalized Adaptive Multilevel Splitting (GAMS) framework in the sequel). The structure of the GAMS framework we describe below is quite similar to the one for the AMS algorithm of Section 2. It iterates over three successive steps: (1) the branching or splitting step, (2) the resampling step and (3) the level computation step. These steps are performed until a suitable stopping criterion is satisfied. We denote by Q_{iter} the number of iterations, which in general is a random variable. At each iteration step $q \geq 0$ of the algorithm the distribution π is approximated by an empirical distribution over a system of weighted replicas $\mathcal{X}^{(q)} := (X^{(n,q)}, G^{(n,q)})_{n \in I^{(q)}} \in \mathcal{P}^{\text{rep}}$, where $I^{(q)} \subset \mathbb{N}^*$ is the (random) finite set of labels at step q of the algorithm and $G^{(n,q)} \in \mathbb{R}_+$ is the (random) weight attached to the replica $X^{(n,q)}$.

As will become clear, in order to obtain a fully implementable algorithm from the GAMS framework, three procedures need to be made precise (i) the stopping criterion, (ii) the computation rule of the branching numbers and (iii) the computation of the stopping levels. These procedures require to define three sets of random variables: $(S^{(q)})_{q \geq 0}$, $(B^{(n,q+1)})_{q \geq 0, n \in I^{(q)}})$ and $(Z^{(q)})_{q \geq 0}$, that are used in the GAMS framework presented in the next section 3.2.1. The precise assumptions on these random variables will be stated in Section 3.2.2 (see Assumption 3 below). As already mentioned above, the AMS algorithm of Section 2 corresponds to specific choices of these three items, but this framework allows for many variants (see Section 3.5). The estimator associated with the GAMS framework is finally defined in Section 3.2.3.

3.2.1 Precise definition of the GAMS framework

We now introduce the Generalized Adaptive Multilevel Splitting (GAMS) framework, which is an iterative procedure on an integer index $q \geq 0$.

The initialization step ($q = 0$)

(i) Choose a finite non empty initial set of labels $I^{(0)} = \{1, \ldots, \operatorname{card} I^{(0)}\} \subset \mathbb{N}^*$.

(ii) Let $(X^{(n,0)})_{n \in I^{(0)}}$ be a sequence of $\mathcal{P}$-valued i.i.d. random variables, and distributed according to the probability measure π.

(iii) Initialize uniformly the weights: for any $n \in I^{(0)}$ set $G^{(n,0)} = 1/\operatorname{card} I^{(0)}$.

(iv) Define the system of weighted replicas $\mathcal{X}^{(0)} = (G^{(n,0)}, X^{(n,0)})_{n \in I^{(0)}}$ and for any $z \in \mathbb{R}$, define the σ-field of events $\mathcal{F}_z^{(0)} = \operatorname{filt}_z^{\mathcal{X}^{(0)}}$.

(v) Sample the initial level $Z^{(0)}$ (it is assumed to be a $(\mathcal{F}_z^{(0)})_{z \in \mathbb{R}}$ – stopping level).

(vi) Define the σ-field of events $\mathcal{F}^{(0)} = \mathcal{F}_{Z^{(0)}}^{(0)}$.

Iteration Iterate on $q \geq 0$, while the stopping criterion is not satisfied.

The stopping criterion Sample the random variable $S^{(q)} \in \{0, 1\}$ (which is assumed to be $\mathcal{F}^{(q)}$-measurable). If $S^{(q)} = 0$ then the algorithm stops and we set $Q_{\text{iter}} = q$. Otherwise, if $S^{(q)} = 1$, the three following steps are performed.

The splitting (branching) step

(i) Conditionally on $\mathcal{F}^{(q)}$, sample the $\mathbb{N}$-valued random branching numbers $\left(B^{(n,q+1)}\right)_{n\in I^{(q)}}$ such that for any $n \in I^{(q)}$

$$\mathbb{E}\left(B^{(n,q+1)}|\mathcal{F}^{(q)}\right) > 0 \text{ a.s.}$$

The random variable $B^{(n,q+1)}$ represents the number of offsprings $X^{(n,q)}$. If $B^{(n,q+1)} \geq 1$, the replica $X^{(n,q)}$ will be split into $B^{(n,q+1)}$ replicas: the old one (parent) $X^{(n,q)}$ with label $n \in I^{(q)}$ and, if $B^{(n,q+1)} > 1$, $B^{(n,q+1)} - 1$ new ones (children) that are defined in the resampling step below. If $B^{(n,q+1)} = 0$, the replica is removed from the system. Let us thus introduce the set of labels of such replicas: $I^{(q+1)}_{\text{killed}} = \left\{n \in I^{(q)} : B^{(n,q+1)} = 0\right\}$.

(ii) Compute the total number of new replicas $K^{(q+1)} = \sum_{n\in I^{(q)}} \max\{B^{(n,q+1)} - 1, 0\}$.

(iii) Introduce the set $I^{(q+1)}_{\text{new}} = \left\{\max I^{(q)} + 1, \ldots, \max I^{(q)} + K^{(q+1)}\right\} \subset \mathbb{N}^* \setminus I^{(q)}$ for new labels and update the total set of labels

$$I^{(q+1)} = \left(I^{(q)} \setminus I^{(q+1)}_{\text{killed}}\right) \sqcup I^{(q+1)}_{\text{new}}. \tag{25}$$

(iv) Set a children-parent map $P^{(q+1)} : I^{(q+1)}_{\text{new}} \to I^{(q)} \setminus I^{(q+1)}_{\text{killed}}$ such that for any $n \in I^{(q)} \setminus I^{(q+1)}_{\text{killed}}$ we have

$$\text{card}\left\{n' \in I^{(q+1)}_{\text{new}} : P^{(q+1)}(n') = n\right\} = \max\{B^{(n,q+1)} - 1, 0\}.$$

This map associates to the label of a new replica the label of its parent. The parent replica (with label $n \in I^{(q)} \setminus I^{(q+1)}_{\text{killed}}$) is used in the resampling procedure to create the new replica with label $n' \in I^{(q+1)}_{\text{new}}$, where n and n' are related through the children-parent map by $P^{(q+1)}(n') = n$. Notice that this map is uniquely determined up to a permutation of $I^{(q+1)}_{\text{new}}$. For notational convenience, we extend the map to $I^{(q+1)}$ as follows: $P^{(q+1)}(n) = n$ for any $n \in I^{(q)} \setminus I^{(q+1)}_{\text{killed}}$.

(v) Update the weights as follows: for all $n' \in I^{(q+1)}$ and $n \in I^{(q)} \setminus I^{(q+1)}_{\text{killed}}$ such that $P^{(q+1)}(n') = n$,

$$G^{(n',q+1)} = \frac{G^{(n,q)}}{\mathbb{E}\left(B^{(n,q+1)}|\mathcal{F}^{(q)}\right)}. \tag{26}$$

The resampling step

(i) Replicas in $I^{(q)} \setminus I^{(q+1)}_{\text{killed}}$ are not resampled *i.e.* for any $n \in I^{(q)} \setminus I^{(q+1)}_{\text{killed}}$, $X^{(n,q+1)} = X^{(n,q)}$.

(ii) For $n' \in I^{(q+1)}_{\text{new}}$, $X^{(n',q+1)}$ is sampled according to the resampling kernel $\pi_{Z^{(q)}}(X^{(P^{(q+1)}(n'),q)}, dx)$, *i.e.* by branching its parent replica $X^{(P^{(q+1)}(n'),q)}$.

Then set $\mathcal{X}^{(q+1)} = (X^{(n,q+1)}, G^{(n,q+1)})_{n\in I^{(q+1)}}$.

The level computation step

(i) For any $z \in \mathbb{R}$, define the σ-field of events

$$\mathcal{F}_z^{(q+1)} = \mathcal{F}^{(q)} \vee \sigma(P^{(q+1)}) \vee \mathrm{filt}_z^{\mathcal{X}^{(q+1)}}. \tag{27}$$

The σ-field generated by $P^{(q+1)}$ contains in particular the σ-field generated by $(B^{(n,q+1)})_{n \in I^{(q)}}$.

(ii) Sample the next level $Z^{(q+1)} \in \mathbb{R}$, which is assumed to satisfy:

- $Z^{(q+1)} \geq Z^{(q)}$ a.s.
- $Z^{(q+1)}$ is a stopping level with respect to $\left(\mathcal{F}_z^{(q+1)}\right)_{z \in \mathbb{R}}$.

(iii) Define the σ-field of events $\mathcal{F}^{(q+1)} = \mathcal{F}_{Z^{(q+1)}}^{(q+1)}$.

Increment Increment $q \leftarrow q+1$ and go back to the stopping criterion.

For theoretical purposes, we need in the following to define the system of weighted replicas $(\mathcal{X}^{(q)})_{q \geq 0}$ and the associated filtration $\left(\mathcal{F}^{(q)}\right)_{q \geq 0}$ for all $q \geq 0$ (and not only up to the iteration Q_{iter}). This is simply done by considering the algorithm above with $S^{(q)} = 0$ for all $q \geq 0$.

Remark 3.3 (On the labeling)**.** The way the replicas are labeled is purely conventional.

3.2.2 From the GAMS framework to a practical algorithm

In the GAMS framework, we have defined (see (27)) a family of σ-fields which is indexed both by the level $z \in \mathbb{R}$ and by the iteration index $q \geq 0$ and which is denoted by $\left(\mathcal{F}_z^{(q)}\right)_{q \geq 0, z \in \mathbb{R}}$. By construction, this family of σ-fields satisfies $\mathcal{F}_z^{(q)} \subset \mathcal{F}_{z'}^{(q')}$ if $q < q'$ or $\{q = q' \text{ and } z \leq z'\}$: in other words, the family $\left(\mathcal{F}_z^{(q)}\right)_{q \geq 0, z \in \mathbb{R}}$ is a filtration if $\mathbb{N} \times \mathbb{R}$ is endowed with the lexicographic ordering. At the end of the q-th iteration of the algorithm ($q \geq 0$), one can think of the σ-field $\mathcal{F}^{(q+1)} = \mathcal{F}_{Z^{(q+1)}}^{(q+1)}$ as containing all the necessary information required to perform the next step of the algorithm.

To make a practical splitting algorithm which enters into the GAMS framework, three sets of random variables need to be defined: $(S^{(q)})_{q \geq 0}$, $(B^{(n,q+1)})_{q \geq 0, n \in I^{(q)}})$ and $(Z^{(q)})_{q \geq 0}$. As already stated above, we assume the following on these random variables.

Assumption 3. *The random variables $((S^{(q)})_{q \geq 0}$, $(B^{(n,q+1)})_{q \geq 0, n \in I^{(q)}})$, and $(Z^{(q)})_{q \geq 0})$ satisfy the following properties:*

- *the sequence of random variables $(S^{(q)})_{q \geq 0}$ needed for the stopping criterion, are such that $S^{(q)}$ is with values in $\{0, 1\}$ and is $\mathcal{F}^{(q)}$-measurable;*
- *the sequence of branching numbers $(B^{(n,q+1)})_{q \geq 0, n \in I^{(q)}}$ are with values in $\mathbb{N}$, are assumed to be sampled conditionally on $\mathcal{F}^{(q)}$ (see Section 1.4 for a precise definition) and such that $\mathbb{E}\left(B^{(n,q+1)} | \mathcal{F}^{(q)}\right) > 0$ a.s.;*
- *the sequence $(Z^{(q)})_{q \geq 0}$ of stopping levels are with values in $\mathbb{R}$, satisfy $Z^{(q+1)} \geq Z^{(q)}$ and are such that $Z^{(q)}$ is a stopping level with respect to $\left(\mathcal{F}_z^{(q)}\right)_{z \in \mathbb{R}}$ (see Definition 3.1).*

As explained above, once these three sets of random variables have been defined, the GAMS framework becomes a practical splitting algorithm which yields an unbiased estimator of (22) (this will be proven in Section 4).

Let us emphasize that the requirement that $Z^{(q)}$ is a $(\mathcal{F}_z^{(q)})_{z\in\mathbb{R}}$-stopping level is fundamental to obtain unbiased estimators. It will be instrumental to apply Doob's optimal stopping Theorem for appropriate martingales in the proof of unbiasedness.

As a consequence of the measurability property of $(S^{(q)})_{q\geq 0}$, one easily gets the following property on Q_{iter}:

Proposition 3.4. *The random variable Q_{iter} is a stopping time with respect to the filtration $\left(\mathcal{F}^{(q)}\right)_{q\geq 0}$.*

Remark 3.5 (On the measurability of the system of replicas with respect to $(\mathcal{F}_z^{(q)})_{z\in\mathbb{R}}$)**.** Let us emphasize that for any $q \geq 0$, the system of replicas $(X^{(n,q)})_{n\in I^{(q)}}$ is $\mathcal{F}$-measurable but it is not measurable with respect to $\mathcal{F}^{(q)} = \mathcal{F}^{(q)}_{Z^{(q)}}$ (which indeed stores the information only up to the stopping level $Z^{(q)}$).

3.2.3 The estimator

For any integer $q \geq 0$ and any bounded test function $\varphi : \mathcal{P} \to \mathbb{R}$, we define the estimator

$$\hat{\pi}^{(q)}(\varphi) = \sum_{n\in I^{(q)}} G^{(n,q)}\varphi(X^{(n,q)}) \tag{28}$$

of $\pi(\varphi)$. As will be proven in Section 4, any algorithm which enters into the GAMS framework is such that $\hat{\pi}^{(q)}(\varphi)$ is an unbiased estimator of $\pi(\varphi)$: for any $q \geq 0$, $\mathbb{E}\left(\hat{\pi}^{(q)}(\varphi)\right) = \pi(\varphi)$. Moreover, under appropriate assumptions (see Theorem 4.1), this statement can be generalized when q is replaced by the random number of iterations Q_{iter} of the algorithm:

$$\mathbb{E}\left(\sum_{n\in I^{(Q_{\text{iter}})}} G^{(n,Q_{\text{iter}})}\varphi(X^{(n,Q_{\text{iter}})})\right) = \pi(\varphi).$$

The proof of this result is given in Sections 4.3 and 4.4 and is based on martingale arguments.

3.3 The AMS algorithm enters into the GAMS framework

In this section, we explain how the GAMS framework encompasses the AMS algorithm of Section 2. We thus go back to the setting described there and prove that the modelling and algorithmic assumptions of sections 3.1 and 3.2 are satisfied in this case.

3.3.1 Modelling assumptions

Let us first check that the so-called dynamical setting (namely the sampling of paths of Markov chains) that we considered in Section 2 for the AMS algorithm enters into the general setting of Section 3.1.

In Section 2, $(\mathcal{P}, \mathcal{B}(\mathcal{P}), \pi)$ is the path space for Markov chains, endowed with the standard topology, as explained in Section 2.1. The filtration $(\text{filt}_z)_{z\in\mathbb{R}}$ on $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ is defined by: for

all $z \in \mathbb{R}$, filt_z is the smallest σ-field which makes the application $x \in \mathcal{P} \mapsto x_{t\wedge(\mathrm{T}_z(x))} \in \mathcal{P}$ measurable:

$$\mathrm{filt}_z = \sigma\big(x \mapsto (x_{t\wedge(\mathrm{T}_z(x))})_{t\geq 0}\big). \tag{29}$$

Finally, for any $z \in \mathbb{R}$ and $x \in \mathcal{P}$, the resampling kernel $\pi_z(x, \cdot)$ is defined by (14)–(15).

Let us now check that Assumptions 1 and 2 are satisfied. The conditions of Assumption 2 are consequences of the strong Markov property applied to the chain $t \mapsto X_t \in S$ defined by (7) at the stopping time τ_z (the strong Markov property always holds true for discrete-time Markov processes).

The right-continuity property of Assumption 1 crucially relies on the definition (12) of $\mathrm{T}_z(x)$ as the entrance time of the path $t \mapsto x_t$ in the level set $\xi^{-1}\big]z, +\infty\big[$: the fact that $]z, \infty[$ is an open set implies $z \mapsto \mathrm{T}_z(x)$ is right continuous. More precisely, we have the following Lemma.

Lemma 3.6. *Assumption 1 is satisfied for the resampling kernel defined by* (14)-(15)*. More precisely, for any $x \in \mathcal{P}$, the resampling kernel $z \in \mathbb{R} \mapsto \pi_z(x, .) \in \mathrm{Proba}(\mathcal{P})$ is piecewise constant and right continuous, where $\mathrm{Proba}(\mathcal{P})$ is endowed with standard convergence in distribution.*

Proof. First, assume that $\mathrm{T}_z(x) = +\infty$, which means that $\Xi(x) \leq z$. Then, for any $\varepsilon \geq 0$ we still have $\mathrm{T}_{z+\varepsilon}(x) = +\infty$. In that case $\pi_z(x, .)$ is a Dirac mass: $\pi_z(x, .) = \pi_{z+\varepsilon}(x, .) = \delta_{(x_{t\wedge \mathrm{T}_A(x)})_{t\geq 0}}$.

Now, assume that $\mathrm{T}_z(x) < +\infty$. Then, for $\varepsilon \in]0, \xi(x_{\mathrm{T}_z(x)}) - z[$, $\mathrm{T}_z(x) = \mathrm{T}_{z+\varepsilon}(x)$, and by the definition of the resampling kernel, $\pi_z(x, .) = \pi_{z+\varepsilon}(x, .)$. □

3.3.2 Algorithmic assumptions

As explained in Section 3.2, to obtain a practical splitting algorithm which enters into the GAMS framework, three procedures need to be made precise: the stopping criterion, the computation rule of the branching numbers and the computation of the stopping levels. These procedures should satisfy the measurability requirements of Assumption 3.

The stopping criterion In the AMS algorithm, we set $S^{(q)} = \mathbb{1}_{Z^{(q)} > z_{\max}}$ which is indeed a $\mathcal{F}^{(q)}$-measurable random variable, since $Z^{(q)}$ is a $(\mathcal{F}_z^{(q)})_{z\in\mathbb{R}}$-stopping level, see Lemma 3.7 below.

The computation rule of the branching numbers The branching numbers $B^{(n,q+1)}$ are defined in the splitting step (iv) of the AMS algorithm by (16), for $n \in I^{(q)}_{\mathrm{on},>Z^{(q)}}$. We extend the definition for $n \in I^{(q)} \setminus I^{(q)}_{\mathrm{on},>Z^{(q)}}$ by simply setting $B^{(n,q+1)} = 1$. It is then easy to check that they satisfy the requirements of Assumption 3. Notice that in the AMS algorithm, the total number of new replicas $K^{(q+1)} = \sum_{n\in I^{(q)}} \max\{B^{(n,q+1)} - 1, 0\}$ is given by $K^{(q+1)} = \mathrm{card}\, I^{(q)}_{\mathrm{on},\leq Z^{(q)}}$. Moreover, all branching numbers are positive, so that $I^{(q+1)}_{\mathrm{killed}} = \emptyset$. Another particular feature of the AMS algorithm is that the map $P^{(q+1)}$ takes values in the strict subset $I^{(q)}_{\mathrm{on},>Z^{(q)}}$ of $I^{(q)}$.

Let us check that the computation rule (17) for the weights in the AMS algorithm is indeed consistent with the formula (26) given in the GAMS framework. First, for $n \in I^{(q+1)}_{\mathrm{off}} = I^{(q)}_{\mathrm{on},\leq Z^{(q)}} \sqcup I^{(q)}_{\mathrm{off}}$, $B^{(n,q+1)} = 1$, $P^{(q+1)}(n) = n$ and, consistently, $G^{(n,q+1)} = G^{(n,q)}$.

Second, for $n \in I^{(q)}_{\mathrm{on},>Z^{(q)}}$, it is clear that $\mathbb{E}\left(B^{(n,q+1)}|\mathcal{F}^{(q)}\right)$ does not depend on n (since the random variables are exchangeable in $n \in I^{(q)}_{\mathrm{on},>Z^{(q)}}$). In addition, by construction, $\sum_{n' \in I^{(q)}_{\mathrm{on},>Z^{(q)}}} B^{(n',q+1)} = n_{\mathrm{rep}}$. Thus, we have by a simple counting argument: for any $n \in I^{(q)}_{\mathrm{on},>Z^{(q)}}$,

$$\begin{aligned}\mathbb{E}\big(B^{(n,q+1)}|\mathcal{F}^{(q)}\big) &= \frac{1}{\operatorname{card} I^{(q)}_{\mathrm{on},>Z^{(q)}}} \sum_{n' \in I^{(q)}_{\mathrm{on},>Z^{(q)}}} \mathbb{E}\big(B^{(n',q+1)}|\mathcal{F}^{(q)}\big) \\ &= \frac{\mathbb{E}\left(\sum_{n' \in I^{(q)}_{\mathrm{on},>Z^{(q)}}} B^{(n',q+1)}|\mathcal{F}^{(q)}\right)}{\operatorname{card} I^{(q)}_{\mathrm{on},>Z^{(q)}}} = \frac{n_{\mathrm{rep}}}{n_{\mathrm{rep}} - K^{(q+1)}}.\end{aligned}$$

Thus for $n \in I^{(q)}_{\mathrm{on},>Z^{(q)}}$ (and since $P^{(q+1)}(n) = n$) the formula $G^{(n,q+1)} = \frac{n_{\mathrm{rep}} - K^{(q+1)}}{n_{\mathrm{rep}}} G^{(n,q)}$ in (17) for the AMS algorithm is indeed consistent with the updating formula (26) formula for the weights in the GAMS framework.

Third, for $n \in I^{(q+1)}_{\mathrm{new}}$, $G^{(n,q+1)} = G^{(P^{(q+1)}(n),q+1)} = \frac{n_{\mathrm{rep}} - K^{(q+1)}}{n_{\mathrm{rep}}} G^{(P^{(q+1)}(n),q)}$ which is again consistent with the updating formula (26) formula for the weights in the GAMS framework since $\frac{n_{\mathrm{rep}} - K^{(q+1)}}{n_{\mathrm{rep}}} = 1/\mathbb{E}\big(B^{(P^{(q+1)}(n),q+1)}|\mathcal{F}^{(q)}\big)$.

Computation of the stopping levels Let us now check that the requirements on $Z^{(q)}$ in Assumption 3 are satisfied. By definition of $Z^{(q+1)}$ (see the level computation step of the AMS algorithm), it is clear that $Z^{(q+1)} \geq Z^{(q)}$ (actually, the strict inequality $Z^{(q+1)} > Z^{(q)}$ holds). It remains to prove that $Z^{(q)}$ is a stopping level for the filtration $(\mathcal{F}^{(q)}_z)_{z \in \mathbb{R}}$.

We start with an elementary result, which again highlights the importance of the strict inequality $> z$ in the definitions (12) and (13) of $\mathrm{T}_z(x)$ and τ_z.

Lemma 3.7. *Let $X : \Omega \to \mathcal{P}$ be a Markov chain over the state space $\mathcal{S}$ (see Equation (6)). Then the random variable $\Xi(X)$ (where, we recall, the maximum level mapping Ξ is defined by (11)) is a $(\mathrm{filt}^X_z)_{z \in \mathbb{R}}$-stopping level: for any $z \in \mathbb{R}$, $\{\Xi(X) \leq z\} \in \mathrm{filt}^X_z$.*

Proof. On the one hand, we clearly have the equality of subsets of $\mathcal{P}$:

$$\{x \in \mathcal{P} : \Xi(x) \leq z\} = \{x \in \mathcal{P} : \mathrm{T}_z(x) = +\infty\}.$$

On the other hand, $\tau_z = \mathrm{T}_z(X)$ is a filt^X_z-measurable random variable. The result is then a consequence of these two facts. □

We are now in position to prove the last results which is needed for Assumption 3 to hold.

Lemma 3.8. *For any $q \geq 0$, $Z^{(q)}$ is a stopping level with respect to the filtration $(\mathcal{F}^{(q)}_z)_{z \in \mathbb{R}}$: for any $z \in \mathbb{R}$, $\{Z^{(q)} \leq z\} \in \mathcal{F}^{(q)}_z$.*

Proof. Set by convention $Z^{(-1)} = -\infty$ and let us consider $q \geq 0$. Let us introduce the k-th order statistics over the maximum levels at iteration q: $L^{(q+1)} = \Xi(X^{(\Sigma^{(q+1)}(k),q+1)})$. Let us

also introduce $M^{(q+1)} = \max\{\Xi(X^{(n,q+1)}) : n \in I_{\text{on}}^{(q+1)}\}$. By definition of $Z^{(q+1)}$ (see the level computation step of the AMS algorithm),

$$Z^{(q+1)} = L^{(q+1)} 1_{\{L^{(q+1)} < M^{(q+1)}\}} + (+\infty) 1_{\{L^{(q+1)} = M^{(q+1)}\}}.$$

Therefore, for any $z \in \mathbb{R}$, (using the partition $\Omega = \{M^{(q+1)} \leq z\} \sqcup \{M^{(q+1)} > z\}$)

$$\begin{aligned}\{Z^{(q+1)} \leq z\} &= \{L^{(q+1)} \leq z\} \cap \{L^{(q+1)} < M^{(q+1)}\} \\ &= \{L^{(q+1)} < M^{(q+1)} \leq z\} \sqcup \{L^{(q+1)} \leq z < M^{(q+1)}\}.\end{aligned}$$

These events are all in the σ-field $\sigma\left(\left\{\Xi(X^{(n,q+1)}) \leq z\right\}, \left\{\Xi(X^{(n,q+1)}) \leq Z^{(q)}\right\}, n \in I^{(q+1)}\right)$ (in particular, the set of labels $I_{\text{on}}^{(q+1)}$ is measurable with respect to $\left\{\Xi(X^{(n,q+1)}) \leq Z^{(q)}\right\}$). To conclude, note that by construction (level computation step, (i)) and thanks to Lemma 3.7: for any $z \in \mathbb{R}$,

$$\sigma\left(\left\{\Xi(X^{(n,q+1)}) \leq z\right\}, \left\{\Xi(X^{(n,q+1)}) \leq Z^{(q)}\right\}, n \in I^{(q+1)}\right) \subset \text{filt}_{z \vee Z^{(q)}}^{\mathcal{X}^{(q+1)}} \subset \mathcal{F}_z^{(q+1)}.$$

□

3.3.3 Almost sure mass conservation

The classical AMS algorithm satisfies an additional nice property, namely it conserves almost surely the mass in the following sense:

Definition 3.9. *A splitting algorithm which enters into the GAMS framework satisfies the almost sure mass conservation property if*

$$\forall q \geq 0, \sum_{n \in I^{(q)}} G^{(n,q)} = 1 \ a.s. \tag{30}$$

Indeed, using the definition (17) of the weights and in particular the fact that all the weights $(G^{(n,q)})_{n \in I_{\text{on},Z^{(q)}}^{(q)}}$ are the same: for any $q \geq 0$,

$$\begin{aligned}\sum_{n' \in I^{(q+1)}} G^{(n',q+1)} &= \sum_{n' \in I_{\text{off}}^{(q+1)}} G^{(n',q+1)} + \sum_{n' \in I_{\text{on}}^{(q+1)}} G^{(n',q+1)} \\ &= \sum_{n' \in I_{\text{on},\leq Z^{(q)}}^{(q)} \sqcup I_{\text{off}}^{(q)}} G^{(n',q)} + \sum_{n \in I_{\text{on},>Z^{(q)}}^{(q)} \sqcup I_{\text{new}}^{(q+1)}} \frac{n_{\text{rep}} - K^{(q+1)}}{n_{\text{rep}}} G^{(P^{(q+1)}(n),q)} \\ &= \sum_{n' \in I_{\text{on},\leq Z^{(q)}}^{(q)} \sqcup I_{\text{off}}^{(q)}} G^{(n',q)} + \sum_{n \in I_{\text{on},>Z^{(q)}}^{(q)}} G^{(n,q)} = \sum_{n' \in I^{(q)}} G^{(n',q)}.\end{aligned}$$

Thus, since $\sum_{n' \in I^{(q)}} G^{(n',0)} = 1$, by induction on q, (30) is satisfied. This property will be useful in Theorem 4.1 below: it is one of the two sufficient conditions to prove the unbiasedness of the estimator $\hat{\varphi} = \hat{\pi}^{(Q_{\text{iter}})}(\varphi)$ ($\hat{\pi}^{(q)}(\varphi)$ being defined, we recall, by (28)).

Notice that this property is not generally satisfied for any algorithm which enters into the GAMS framework. Actually, it is only true in general on average: by taking $\varphi(x) = 1$ and $Q_{\text{iter}} = q$ in Theorem 4.1 below, one indeed obtains: $\forall q \geq 0$, $\mathbb{E}\left(\sum_{n' \in I^{(q)}} G^{(n',q)}\right) = 1$.

3.4 Reformulation of the AMS algorithm as a Sequential Monte-Carlo method

The aim of this section is to make more explicit the link between the AMS algorithm and a Sequential Monte Carlo (SMC) sampler, for readers who are familiar with SMC methods. For those who are not, this section can be easily skipped.

For a reaction coordinate with discrete values, the AMS algorithm presented in Section 2 can be understood as a sequential importance sampling algorithm, where weights are assigned to replicas, and replicas are then duplicated and killed to compensate for these weights and obtain unbiased estimators (see for example [12] for a nice introduction to SMC methods and [7] for a discussion of the relationship between SMC algorithms and multilevel splitting algorithms).

To highlight the similarity between the AMS algorithm and a SMC sampler, let us assume that the reaction coordinate takes values in the finite set $\{1, 2, \ldots, z_{\max}\}$

$$\xi : \mathcal{S} \to \{1, 2, \ldots, z_{\max}\}.$$

Let us now introduce a new way to label the successive iterations of the algorithm, by using the levels $z \in \{1, 2, \ldots, z_{\max}\}$ rather than the iteration index $q \geq 0$. Notice indeed that for each z, there exists a unique iteration index $q \geq 0$ such that $z \in [Z^{(q-1)}, Z^{(q)})$. Let us then set: for all $z \in \{1, 2, \ldots, z_{\max}\}$ and q such that $z \in [Z^{(q-1)}, Z^{(q)})$,

$$\begin{cases} J^{(z)} = I^{(q)}, \\ Y^{(n,z)} = (X^{(n,q)}_{t \wedge \mathrm{T}_z(X^{(n,q)})}, t \in \mathbb{N}), \quad \forall n \in I^{(q)}, \\ H^{(n,z)} = G^{(n,q)}, \quad \forall n \in I^{(q)}. \end{cases}$$

The random variables $J^{(z)}$, $Y^{(n,z)}$ and $H^{(n,z)}$ are thus respectively the new set of labels, the new system of replicas and the new system of weights, indexed by the levels z rather than the iteration index q. One can then check that the sequence of weighted replicas $(Y^{(n,z)}, H^{(n,z)})_{n \in J^{(z)}}$ is obtained by applying a standard sequential Monte Carlo algorithm which iterates the following two steps:

1. A splitting step, equivalent to the splitting step of Section 2.5: replicas that have reached the z-level set are split and weighted according to the splitting rule which conserves the total number of replicas above z. The weights of replicas that have not reached the z-level set are not modified.

2. A mutation step, where all replicas are resampled independently according to π_z, but with paths stopped at the stopping time T_{z+1} (defined by (12)).

In the SMC algorithm presented above, all the replicas are resampled which is not the case for the classical AMS algorithm. The following lemma is then crucial to reformulate the AMS algorithm as a SMC sampler. We recall that the resampling kernel $\pi_z(x, dx')$ has been defined in Section 2.4. Moreover, the children-parent mapping has been extended to $I^{(q+1)}_{\mathrm{on}}$ by setting $P^{(q+1)}(n) = n$ for $n \in I^{(q+1)}_{\mathrm{on}} \setminus I^{(q+1)}_{\mathrm{new}}$.

Lemma 3.10. *Consider the algorithm AMS introduced in Section 2.5. Assume that in the resampling step,* all *replicas are resampled. More precisely, replace the two items* (i) *and* (ii) *in the resampling step by a single one:*

(i) For all $n \in I_{\text{on}}^{(q+1)}$, $X^{(n,q+1)}$ is sampled with the resampling kernel $\pi_{Z^{(q)}}(X^{(P^{(q+1)}(n),q)}, dx')$.

Then, the probability distribution of the algorithm is unchanged: the random variables $(X^{(n,Q_{\text{iter}})}, G^{(n,Q_{\text{iter}})})_{n \in I^{(Q_{\text{iter}})}}$ have the same law for the modified algorithm as for the original one.

This lemma is easily checked following the same induction argument as in the proof of Proposition 4.3.

The discussion above thus shows that the AMS algorithm can be recast in the framework of sequential sampling. The interpretation of multi-level splitting methods as a sequential sampling method is not new (see *e.g.* [16]). We refer to the classical monographs [12, 11] for respectively applications of Sequential Monte-Carlo methods in Bayesian statistics, and a comprehensive associated mathematical analysis. In particular, from the point view of [11], the Adaptive Multi-level Splitting method for Markov chains (namely the dynamical setting) considered here can be interpreted as a time-dependent Feynman-Kac particle model with hard obstacles where: (i) the time index is given by the discrete levels z, (ii) the particles are paths of the Markov chain stopped at T_z, and (iii) the hard obstacle at level z corresponds to reach A before the z-level set. Note however that strictly speaking, the version presented in the present section slightly differs from the classical presentation of Feynman-Kac particle models in [11] since all the replicas are used in the estimators, including those who have reached A. But this does not change the global picture.

To conclude, let us recall that the construction of unbiased estimators for averages of the form (9) is standard for SMC algorithms. In the SMC language, averages such as (9) are called non-normalized averages, normalized averages being conditional expectations, namely ratios of two such averages. From this point of view, the unbiasedness result of the present work (see Theorem 4.1) is therefore not a surprise. Actually, another strategy of proof of Theorem 4.1 would be to rely on general unbiasedness results for SMC samplers, using the equivalence between AMS and SMC described above for discrete reaction coordinates, and then to extend the result to continuous reaction coordinates by considering a continuous limit of discrete levels.

3.5 Examples of algorithmic variants

In this section, we consider the setting and the AMS algorithm of Section 2, and we propose variants which fit into the Generalized Adaptive Multilevel Splitting framework and may improve the efficiency of the algorithm in several directions (see Sections 3.5.1, 3.5.2 and 3.5.3). In particular, Theorem 4.1 applies to the three examples detailed below. Moreover, we also illustrate the interest of the general setting we have introduced by providing in Section 3.5.4 an example which does not enter into the standard dynamical setting (sampling of paths of Markov chains) and for which the AMS algorithm could be used.

3.5.1 Removing extinction

We first introduce a variant of the AMS algorithm in the Markov chain setting (Section 2), which is designed in order to remove the possibility of equality of levels for two different replicas – this phenomenon is explained in Remark 2.4 for the AMS algorithm. This variant enters into the GAMS framework and thus leads to unbiased estimators. With this variant, exactly k replicas are resampled at each iterations. In particular, extinction of the system of

replicas cannot occur. However, the algorithm requires the use of a rejection procedure for each resampling, which may slow down the simulation. Let us now describe this variant in detail.

Let $z \in \mathbb{R}$ be a level and $x \in \mathcal{P}$. The definition (see (12)) of $\mathrm{T}_z(x)$ remains the same, but the resampling kernel π_z defined by (14)–(15) is modified as follows. Given $x \in \mathcal{P}$, the probability law $\pi_z(x, dx')$ is the distribution of a random variable $Y \in \mathcal{P}$ sampled as follows:

- For $t \leq \mathrm{T}_z(x) - 1$, $Y_t = x_t$.
- When $\mathrm{T}_z(x) < +\infty$, $Y_{\mathrm{T}_z(x)}$ is sampled using the transition kernel $P(x_{\mathrm{T}_z(x)-1}, \,.\,)$ of the Markov chain, conditionally on $\xi(Y_{\mathrm{T}_z(x)}) > z$. This can be done for example using a rejection procedure: a sequence $(\mathcal{Y}_\ell)_{\ell \in \mathbb{N}^*}$ of i.i.d. random variables distributed according to $P(x_{\mathrm{T}_z(x)-1}, \,.\,)$ is sampled, and one considers $Y_{\mathrm{T}_z(x)} = \mathcal{Y}_L$ where $L = \inf\{\ell \in \mathbb{N}^* : \xi(\mathcal{Y}_\ell) > z\}$.
- For $t > \mathrm{T}_z(x)$, the Markov transition kernel P is used to sample the end of the trajectory, up to the stopping time $\mathrm{T}_A(Y)$ where the path is stopped:

$$\mathrm{Law}(Y_t | (Y_s)_{0 \leq s \leq t-1}) = P(Y_{t-1}, \,.\,).$$

The definition of the filtration $(\mathrm{filt}_z)_{z \in \mathbb{R}}$ needs to be adapted in order to check Assumption 2. The filtration filt_z is the smallest σ-field which makes the application $x \in \mathcal{P} \mapsto (\mathrm{T}_z(x), (x_{t \wedge (\mathrm{T}_z(x)-1)})_{t \geq 0}) \in \mathbb{N} \times \mathcal{P}$ measurable:

$$\mathrm{filt}_z = \sigma(x \mapsto (\mathrm{T}_z(x), (x_{t \wedge (\mathrm{T}_z(x)-1)})_{t \geq 0})). \tag{31}$$

Notice that we need $\mathrm{T}_z(x)$ in addition to $(x_{t \wedge (\mathrm{T}_z(x)-1)})_{t \geq 0}$ since we need to know the time at which the chain reaches the level z. In order to check Assumption 2, let us introduce the auxiliary Markov chain with values in $\mathcal{S} \times \{0, 1\}$: $\tilde{X}_t = (X_t, \mathbb{1}_{\xi(X_{t+1}) > z})$. Then Assumption 2 follows from the strong Markov property applied to $\tilde{X}_t$ and the family of stopping times indexed by z defined by $\tilde{\tau}_z = \inf\left\{t \geq 0 : \mathbb{1}_{\xi(X_{t+1}) > z} = 1\right\}$. Indeed, $\mathcal{F}_{\tilde{\tau}_z} = \mathrm{filt}_z^X$, and $\mathbb{E}(\varphi(X) | \mathcal{F}_{\tilde{\tau}_z}) = \pi_z(\varphi)(X)$.

With this modification of the classical AMS algorithm of Section 2, it is easy to check that the event that two replicas have the same maximum level is of probability zero, at least if the natural additional property is satisfied: if Y_1 and Y_2 are generated according to $P(x_{\mathrm{T}_z(x)-1}, \,.\,)$, where $x \in \mathcal{P}$ is such that $\mathrm{T}_z(x) < +\infty$, then $\mathbb{P}(\xi(Y_1) = \xi(Y_2)) = 0$. This additional condition is satisfied in many practical cases, for example if the Markov Chain is defined as the Euler-Maruyama discretization of a Langevin dynamics, see (1).

3.5.2 Randomized level computation

To run the AMS algorithm of Section 2, a sorting procedure of the replicas according to their maximum levels is required. More precisely, at the initialization step, all replicas must be sorted according to their maximum levels; at further iterations, the procedure is faster, since only the new replicas that have been resampled need to be sorted.

It is possible to propose algorithms within the GAMS framework which never require the sorting of the entire system of replicas. The idea is to sample at iteration q a (small) random subset $\mathcal{I}^{(q+1)} \subset I_{\mathrm{on}}^{(q+1)}$. The level $Z^{(q+1)}$ is then defined as the k-th order statistics of maximum levels computed only on the replicas with labels in $\mathcal{I}^{(q+1)}$. Notice that such

algorithms introduce some flexibility in the implementation of the level computation, which may be useful to design efficient parallelization strategies to speed up the computation.

Notice that Assumption 3 on the stopping-levels $(Z^{(q)})_{q\geq 0}$ is then satisfied by slightly modifying the definition of the σ-fields indexed by z in the level computation step as follows:

$$\mathcal{F}_z^{(q+1)} = \mathcal{F}^{(q)} \vee \sigma(P^{(q+1)}) \vee \mathrm{filt}_z^{\mathcal{X}^{(q+1)}} \vee \sigma(\mathcal{I}^{(q+1)}).$$

3.5.3 Additional selection

It is also possible to modify the branching rules so that larger branching numbers are affected to replicas which are in areas which have been identified as important to get an accurate estimate of $\pi(\varphi)$ (in the spirit of a sequential importance sampling algorithm). For instance, in the bi-channel case of Section 5.2, it is possible to enforce a higher probability of branching for replicas which visit the channel which is not sampled sufficiently well. The only requirements to implement these strategies is that the branching numbers are defined in such a way that Assumption 3 is satisfied.

3.5.4 Application to the sampling of a Gaussian bridge

We presented above variants of the AMS algorithm. The GAMS framework also allows for different general setting: splitting algorithms can be used to sample other random variables than paths of Markov chains with levels defined as $\sup\{\xi(X_{t\wedge\tau_A})_{t\geq 0}\}$ for some stopping time τ_A and some reaction coordinate function ξ. Actually, under appropriate assumptions, the following cases also enter into the setting of the GAMS framework: path-dependent reaction coordinates (duration of the path, integral over the path), sampling of continuous time stochastic processes (diffusions, jump processes, branching processes), sampling of non-homogeneous stochastic processes, etc... Let us discuss in this section as an example the sampling of a Gaussian bridge.

Let $\kappa \in \mathbb{N}^*$ be given, and consider the following Gaussian bridge distribution in $\mathcal{P} = \mathbb{R}^\kappa$:

$$\pi(dx_1 \dots dx_\kappa) = \frac{1}{Z_\kappa} \mathrm{e}^{-\frac{1}{2}\left(x_1^2 + (x_1-x_2)^2 + \dots + (x_{\kappa-1}-x_\kappa)^2 + x_\kappa^2\right)} dx_1 \dots dx_\kappa$$

where $Z_\kappa > 0$ is the appropriate normalization constant. This distribution is a discrete version of a Brownian Bridge, and can be interpreted as a Gaussian random walk $(X_1, X_2, \dots, X_{\kappa+1})$ starting from $X_0 = 0$ and conditioned on $\{X_{\kappa+1} = 0\}$.

The definition of the maximum level is $\Xi(x) = \max\{x_i : i \in \{1, \dots, \kappa\}\}$ and we wish to implement the AMS algorithm to compute small probabilities of the form $\mathbb{P}(\Xi(X) > z)$ for some $z > 0$.

For this purpose, let us define $\mathrm{T}_z(x) = \inf\{i \in \{1, \dots, \kappa\} : x_i > z\}$, and consider the filtration $\mathrm{filt}_z = \sigma(x_1, \dots, x_{\mathrm{T}_z(x)})$.

Let us now define the resampling kernels $\pi_z(x, dx')$. For a given $x \in \mathbb{R}^\kappa$ assuming that $\mathrm{T}_z(x) < +\infty$, let us introduce a random variable $X' \in \mathbb{R}^\kappa$ such that $X'_i = x_i$ for $i \in \{1, \dots, \mathrm{T}_z(x)\}$ and $(X'_{\mathrm{T}_z(x)+1}, \dots, X'_\kappa) \sim \mathcal{B}_{\kappa-\mathrm{T}_z(x)}(x_{\mathrm{T}_z(x)}, 0)$ where for each $m \geq 1$, and $x_0, x_{m+1} \in \mathbb{R}$, $\mathcal{B}_m(x_0, x_{m+1})$ denotes the Gaussian bridge distribution

$$\mathcal{B}_m(x_0, x_{m+1}) = \frac{1}{Z_m} \mathrm{e}^{-\frac{1}{2}\left((x_0-x_1)^2 + (x_1-x_2)^2 + \dots + (x_{m-1}-x_m)^2 + (x_m-x_{m+1})^2\right)} dx_1 \dots dx_m.$$

We then define

$$\pi_z(x, dx') = \mathrm{Law}(X').$$

This general setting enters into the GAMS framework, and satisfies in particular Assumptions 1 and 2 above. Assumption 2 is a consequence of the following Lemma, applied to $K = \mathrm{T}_z(X) \wedge (\kappa - 1)$ (using the fact that $\sigma(X_1, \ldots, X_{\mathrm{T}_z(X)}) = \mathrm{filt}_z^X$).

Lemma 3.11. *Let $(X_1, \ldots, X_\kappa) \sim \mathcal{B}_\kappa(x_0, x_{\kappa+1})$ and let K be a stopping time with respect to the natural filtration of $(X_1, \ldots, X_\kappa)$ (i.e. for any $l \in \{1, \ldots, \kappa\}$, $\{K \leq l\} \subset \sigma(X_1, \ldots, X_l)$) such that $K < \kappa$. Then,*

$$\mathrm{Law}((X_{K+1}, \ldots, X_\kappa)|(X_1, \ldots, X_K)) = \mathcal{B}_{\kappa-K}(X_K, x_{\kappa+1}).$$

Proof. First, the lemma is easily checked for K a deterministic integer, using the the formula for conditional densities. Then the result is proven by conditioning on each value of K and using the fact that K is a stopping time. □

This example can be generalized in various ways. First, it is possible to build resampling kernels $\pi_z(x, dx')$ such that only the components x_i such that $x_i > k$ are resampled. Second, the same kind of algorithms can be applied to discrete Gaussian Markov random fields.

4 The unbiasedness theorem

In the present section, the unbiasedness of the replicas weighted empirical distribution is proven. This is the content of Theorem 4.1. We first provide in Section 4.1 a summary of the notation used in the GAMS framework of Section 3. The latter will be helpful to follow the statements and proofs of the present section. The main result is stated in Section 4.2 and the last two sections 4.3 and 4.4 are devoted to the proof of this result.

4.1 Summary of GAMS notation

We follow the algorithmic order of the GAMS framework of Section 3. In the following, $\varphi : \mathcal{P} \to \mathbb{R}$ denotes a bounded test function. We also introduce below a new notation for an intermediate empirical distribution (see (32)).

The initialization step ($q = 0$) The system of weighted replicas is denoted by $\mathcal{X}^{(0)} = (X^{(n,0)}, G^{(n,0)})_{n \in I^{(0)}}$ with constant weights $G^{(n,0)} = 1/\operatorname{card} I^{(0)}$ for $n \in I^{(0)}$. The first level is $Z^{(0)}$ with the associated σ-field $\mathcal{F}^{(0)} = \mathrm{filt}_{Z^{(0)}}^{\mathcal{X}^{(0)}}$.

Iterations Iterate on $q \geq 0$ the following steps.

The stopping criterion If the stopping criterion is satisfied, the algorithm stops at this stage, and we set $q = Q_{\mathrm{iter}}$. At the beginning of iteration q, the weighted empirical distribution estimator (defined by (28)) is:

$$\hat{\pi}^{(q)}(\varphi) = \sum_{n \in I^{(q)}} G^{(n,q)} \varphi(X^{(n,q)}).$$

The splitting (branching) step The random branching numbers are denoted by $(B^{(n,q+1)})_{n\in I^{(q)}}$, the updated set of labels $I^{(q+1)} = (I^{(q)} \setminus I^{(q+1)}_{\text{killed}}) \sqcup I^{(q+1)}_{\text{new}}$, the associated children-parent map $P^{(q+1)} : I^{(q+1)} \to I^{(q)} \setminus I^{(q+1)}_{\text{killed}}$, and the associated new weights $(G^{(n',q+1)})_{n'\in I^{(q+1)}}$. All the latter variables are sampled conditionally on $\mathcal{F}^{(q)}$, and are $\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})$ measurable. The weighted empirical distribution at this stage is denoted by

$$\hat{\pi}^{(q+1/2)}(\varphi) = \sum_{n'\in I^{(q+1)}} G^{(n',q+1)}\varphi(X^{(P^{(q+1)}(n'),q)}). \tag{32}$$

The resampling step The system of weighted replicas after resampling is denoted by $\mathcal{X}^{(q+1)} = (X^{(n',q+1)}, G^{(n',q+1)})_{n'\in I^{(q+1)}}$.

The level computation step The new level is denoted by $Z^{(q+1)}$, the associated σ-field is defined as $\mathcal{F}^{(q+1)} = \mathcal{F}^{(q)} \vee \sigma(P^{(q+1)}) \vee \text{filt}^{\mathcal{X}^{(q+1)}}_{Z^{(q+1)}}$.

As already explained at the end of Section 3.2.1, we will use use in the following the whole sequence $(\mathcal{X}^{(q)})_{q\geq 0}$ of weighted replicas as well as the whole sequence of related filtrations, which are simply obtained by considering the algorithm without stopping criterion.

4.2 Statement of the main result

The main theoretical result of this paper is the following.

Theorem 4.1. *Let $(\mathcal{X}^{(q)})_{0\leq q\leq Q_{\text{iter}}}$ be the sequence of random systems of weighted replicas generated by an algorithm which enters into the GAMS framework of Section 3. In particular, the Assumptions 1 and 2 on the general setting (see Section 3.1) as well as the Assumption 3 on the stopping criterion, branching numbers and level computations (see Section 3.2.2) are supposed to hold.*

Assume moreover that the number of iterations Q_{iter} is almost surely finite (i.e. $\mathbb{P}(Q_{\text{iter}} < +\infty) = 1$) and that one of the following conditions is satisfied:

- *Q_{iter} is bounded from above by a deterministic constant,*
- *or the almost sure mass conservation* (30) *is satisfied.*

Then, for any bounded measurable test function $\varphi : \mathcal{P} \to \mathbb{R}$,

$$\mathbb{E}\left(\hat{\pi}^{(Q_{\text{iter}})}(\varphi)\right) = \pi(\varphi).$$

Notice that a deterministic number of iterations $Q_{\text{iter}} = q_0 \in \mathbb{N}$ satisfy the assumptions[2] of Theorem 4.1 so that in the above setting (namely under Assumptions 1-2-3):

$$\forall q_0 \geq 0,\ \mathbb{E}\left(\hat{\pi}^{(q_0)}(\varphi)\right) = \pi(\varphi).$$

As a corollary of Theorem 4.1 and thanks to the discussion in Section 3.3 which shows that the AMS algorithm of Section 2.5 enters into the GAMS framework, we also obtain that

[2]To obtain $Q_{\text{iter}} = q_0$, one simply has to choose $S^{(q)} = \begin{cases} 0 & \text{if } q < q_0 \\ 1 & \text{if } q \geq q_0 \end{cases}$.

the AMS estimator $\hat{\varphi} = \hat{\pi}^{(Q_{\text{iter}})}(\varphi)$ defined by (19) in Section 2.6 is an unbiased estimator of $\pi(\varphi)$.

The strategy to prove this theorem is to introduce the sequence of random variables

$$M^{(q)}(\varphi) = \mathbb{E}\left(\hat{\pi}^{(q)}(\varphi)|\mathcal{F}^{(q)}\right) \tag{33}$$

for a fixed bounded measurable test function $\varphi : \mathcal{P} \to \mathbb{R}$ and to show that the process $\left(M^{(q)}(\varphi)\right)_{q\in\mathbb{N}}$ indexed by q is a martingale with respect to the filtration $\left(\mathcal{F}^{(q)}\right)_{q\in\mathbb{N}}$. Since, by Proposition 3.4, Q_{iter} is a stopping time for this filtration, Doob's stopping theorem for discrete-time martingales can then be applied to obtain Theorem 4.1. The next two sections are devoted to the proof of Theorem 4.1.

4.3 Proof of Theorem 4.1

The following definition of conditionally independent replicas will be useful in the proof.

Definition 4.2. *Let Z be a random level, $I \subset \mathbb{N}^*$ a finite random set of indices and $\mathcal{F}$ a σ-field of events. We assume that $\sigma(I) \vee \sigma(Z) \subset \mathcal{F}$. We say that the random system of replicas $(X^{(n)})_{n\in I}$ is independently distributed with distribution $(\pi_Z(X^{(n)}, .))_{n\in I}$ conditionally on $\mathcal{F}$, if for any sequence of bounded measurable functions $(\varphi_n)_{n\in I}$ from $\mathcal{P}$ to $\mathbb{R}$, we have*

$$\mathbb{E}\left(\prod_{n\in I}\varphi_n(X^{(n)})|\mathcal{F}\right) = \prod_{n\in I}\pi_Z(\varphi_n)(X^{(n)}).$$

Let us now state two intermediate propositions before proving Theorem 4.1. The first proposition states that, in the sense of Definition 4.2, the set of replicas with indices in $I^{(q)}$ (resp. $I^{(q+1)}$) are $\mathcal{F}^{(q)}$-conditionally independent (resp. $\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})$-conditionally independent) with explicit distributions.

Proposition 4.3. *Let us consider the setting of Theorem 4.1. For any integer $q \geq 0$,*

$(i)_q$ *The replicas $(X^{(n,q)})_{n\in I^{(q)}}$ are independent with distribution $\left(\pi_{Z^{(q)}}(X^{(n,q)}, .)\right)_{n\in I^{(q)}}$ conditionally on $\mathcal{F}^{(q)}$.*

$(ii)_q$ *The replicas $(X^{(n',q+1)})_{n'\in I^{(q+1)}}$ are independent with distribution $\left(\pi_{Z^{(q)}}(X^{(n',q+1)}, .)\right)_{n'\in I^{(q+1)}}$ conditionally on $\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})$.*

The second proposition states intermediate equalities between conditional averages of the empirical distributions, required to obtain the desired martingale property of $\left(M^{(q)}(\varphi)\right)_{q\geq 0}$, and which are easily obtained from Proposition 4.3.

Proposition 4.4. *Let us consider the setting of Theorem 4.1. For any integer $q \geq 0$ and for any bounded measurable test function $\varphi : \mathcal{P} \to \mathbb{R}$,*

$(iii)_q$ $\mathbb{E}\left(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)}\right) = \mathbb{E}\left(\hat{\pi}^{(q)}(\varphi)|\mathcal{F}^{(q)}\right)$.

$(iv)_q$ $\mathbb{E}\left(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})\right) = \mathbb{E}\left(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})\right)$.

The proofs of both Proposition 4.3 and Proposition 4.4 are postponed to Section 4.4. We are now in position to prove Theorem 4.1.

Proof of Theorem 4.1. The proof consists in first proving that $\left(M^{(q)}(\varphi)\right)_{q\geq 0}$ defined by (33) is a $\left(\mathcal{F}^{(q)}\right)_{q\geq 0}$-martingale and then applying the Doob's optional stopping theorem.

Notice that $\mathbb{E}\left(M^{(q+1)}(\varphi)|\mathcal{F}^{(q)}\right) = \mathbb{E}(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)})$ and let us compute the right-hand side. First, from point $(iv)_q$ of Proposition 4.4 and since $\mathcal{F}^{(q)} \subset \mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})$, we get

$$\begin{aligned}\mathbb{E}\left(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)}\right) &= \mathbb{E}\left(\mathbb{E}\left(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})\right)|\mathcal{F}^{(q)}\right) \\ &= \mathbb{E}\left(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)}\right).\end{aligned}$$

Second, from point $(iii)_q$ of Proposition 4.4 we have

$$\mathbb{E}\left(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)}\right) = \mathbb{E}\left(\hat{\pi}^{(q)}(\varphi)|\mathcal{F}^{(q)}\right).$$

We thus have for any $q \geq 0$,

$$\mathbb{E}\left(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)}\right) = \mathbb{E}\left(\hat{\pi}^{(q)}(\varphi)|\mathcal{F}^{(q)}\right) \tag{34}$$

and $\left(M^{(q)}(\varphi)\right)_{q\in\mathbb{N}}$ is therefore a $\left(\mathcal{F}^{(q)}\right)_{q\in\mathbb{N}}$-martingale.

We now focus on stopping the latter martingale at the random index Q_{iter}. By assumption, either the almost sure mass conservation property (30) is satisfied, in which case $\left(M^{(q)}(\varphi)\right)_{q\in\mathbb{N}}$ is a bounded martingale (since $\left|M^{(q)}(\varphi)\right| \leq \|\varphi\|_\infty$), or $Q_{\text{iter}} \leq q_{\max}$ for some deterministic real number $q_{\max} \in \mathbb{R}$. In both cases, we apply the Doob's optional stopping Theorem (see for instance [21], Chapter 7, Section 2, Theorem 1 and Corollaries 1 and 2) to the martingale $\left(M^{(q)}(\varphi)\right)_{q\in\mathbb{N}}$ and with the stopping time Q_{iter} with respect to the filtration $\left(\mathcal{F}^{(q)}\right)_{q\in\mathbb{N}}$. We obtain

$$\mathbb{E}\left(\hat{\pi}^{(Q_{\text{iter}})}(\varphi)\right) = \mathbb{E}\left(M^{(0)}\right) = \pi(\varphi)$$

which concludes the proof of Theorem 4.1. □

4.4 Proofs of Propositions 4.3 and 4.4

Proposition 4.3 requires an additional intermediate result, namely the propagation Lemma 4.5 below. This lemma gives rigorous conditions under which the property on a system of replicas $(X^{(n)})_{n\in I}$ of being independently distributed with distribution $(\pi_Z(X^{(n)}, .))_{n\in I}$ conditionally on $\mathcal{F}$ can be transported from the σ-field $\mathcal{F}$ to a larger σ-field. It is based on Doob's optional stopping theorem for martingales indexed by the level variable z. Notice that it is the only result where the right continuity stated in Assumption 1 is explicitly used.

Lemma 4.5. *Let us assume that Assumptions 1 and 2 hold. Let $Z \in \mathbb{R} \cup \{-\infty, +\infty\}$ be a random level, $\mathcal{G}$ a σ-field, and $I \subset \mathbb{N}^*$ a finite random set of labels. Assume that $\sigma(I) \vee \sigma(Z) \subset \mathcal{G}$. Consider a random system of replicas $\mathcal{X} = (X^{(n)})_{n\in I}$, which is independently distributed with distribution $(\pi_Z(X^{(n)}, .))_{n\in I}$ conditionally on $\mathcal{G}$ (in the sense of Definition 4.2). Set*

$$\forall z \in \mathbb{R},\ \mathcal{G}_z = \mathcal{G} \vee \text{filt}_z^{\mathcal{X}}, \tag{35}$$

and assume that $Z' \in \mathbb{R} \cup \{-\infty, +\infty\}$ is a stopping level for the filtration $(\mathcal{G}_z)_{z\in\mathbb{R}}$ such that, almost surely, $Z' \geq Z$.

Then the replicas $(X^{(n)})_{n\in I}$ are independently distributed with distribution $\left(\pi_{Z'}(X^{(n)}, .)\right)_{n\in I}$ conditionally on $\mathcal{G}_{Z'}$.

Proof of Lemma 4.5. Step 1. The first step consists in proving that for any fixed $z \in \mathbb{R}$, the system of replicas is independently distributed with distribution $\left(\pi_{Z\vee z}(X^{(n)}, \,.\,)\right)_{n\in I}$ conditionally on $\mathcal{G} \vee \mathrm{filt}_z^{\mathcal{X}}$. By a standard monotone class argument, it is sufficient to show that

$$\mathbb{E}\left(\prod_{n\in I} \varphi_n(X^{(n)})\psi_n(X^{(n)})\, Y\right) = \mathbb{E}\left(\prod_{n\in I} \pi_{Z\vee z}(\varphi_n)(X^{(n)})\psi_n(X^{(n)})\, Y\right),$$

where $(\varphi_n)_{n\geq 1}$ ranges over bounded measurable test functions from $\mathcal{P}$ to $\mathbb{R}$, $(\psi_n)_{n\geq 1}$ ranges over filt_z-measurable test functions from $\mathcal{P}$ to $\mathbb{R}$, and Y over bounded $\mathcal{G}$-measurable random variables. Let us denote $\mathcal{I} = \mathbb{E}\left(\prod_{n\in I} \varphi_n(X^{(n)})\psi_n(X^{(n)})\, Y\right)$ the left-hand side. Since Y is $\mathcal{G}$-measurable, by Definition 4.2 of the conditional independence we get that

$$\mathcal{I} = \mathbb{E}\left(\prod_{n\in I} \pi_Z(\varphi_n\psi_n)(X^{(n)})Y\right).$$

The functions $(\psi_n)_{n\geq 1}$ being filt_z-measurable, they are *a fortiori* $\mathrm{filt}_{z\vee z'}$-measurable for any $z' \in \mathbb{R}$. Assumption 2 on the resampling family $(\pi_z)_{z\in\mathbb{R}}$ then yields

$$\pi_{z'}(\varphi_n\psi_n)(x) = \pi_{z'}(\pi_{z'\vee z}(\psi_n\varphi_n))(x) = \pi_{z'}(\psi_n\pi_{z'\vee z}(\varphi_n))(x).$$

As a consequence, using again that the system of replicas $(X^{(n)})_{n\in I}$ is independently distributed with distribution $\left(\pi_Z(X^{(n)}, \,.\,)\right)_{n\in I}$ conditionally on $\mathcal{G}$ and that Y is $\mathcal{G}$-measurable, we get the following identity

$$\mathcal{I} = \mathbb{E}\left(\prod_{n\in I} \pi_Z\left(\psi_n\pi_{Z\vee z}(\varphi_n)\right)(X^{(n)})Y\right) = \mathbb{E}\left(\prod_{n\in I} \pi_{Z\vee z}(\varphi_n)(X^{(n)})\psi_n(X^{(n)})Y\right),$$

and this concludes the first step.

Step 2. We now prove the main claim of this lemma, namely the fact that the replicas $(X^{(n)})_{n\in I}$ are independent with distribution $\left(\pi_{Z'}(X^{(n)}, \,.\,)\right)_{n\in I}$ conditionally on $\mathcal{G}_{Z'}$. Let us first assume that the test functions $(\varphi_n)_{n\in I}$ are continuous from $\mathcal{P}$ to $\mathbb{R}$.

In order to come back to a classical setting to apply Doob's optional stopping theorem, let us introduce a continuous, one-to-one and strictly increasing change of level parametrization $\mathcal{Z} : [0,1] \to \mathbb{R} \cup \{-\infty, +\infty\}$. Let us consider the following stochastic process indexed by $t \in [0,1]$:

$$N_t = \mathbb{E}\left(\prod_{n\in I} \varphi_n(X^{(n)}) \,\middle|\, \mathcal{G}_{\mathcal{Z}(t)}\right).$$

It is a bounded (since I is $\mathcal{G}_{\mathcal{Z}(t)}$-measurable for all t) and thus uniformly integrable martingale with respect to the filtration $(\mathcal{G}_{\mathcal{Z}(t)})_{t\in[0,1]}$. In addition, $N_1 = \mathbb{E}\left(\prod_{n\in I} \varphi_n(X^{(n)})|\mathcal{G}_{+\infty}\right)$ where $\mathcal{G}_{+\infty} = \mathcal{G} \vee \mathrm{filt}_{+\infty}^{\mathcal{X}}$. Thanks to Step 1 above, we get: almost surely, for all $t \in [0,1]$,

$$N_t = \prod_{n\in I} \pi_{Z\vee\mathcal{Z}(t)}(\varphi_n)(X^{(n)}).$$

Therefore, N_t is almost surely a right-continuous bounded martingale from Assumption 1 on $(\pi_z)_{z\in\mathbb{R}}$. By assumption, $T' = \mathcal{Z}^{-1}(Z')$ is a $\left(\mathcal{G}_{\mathcal{Z}(t)}\right)_{t\in[0,1]}$-stopping level, and we can

use a Doob's optional stopping argument for right continuous bounded martingales (see for instance [17, Theorem 3.22]) to get

$$\mathbb{E}\left(N_1|\mathcal{G}_{\mathcal{Z}(T')}\right) = N_{T'}$$

which can be rewritten as (since $Z' \geq Z$ a.s.)

$$\mathbb{E}\left(\prod_{n\in I}\varphi_n(X^{(n)})|\mathcal{G}_{Z'}\right) = \prod_{n\in I}\pi_{Z'}(\varphi_n)(X^{(n)}).$$

This equality actually holds for any sequence of bounded measurable functions $(\varphi_n)_{n\in I}$ since continuous bounded functions are separating. This concludes the proof of Lemma 4.5.

□

Thanks to Lemma 4.5, we can now prove Proposition 4.3.

Proof of Proposition 4.3. We proceed by induction on the iteration index $q \geq 0$. We first prove directly the statement $(i)_q \Rightarrow (ii)_q$ and then $(ii)_q \Rightarrow (i)_{q+1}$ using Lemma 4.5. The initialization step consists in proving $(i)_0$ using Lemma 4.5. In this proof, $(\varphi_n)_{n\geq 1}$ denotes a sequence of bounded measurable test functions from $\mathcal{P}$ to $\mathbb{R}$.

Proof of $(i)_0$. The statement $(i)_0$ reads

$$\mathbb{E}\left(\prod_{n\in I^{(0)}}\varphi_n\left(X^{(n,0)}\right)|\mathcal{F}^{(0)}\right) = \prod_{n\in I^{(0)}}\pi_{Z^{(0)}}(\varphi_n)(X^{(n,0)}),$$

where $\mathcal{F}^{(0)} = \mathrm{filt}^{\mathcal{X}^{(0)}}_{Z^{(0)}}$. This is exactly the result of Lemma 4.5, taking $Z = -\infty$, $Z' = Z^{(0)}$, $\mathcal{G} = \sigma(I^{(0)})$, and recalling that the replicas are initially independent and distributed according to π.

Proof of $(i)_q \Rightarrow (ii)_q$. Assume that $(i)_q$ holds. We rewrite property $(ii)_q$ as follows

$$\mathbb{E}\left(\prod_{n'\in I^{(q+1)}}\varphi_{n'}\left(X^{(n',q+1)}\right)|\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})\right) = \prod_{n'\in I^{(q+1)}}\pi_{Z^{(q)}}(\varphi_{n'})(X^{(n',q+1)}), \tag{36}$$

and we now prove this identity.

Let us recall that in the resampling step, the replicas with labels in $I^{(q)}$ are not resampled and the replicas $(X^{(n',q+1)})_{n'\in I^{(q+1)}_{\mathrm{new}}}$ are sampled in such a way that they are independently distributed with distribution $(\pi_{Z^{(q)}}(X^{(P^{(q+1)}(n'),q)},\,.\,))_{n'\in I^{(q+1)}_{\mathrm{new}}}$ conditionally on $\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})$. Therefore, by definition of the total set of labels $I^{(q+1)}$ given in (25), one obtains

$$\begin{aligned}
&\mathbb{E}\left(\prod_{n'\in I^{(q+1)}}\varphi_{n'}\left(X^{(n',q+1)}\right)|\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})\right)\\
&\quad= \mathbb{E}\left(\prod_{n\in I^{(q)}\setminus I^{(q+1)}_{\mathrm{killed}}}\varphi_n\left(X^{(n,q)}\right)\prod_{n'\in I^{(q+1)}_{\mathrm{new}}}\varphi_{n'}\left(X^{(n',q+1)}\right)|\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})\right)\\
&\quad= \prod_{n'\in I^{(q+1)}_{\mathrm{new}}}\pi_{Z^{(q)}}(\varphi_{n'})(X^{(P^{(q+1)}(n'),q)})\mathbb{E}\left(\prod_{n\in I^{(q)}\setminus I^{(q+1)}_{\mathrm{killed}}}\varphi_n\left(X^{(n,q)}\right)|\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})\right).
\end{aligned}$$

Next, from the induction hypothesis $(i)_q$, the replicas $(X^{(n,q)})_{n\in I^{(q)}}$ are independent with distribution $\left(\pi_{Z^{(q)}}(X^{(n,q)},\,.\,)\right)_{n\in I^{(q)}}$ conditionally on $\mathcal{F}^{(q)}$. Since $P^{(q+1)}$ is sampled conditionally on $\mathcal{F}^{(q)}$, the replicas $(X^{(n,q)})_{n\in I^{(q)}}$ are also independent conditionally on $\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})$, with the same distributions. Therefore (notice that $I^{(q)}$ and $I^{(q+1)}_{\text{killed}}$ are $\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})$-measurable)

$$\begin{aligned}\mathbb{E}\left(\prod_{n\in I^{(q)}\setminus I^{(q+1)}_{\text{killed}}}\varphi_n\left(X^{(n,q)}\right)|\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})\right) &= \prod_{n\in I^{(q)}\setminus I^{(q+1)}_{\text{killed}}}\pi_{Z^{(q)}}(\varphi_n)(X^{(n,q)})\\ &= \prod_{n'\in I^{(q)}\setminus I^{(q+1)}_{\text{killed}}}\pi_{Z^{(q)}}(\varphi_{n'})(X^{(P^{(q+1)}(n'),q)}).\end{aligned}$$

Gathering the results leads to

$$\mathbb{E}\left(\prod_{n'\in I^{(q+1)}}\varphi_{n'}\left(X^{(n',q+1)}\right)|\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})\right) = \prod_{n'\in I^{(q+1)}}\pi_{Z^{(q)}}(\varphi_{n'})(X^{(P^{(q+1)}(n'),q)}).$$

From the resampling step and Assumption 2, the following identity holds:

$$\forall q\geq 0,\ \forall n'\in I^{(q+1)},\qquad \pi_{Z^{(q)}}(X^{(n',q+1)},\,.\,) = \pi_{Z^{(q)}}(X^{(P^{(q+1)}(n'),q)},\,.\,). \tag{37}$$

This concludes the proof of (36).

Proof of $(ii)_q \Rightarrow (i)_{q+1}$. Let us now assume that $(ii)_q$ holds. To prove that $(i)_{q+1}$ holds, it is sufficient to check that

$$\mathbb{E}\left(\prod_{n\in I^{(q+1)}}\varphi_n\left(X^{(n,q+1)}\right)|\mathcal{F}^{(q+1)}\right) = \prod_{n\in I^{(q+1)}}\pi_{Z^{(q+1)}}(\varphi_n)(X^{(n,q+1)}).$$

This is again exactly the result of Lemma 4.5 applied to $\mathcal{X}^{(q+1)}$, taking $Z = Z^{(q)}$, $Z' = Z^{(q+1)}$ and $\mathcal{G} = \mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})$ so that $\mathcal{G}_z = \mathcal{F}^{(q+1)}_z$ (where, we recall, $\mathcal{F}^{(q+1)}_z$ is defined by Equation (27)). □

Finally, let us prove Proposition 4.4.

Proof of Proposition 4.4. The first equality $(iii)_q$ is a direct consequence of the definition of the branching numbers. The second equality $(iv)_q$ is obtained as a consequence of Proposition 4.3 by combining $(i)_q$ and $(ii)_q$.

Proof of $(iii)_q$. The proof of this assertion is a direct application of the branching rule. Indeed, by definition of the weights $G^{(n',q+1)}$ given in (26), by definition of the branching numbers $(B^{(n,q+1)})_{n\in I^{(q)}}$ as the number of offsprings of the n-th replica, and because these branching numbers are independent of $(G^{(n,q)},X^{(n,q)})_{n\in I^{(q)}}$ conditionally on $\mathcal{F}^{(q)}$, we get

$$\begin{aligned}\mathbb{E}\left(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)}\right) &= \mathbb{E}\left(\sum_{n'\in I^{(q+1)}}\frac{G^{(P^{(q+1)}(n'),q)}}{\mathbb{E}\left(B^{(P^{(q+1)}(n'),q+1)}|\mathcal{F}^{(q)}\right)}\varphi(X^{(P^{(q+1)}(n'),q)})|\mathcal{F}^{(q)}\right)\\ &= \mathbb{E}\left(\sum_{n\in I^{(q)}}\frac{G^{(n,q)}}{\mathbb{E}\left(B^{(n,q+1)}|\mathcal{F}^{(q)}\right)}B^{(n,q+1)}\varphi(X^{(n,q)})|\mathcal{F}^{(q)}\right)\\ &= \mathbb{E}\left(\hat{\pi}^{(q)}(\varphi)|\mathcal{F}^{(q)}\right).\end{aligned}$$

Proof of $(i)_q + (ii)_q \Rightarrow (iv)_q$. Using successively $(i)_q$, the identity (37) and $(ii)_q$, we have:

$$\begin{aligned}\mathbb{E}\left(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})\right) &= \sum_{n' \in I^{(q+1)}} G^{(n',q+1)} \pi_{Z^{(q)}}(\varphi)(X^{(P^{(q+1)}(n'),q)}) \\ &= \sum_{n' \in I^{(q+1)}} G^{(n',q+1)} \pi_{Z^{(q)}}(\varphi)(X^{(n',q+1)}) \\ &= \mathbb{E}\left(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})\right)\end{aligned}$$

□

5 Numerical illustration

The aim of this section is to illustrate the behavior of the AMS algorithm as defined in Section 2, in various situations involving discrete-time approximations of overdamped Langevin dynamics (1) in dimension 1 and 2.

We would like to discuss in particular the unbiasedness of the AMS estimator $\hat{p}$ of $p = \mathbb{P}(\tau_B < \tau_A)$ (see the formula (20)) whatever the choice of the reaction coordinate ξ, the number of replicas n_{rep} and the minimal number k of replicas which are declared retired and resampled at each iteration of the AMS algorithm. Indeed, from Theorem 4.1, we know that

$$\forall\ \xi,\ n_{\text{rep}},\ k, \quad \mathbb{E}\,(\hat{p}) = p = \mathbb{P}(\tau_B < \tau_A).$$

In the following, $(\hat{p}_m)_{1 \leqslant m \leqslant N}$ refers to independent realizations of the estimator $\hat{p}$ obtained by N independent runs of the algorithm and the associated empirical mean is denoted by

$$\overline{p}_N = \frac{1}{N} \sum_{m=1}^{N} \hat{p}_m. \tag{38}$$

The variance of the estimator $\hat{p}$ is also investigated numerically, and it is shown in some two-dimensional situations that the variance heavily depends in particular on the choice of the reaction coordinate. In the following, we will denote by

$$\delta_N = 2 \times \frac{1.96}{\sqrt{N}} \times \sqrt{\frac{1}{N} \sum_{m=1}^{N} (\hat{p}_m)^2 - (\overline{p}_N)^2} \tag{39}$$

the size of the 95% empirical confidence interval computed using the empirical variance obtained over N independent runs of the algorithm.

The section is organized as follows. In Section 5.1, we illustrate on a simple one-dimensional test case the importance of a proper implementation of the branching and splitting steps in order to obtain an unbiased estimator of (9). Then, in Sections 5.2 and 5.3, we discuss more precisely on two examples in dimension 2 the efficiency of the AMS algorithm by studying how the convergence of the estimator depends on the parameters ξ, n_{rep} and k. Finally, in Section 5.4, we draw some conclusions and practical recommendations from these numerical experiments.

5.1 One-dimensional example: Brownian-drift dynamics

Let us first consider a one-dimensional example, with a reaction coordinate $\xi : \mathbb{R} \to \mathbb{R}$ which is an increasing function. Of course, in this situation, the AMS algorithm does not depend on ξ. The aim is thus here to show the unbiasedness of the estimator whatever n_{rep} and k. Moreover, we would like to illustrate the fact that incorrect implementations of the branching and splitting steps may lead to strongly biased results.

The model Let $(X_t)_{t\geqslant 0} \in \mathbb{R}$ be a drifted Brownian motion, starting at x_0, with drift $-\mu < 0$ and inverse temperature β: for any $t \geq 0$, $X_t = x_0 - \mu t + \sqrt{2\beta^{-1}} W_t$, where $(W_t)_{t\geqslant 0}$ is a standard Brownian motion. We use the explicit Euler-Maruyama method with a time-step size $\Delta t > 0$: for any $i \in \mathbb{N}$

$$X_{i+1} = X_i - \mu \Delta t + \sqrt{2\beta^{-1}\Delta t}\, G_i, \quad X_0 = x_0,$$

where the random variables $(G_i)_{i\in\mathbb{N}}$ are independent standard Gaussian random variables. Given $a < x_0 < b$, we consider the estimation of

$$p = \mathbb{P}(\tau_b < \tau_a)$$

where $\tau_b \in \mathbb{N}$ and $\tau_a \in \mathbb{N}$ are the first hitting times of $]b, +\infty[$ and $]-\infty, a[$ respectively.

In the sequel, we choose the initial condition $x_0 = 1$, as well as the two barriers $a = 0.1$ and $b = z_{\max} = 1.9$. We choose $\mu = 1$ and we use the following values for the inverse temperature $\beta \in \{8, 24\}$ in order to have a range of estimated probabilities over several orders of magnitude. Moreover the time-step size is $\Delta t = 0.1$. The number of independent runs is $N = 6.10^6$.

Biased algorithms In order to highlight the importance of a proper implementation of the splitting and resampling steps when many replicas have $Z^{(q)}$ as a maximum level, we perform tests with slightly modified versions of the AMS algorithm, which happen to yield biased estimators. Two biased versions are considered.

- Version 1: We first consider the algorithm where the number of resampled replicas is exactly k even if $K^{(q+1)} > k$, namely if more than k replicas have a maximum level smaller than the current level $Z^{(q)}$ of the algorithm.
- Version 2: We then consider a situation where the replicas are possibly resampled from a state with a reaction coordinate equal to the current level $Z^{(q)}$.

More precisely, the first version (Version 1) of a biased algorithm is obtained by modifying the AMS algorithm of Section 2.5 as follows:

(i) At iteration q, the working replicas are ordered according to their maximum level (possibly with equalities) and the k working replicas with smallest maximum level are declared retired, and the others remain working replicas. The level $Z^{(q)}$ is still defined as the maximum level of the k-th newly retired replica. As explained in Remark 2.4, some of the remaining working replicas may have their maximum levels equal to $Z^{(q)}$. Then, as in the classical AMS algorithm, k new replicas are resampled by picking (randomly) an initial condition among the current working replicas with maximum level strictly larger than $Z^{(q)}$ (namely with labels in $I^{(q)}_{\text{on},>Z^{(q)}}$). Replicas with maximum equal to $Z^{(q)}$ are not split. The resampling kernel π_z is the same as in the AMS algorithm, see Section 2.4.

(ii) At the end of the algorithm, the probability p is estimated by

$$\left(\frac{n_{\text{rep}}-k}{n_{\text{rep}}}\right)^{Q_{\text{iter}}}\left(\frac{1}{n_{\text{rep}}}\sum_{n\in I_{\text{on}}^{(Q_{\text{iter}})}}\mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\text{iter}})})<\mathrm{T}_A(X^{(n,Q_{\text{iter}})})}\right), \tag{40}$$

which is consistent with the general updating formula (26) for the weights in the GAMS framework.

The second version (Version 2) of a biased algorithm is obtained by modifying the AMS algorithm of Section 2.5 exactly as for Version 1, except for the resampling step. Item (i) is modified as:

(i-bis) At iteration q, the k working replicas with smallest maximum are declared "retired". Then, k new replicas are resampled by picking (randomly) an initial condition among the states of *all* the current replicas with maximum levels larger than $Z^{(q)}$, including those with maximum levels equal to $Z^{(q)}$. The resampling kernel π_z (defined by (14)–(15)) is modified accordingly: the parent trajectory x is copied up to the time $\tilde{T}_z(x) = \inf\{t \geq 0 : \xi(x_t) \geq z\}$ (with a large inequality) instead of $\mathrm{T}_z(x)$ (which involves a strict inequality, see (12)) and then completed using the Markov dynamics.

These biased versions do not enter into the GAMS framework. Note that for both modified versions and contrary to the classical AMS algorithm, the sequence of levels $Z^{(q)}$ is not necessarily strictly increasing. These modifications will increase the number of iterations of the algorithm, which in view of (40) explains the bias towards lower probability (see results below).

Results In Table 1 and Table 2, estimations with the AMS algorithm are given. We observe that the estimated probability is stable under changes of n_{rep} and k, with a small confidence interval. This yields the reference values $p = 3.60\text{e-}4$ for $\beta = 8$, and $p = 1.2\text{e-}10$ for $\beta = 24$ (where e-n stands for 10^{-n}). For $\beta = 8$, we have checked that these results are in agreement with those obtained by a standard direct Monte-Carlo estimation with 6.10^8 realizations.

In Table 3, estimations with the two biased versions of the AMS algorithm with $k = 1$ are given. Even for $\beta = 8$ (for which the target probability is $p = 3.60\text{e-}4$), the bias is non-negligible. For $\beta = 24$, the probability can be underestimated by a multiplicative factor 100 with Version 1. The bias induced by using incorrect implementations of the AMS algorithm can thus be very large.

Notice that we have chosen a relatively large timestep ($\Delta t = 0.1$) in order to easily extract the bias introduced by inappropriate modifications of the AMS algorithm. When the timestep size Δt goes to 0, the two variants we discussed above get close to the classical AMS algorithm and the bias disappears, since the probability to observe two replicas with the same maximum level goes to zero (see Remark 2.4) and the probability that $\tilde{T}_z(x)$ is different of $\mathrm{T}_z(x)$ for a parent trajectory also goes to zero. Finally, we recall that an unbiased variant where exactly k replicas are resampled at each iteration has been presented in Section 3.5.1.

5.2 The first two-dimensional example: the bi-channel problem

The aim of this two-dimensional example is to investigate the importance of the choice of the reaction coordinate on the efficiency of the algorithm, on a typical example which has been used in previous numerical studies, see [19, 18, 10].

n_{rep}	10	50	50	50	100	200
k	1	1	10	20	1	1
$\overline{p}_N$	3.60e-4	3.596e-4	3.596e-4	3.597e-4	3.597e-4	3.597e-4
δ_N	0.01e-4	0.004e-4	0.004e-4	0.006e-4	0.003e-4	0.002e-4

Table 1: Results obtained on the 1d test case with the classical AMS algorithm and $\beta = 8$.

n_{rep}	10	50	50	50	100	200
k	1	1	10	20	1	1
$\overline{p}_N$	1.20e-10	1.21e-10	1.21e-10	1.21e-10	1.205e-10	1.203e-10
δ_N	0.3e-10	0.03e-10	0.03e-10	0.03e-10	0.01e-10	0.005e-10

Table 2: Results obtained on the 1d test case with the classical AMS algorithm and $\beta = 24$.

5.2.1 The model

We consider the following two-dimensional overdamped Langevin dynamics:

$$dX(t) = -\nabla\mathcal{E}(X(t))dt + \sqrt{2\beta^{-1}}dW_t, \tag{41}$$

where W is a two-dimensional Wiener process and $\beta > 0$ is the inverse temperature.

We use again the Euler-Maruyama method for the time-discretization of the process X. The time step is denoted by Δt. For our numerical simulations we take $\Delta t = 0.05$. For an initial condition $X_0 = x_0 \in \mathbb{R}^2$ and for $m \in \mathbb{N}$, the numerical scheme reads

$$X_{m+1} = X_m - \Delta t\, \nabla\mathcal{E}(X_m) + \sqrt{2\beta^{-1}}\,(W_{(m+1)\Delta t} - W_{m\Delta t}),$$

where $W_{(m+1)\Delta t} - W_{m\Delta t} = \sqrt{\Delta t}\, G_m$ and $(G_m)_{n\in\mathbb{N}}$ are independent Gaussian random variables in $\mathbb{R}^2$ with zero mean and covariance Id.

In the simulations below, the initial condition is $X_0 = x_0 = (-0.9, 0)$. The potential $\mathcal{E} : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\mathcal{E}(x,y) = 0.2x^4 + 0.2\left(y - \frac{1}{3}\right)^4 + 3e^{-x^2-\left(y-\frac{1}{3}\right)^2} - 3e^{-x^2-\left(y-\frac{5}{3}\right)^2} - 5e^{-(x-1)^2-y^2} - 5e^{-(x+1)^2-y^2}.$$

This potential is plotted on Figure 3. This potential has two global minima connected one to another by two channels: the upper channel (which goes through the shallow minimum around $(0, 1.5)$) and the lower channel (which goes through the saddle point around $(0, -0.5)$).

Version	1	2	2	1	2	2
n_{rep}	100	100	10	100	100	10
β	8	8	8	24	24	24
p	3.60e-4	3.60e-4	3.60e-4	1.2e-10	1.2e-10	1.2e-10
$\overline{p}_N$	1.74e-4	3.257e-4	2.96e-4	1.40e-12	6.05e-11	5e-11
δ_N	0.03e-4	0.003e-4	0.01e-4	0.1e-12	0.07e-11	1.5e-11

Table 3: Results obtained on the 1d test case with biased versions of AMS.

The two global minima are close to $m_A = (x_A, y_A) = (-1, 0)$ and $m_B = (x_B, y_B) = (1, 0)$. For some $\rho \in]0, 1[$, we consider the sets A and B defined as the Euclidean open balls of radius ρ around the two minima m_A and m_B, namely

$$\begin{cases} A = \mathcal{B}(m_A, \rho) = \left\{(x, y) \in \mathbb{R}^2 : \sqrt{(x - x_A)^2 + (y - y_A)^2} < \rho\right\} \\ B = \mathcal{B}(m_B, \rho) = \left\{(x, y) \in \mathbb{R}^2 : \sqrt{(x - x_B)^2 + (y - y_B)^2} < \rho\right\}. \end{cases}$$

In the numerical applications, we take $\rho = 0.05$. Most of the trajectories starting from x_0 hit A before B. Moreover, A and B are metastable states: in the small temperature regime, starting from A (resp. B), it takes a lot of time to leave A (resp. B).

We are interested in the estimation of the probability $p = \mathbb{P}(\tau_B < \tau_A)$, where the first hitting times τ_A and τ_B are defined by

$$\tau_A = \inf\{m \in \mathbb{N} : X_m \in A\} \quad \text{and} \quad \tau_B = \inf\{m \in \mathbb{N} : X_m \in B\}.$$

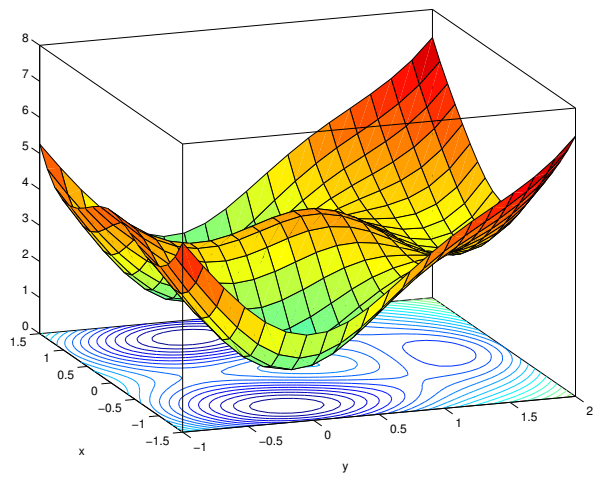

Figure 3: Plot of the potential function for the bi-channel problem.

We will consider the results of the AMS algorithm for the three reaction coordinates ξ^i with $i \in \{1, 2, 3\}$:

1. the *norm to the initial point* m_A*:* $\xi^1(x, y) = \sqrt{(x - x_A)^2 + (y - y_A)^2}$,
2. the *norm to the final point* m_B*:* $\xi^2(x, y) = \xi^1(x_B, y_B) - \sqrt{(x - x_B)^2 + (y - y_B)^2}$,
3. the *abscissa:* $\xi^3(x, y) = x$.

The maximum levels used in the stopping criterion of the algorithm are $z^1_{\max} = z^2_{\max} = 1.9$ and $z^3_{\max} = 0.9$. Notice that for $i \in \{1, 2, 3\}$, we have $B \subset (\xi^i)^{-1}(]z^i_{\max}, +\infty[)$ (see (10)).

In Section 5.2, we take $k = 1$ and the number of replicas is $n_{\text{rep}} = 100$. The values of β belong to the set $\{8.67, 9.33, 10\}$ which are associated with probabilities p ranging approximately from 2e-9 to 1e-10.

5.2.2 Evolution of the empirical mean

Let us first perform simulations with N independent runs of the algorithm, N varying between 1 and 6.10^6. We represent on Figure 4 the evolution as a function of N of the empirical mean $\overline{p}_N$ (defined by (38)) and of the associated 95% confidence intervals $[\overline{p}_n - \delta_N/2, \overline{p}_n + \delta_N/2]$ computed using the empirical variance, see (39).

The colors in the figures are as follows: green (solid line) for ξ^1, red (line with crosses) for ξ^2 and blue (line with circles) for ξ^3. The full lines represent the evolution of the upper and lower bounds of the confidence intervals, while dotted lines represent the evolution of the empirical means.

From these simulations, we observe that:

- When N is sufficiently large, the confidence intervals overlap. This is in agreement with the fact that $\hat{p}$ is an unbiased estimator of p whatever the choice of the reaction coordinate.
- The statistical fluctuations depend a lot on the reaction coordinate. In particular, the results obtained with ξ^1 seem much better than with ξ^2 or ξ^3. We will come back to this in Section 5.2.4.
- The confidence interval being computed empirically, one may conclude that the algorithm is biased by considering the results for N too small (see for example the graphs in the right column in Figure 4). This is due to the fact that the empirical variance dramatically underestimates the real variance if N is too small. This is a well-known phenomenon for splitting algorithms in general called "apparent bias", see [14]. As β gets larger (namely as the temperature gets smaller), the number of independent runs N required to observe overlapping confidence intervals gets larger. For example, for $\beta = 13$, and $N = 6.10^6$, we were not able to get overlapping confidence intervals for the three reaction coordinates: this is actually an indication of the fact that one should pursue computations with larger N in order to get reliable estimates.

We observe that there are some realizations for which the estimator of the probability is very large. These realizations have small probability but they dramatically increase the value of the empirical mean and of the empirical variance. This explains the jumps which are observed on the empirical average and confidence interval as a function of the number of realizations, see Figure 4. In the next section, we illustrate this aspect. As is usually the case with Monte Carlo simulations for rare event simulations, it is impossible to decide *a priori* if the sample size N is sufficiently large to give an accurate estimation.

5.2.3 Heavy tails

In this section, we give a more quantitative interpretation of the above observations on the evolution of the empirical mean. For a given inverse temperature β, and a given reaction coordinate, we sample N independent realizations of the algorithm denoted by $(\hat{p}_m)_{1\leq m\leq N}$.

To illuminate the importance of the largest values of the estimator in the empirical average, we compute partial empirical averages over the largest values or over the smallest values among the N realizations. More precisely, for a fixed N_0, we compute

$$\overline{p}_N^{N_0,\text{large}} = \frac{1}{N_0} \sum_{\ell=N-N_0+1}^{N} \hat{p}_{(\ell)} \text{ and } \overline{p}_N^{N_0,\text{small}} = \frac{1}{N-N_0} \sum_{\ell=1}^{N-N_0} \hat{p}_{(\ell)},$$

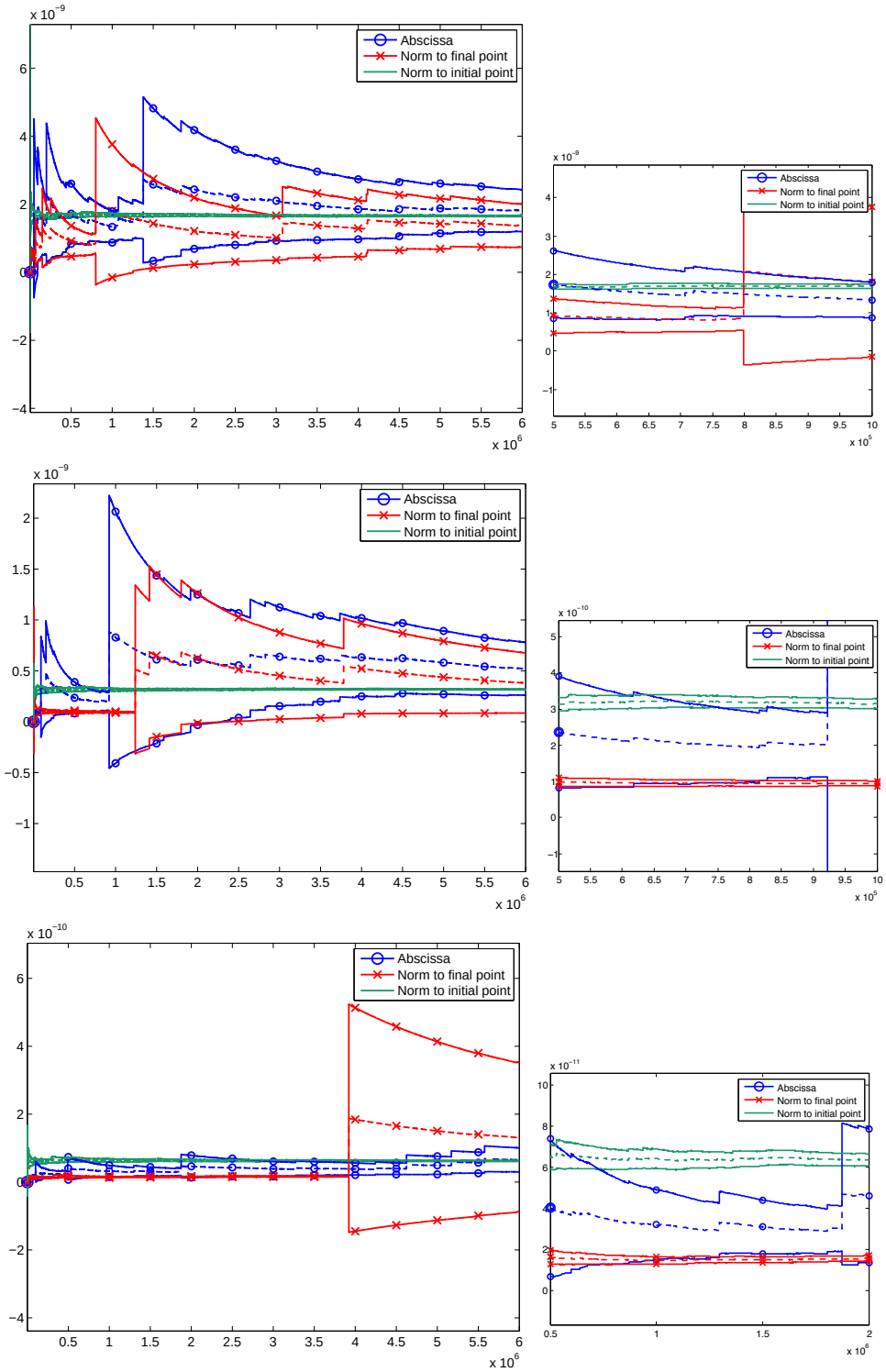

Figure 4: Evolution as a function of N of the empirical mean $\overline{p}_N$ and of the associated 95% confidence intervals $[\overline{p}_n - \delta_N/2, \overline{p}_n + \delta_N/2]$. Upper to lower $\beta = 8.67, 9.33, 10$. The right inserts are zooms of the left graphs on smaller values of N, in order to illustrate the "apparent bias" phenomenon.

where $\left(\hat{p}_{(\ell)}\right)_{1\leq \ell \leq N}$ denotes the order statistics of $(\hat{p}_m)_{1\leq m\leq N}$. In other words, $\overline{p}_N^{N_0,\text{large}}$ is the average over the N_0 largest realizations of $\hat{p}_m$, while $\overline{p}_N^{N_0,\text{small}}$ is the average over all the other values. In particular,

$$\overline{p}_N = \frac{N_0}{N}\overline{p}_N^{N_0,\text{large}} + \left(1 - \frac{N_0}{N}\right)\overline{p}_N^{N_0,\text{small}}.$$

In Table 4, we use $N = 6.10^6$, and $N_0 = 100$ or $N_0 = 1000$. The results are obtained with the same realizations as those used in the previous sections. We observe that the largest values contribute a lot to the empirical average over all the runs whatever the reaction coordinate. This is characteristic of a heavy tailed distribution. Indeed, for $\beta = 10$ for example, the value of $\overline{p}_N^{N_0,\text{small}}$ differs a lot from the value of $\overline{p}_N$. In addition, notice that this difference is more important when using ξ^2 and ξ^3, while when using ξ^1. This is in agreement with the fact that we observed better results (in terms of statistical fluctuations) with ξ^1 than with ξ^2 and ξ^3 in Figure 4. Indeed, distributions with heavier tails typically lead to larger fluctuations.

	β	$\overline{p}_N$	N_0	$\overline{p}_N^{N_0,\text{large}}$	$\overline{p}_N^{N_0,\text{small}}$	N_0	$\overline{p}_N^{N_0,\text{large}}$	$\overline{p}_N^{N_0,\text{small}}$
ξ^1	8.67	1.7e-09	1000	1.3e-e-06	1.5e-09	100	4.0e-06	1.6e-09
ξ^2	8.67	1.3e-09	1000	5.7e-06	4.2e-10	100	5.2e-05	5.1e-10
ξ^3	8.67	1.8e-09	1000	8.0e-06	4.8e-10	100	6.5e-05	7.3e-10
ξ^1	9.33	3.2e-10	1000	3.0e-07	2.7e-10	100	1.0e-06	3.0e-10
ξ^2	9.33	3.8e-10	1000	1.9e-06	6.4e-11	100	1.8e-05	8.2e-11
ξ^3	9.33	5.2e-10	1000	2.7e-06	7.2e-11	100	2.5e-05	1.0e-10
ξ^1	10	6.2e-11	1000	8.9e-08	4.8e-11	100	4.0e-07	5.6e-11
ξ^2	10	1.3e-10	1000	7.0e-07	9.1e-12	100	7.2e-06	1.3e-11
ξ^3	10	6.5e-11	1000	3.0e-07	1.0e-11	100	3.0e-06	1.5e-11

Table 4: Comparisons of the empirical averages $\overline{p}_N$ with partial empirical averages obtained over the N_0 largest and $N - N_0$ smallest values among N realizations.

Remark 5.1 (On heavy tails in the idealized setting)**.** Let us mention another setting where it can be analytically checked that the distribution of the estimator $\hat{p}$ has heavy tails in some regime. In a one-dimensional case (or when the committor function is used as a reaction coordinate), one can show that the AMS algorithm can be related to the so-called exponential case, namely when $\Xi(X)$ is distributed with exponential law with parameter one (see for example [5]). It is then easy to show that for $k = 1$, and in the regime where $n_{\text{rep}} \to +\infty$ and the target probability is $p = \exp(-\sigma^2 n_{\text{rep}})$ for some fixed $\sigma > 0$, then $\frac{\hat{p}}{p}$ converges in distribution to the log-normal distribution $\exp(\sigma Z - \sigma^2/2)$ where $Z \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable (see [2, Proposition 3.4]). In particular, $\hat{p}$ has heavy tails (the median $\exp(-\sigma^2/2)$ of this distribution is smaller than its expectation 1, so that for a large σ the empirical mean under-estimates the expectation).

5.2.4 Fluctuations induced by the two channels

In this section, we compare the results when using two reaction coordinates: ξ^1 (norm to m_A) and ξ^3 (abscissa). Since the typical behavior we observe in Figure 4 and in Table 5 is the same for ξ^2 and ξ^3, we do not repeat the analysis for ξ^2.

As explained above, there are two possible channels for the reactive trajectories going from A to B: the upper channel and the lower channel. For each realization m, one can distinguish the contributions to the estimator $\hat{p}_m$ of the replicas going through the upper channel and the ones going through the lower channel. In the following, for a given path, the trajectory is associated to the upper (resp. lower) channel if the first hitting point of the y-axis is such that $y > 0.5$ (resp. such that $y \leq 0.5$). More precisely, let us define $\Pi_1(x, y) = x$ and $\Pi_2(x, y) = y$ for any $(x, y) \in \mathbb{R}^2$. For a replica $X = (X_t)_{t\in\mathbb{N}}$ such that $\tau = \inf\{t \in \mathbb{N} : \Pi_1(X_t) > 0\} < \infty$, $X \in$ Upper if $\Pi_2(X_\tau) > 0.5$ and $X \in$ Lower if $\Pi_2(X_\tau) \leq 0.5$.

For each run of the algorithm, we compute the three following quantities:

- the number of replicas which reach B before A:

$$\mathrm{M}^B = \sum_{n\in I_{\mathrm{on}}^{(Q_{\mathrm{iter}})}} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\mathrm{iter}})})<\mathrm{T}_A(X^{(n,Q_{\mathrm{iter}})})}$$

- the number of replicas which reach B before A and go through the upper channel:

$$\mathrm{M}^{B,\mathrm{upper}} = \sum_{n\in I_{\mathrm{on}}^{(Q_{\mathrm{iter}})}} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\mathrm{iter}})})<\mathrm{T}_A(X^{(n,Q_{\mathrm{iter}})})} \mathbb{1}_{X^{(n,Q_{\mathrm{iter}})}\in\mathrm{Upper}}$$

- the number of replicas which reach B before A and go through the lower channel:

$$\mathrm{M}^{B,\mathrm{lower}} = \sum_{n\in I_{\mathrm{on}}^{(Q_{\mathrm{iter}})}} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\mathrm{iter}})})<\mathrm{T}_A(X^{(n,Q_{\mathrm{iter}})})} \mathbb{1}_{X^{(n,Q_{\mathrm{iter}})}\in\mathrm{Lower}}$$

Notice that $\mathrm{M}^B = \mathrm{M}^{B,\mathrm{upper}} + \mathrm{M}^{B,\mathrm{lower}}$ and that $\mathrm{M}^B \neq 0$ is equivalent to $\hat{p} \neq 0$. When needed, we explicitly indicate the dependence of these quantities on the realization by a lowerscript m: for $m \in \{1, \ldots, N\}$, we thus denote M_m^B, $\mathrm{M}_m^{B,\mathrm{upper}}$ and $\mathrm{M}_m^{B,\mathrm{lower}}$ the m-th realization of M^B, $\mathrm{M}^{B,\mathrm{upper}}$ and $\mathrm{M}^{B,\mathrm{lower}}$.

Let us introduce the set $\mathcal{E}_N = \{m : \hat{p}_m \neq 0\}$ of realizations which lead to a non zero $\hat{p}$ and the proportion $R_N = \operatorname{card}\mathcal{E}_N/N$ of such realizations. We now divide the realizations in $\mathcal{E}_N$ into three disjoint subsets, with associated proportions.

- All replicas reaching B before A go through the upper channel:

$$\mathcal{E}_N^{\mathrm{upper}} = \left\{m \in \mathcal{E}_N : \mathrm{M}_m^{B,\mathrm{lower}} = 0\right\} \text{ and } \rho_N^{\mathrm{upper}} = \frac{\operatorname{card}\mathcal{E}_N^{\mathrm{upper}}}{\operatorname{card}\mathcal{E}_N}.$$

- All replicas reaching B before A go through the lower channel:

$$\mathcal{E}_N^{\mathrm{lower}} = \left\{m \in \mathcal{E}_N : \mathrm{M}_m^{B,\mathrm{upper}} = 0\right\} \text{ and } \rho_N^{\mathrm{lower}} = \frac{\operatorname{card}\mathcal{E}_N^{\mathrm{lower}}}{\operatorname{card}\mathcal{E}_N}.$$

- Both channels are used by the replicas reaching B before A :

$$\mathcal{E}_N^{\mathrm{mix}} = \mathcal{E}_N \setminus \left(\mathcal{E}_N^{\mathrm{upper}} \cup \mathcal{E}_N^{\mathrm{lower}}\right) \text{ and } \rho_N^{\mathrm{mix}} = \frac{\operatorname{card}\mathcal{E}_N^{\mathrm{mix}}}{\operatorname{card}\mathcal{E}_N}.$$

Obviously, $\rho_N^{\text{mix}} = 1 - \rho_N^{\text{upper}} - \rho_N^{\text{lower}}$. Finally, we define conditional estimators for $\hat{p}$ associated with the partition of $\mathcal{E}_N$ defined above:

$$\tilde{p}_N^{\text{upper}} = \frac{\sum_{m \in \mathcal{E}_N^{\text{upper}}} \hat{p}_m}{\text{card}\, \mathcal{E}_N^{\text{upper}}}, \quad \tilde{p}_N^{\text{lower}} = \frac{\sum_{m \in \mathcal{E}_N^{\text{lower}}} \hat{p}_m}{\text{card}\, \mathcal{E}_N^{\text{lower}}} \quad \text{and} \quad \tilde{p}_N^{\text{mix}} = \frac{\sum_{m \in \mathcal{E}_N^{\text{mix}}} \hat{p}_m}{\text{card}\, \mathcal{E}_N^{\text{mix}}}.$$

Notice that

$$\overline{p}_N = R_N \left(\rho_N^{\text{upper}} \tilde{p}_N^{\text{upper}} + \rho_N^{\text{lower}} \tilde{p}_N^{\text{lower}} + \rho_N^{\text{mix}} \tilde{p}_N^{\text{mix}} \right).$$

In other words, we have separated the non-zero contributions to $\overline{p}_N$ into (i) realizations for which all the replicas go through the upper channel (first term in the parenthesis), (ii) realizations for which all the replicas go through the lower channel (second term in the parenthesis), and finally (iii) realizations for which the two channels are used by the replicas (third term in the parenthesis).

Contrary to $\overline{p}_N$, the limit when $N \to \infty$ of the estimators R_N, ρ_N^{upper}, ρ_N^{lower}, ρ_N^{mix}, $\tilde{p}_N^{\text{upper}}$, $\tilde{p}_N^{\text{mix}}$ or $\tilde{p}_N^{\text{lower}}$ (for a given value of n_{rep}) depends on the choice of the reaction coordinate ξ.

From Table 5, we observe that for ξ^1, approximately half of the realizations use exclusively the upper channel and the other half use the lower channel. The associated conditional estimators $\tilde{p}_N^{\text{upper}}$ and $\tilde{p}_N^{\text{lower}}$ are very close. This is not the case for ξ^3: only very few realizations go through the upper channel while the associated probability $\tilde{p}_N^{\text{upper}}$ is much larger than the two other ones $\tilde{p}_N^{\text{lower}}$ and $\tilde{p}_N^{\text{mix}}$. This means that a few realizations contribute a lot to the empirical average $\overline{p}_N$. This explains the very large confidence intervals observed with ξ^3 (in comparison with those observed for ξ^1) on Figure 4.

	β	N	R_N	ρ_N^{upper}	ρ_N^{mix}	ρ_N^{lower}	$\tilde{p}_N^{\text{upper}}$	$\tilde{p}_N^{\text{mix}}$	$\tilde{p}_N^{\text{lower}}$	$\overline{p}_N$
ξ^1	8.67	2.10^6	0.81	0.45	0.03	0.52	2.7e-09	3.0e-09	2.3e-09	1.7e-09
ξ^3	8.67	2.10^6	0.99	0.0008	0.02	0.98	2.3e-06	5.9e-10	5.5e-10	2.4e-09
ξ^1	9.33	4.10^6	0.72	0.51	0.02	0.47	6.2e-10	6.3e-10	2.5e-10	3.2e-10
ξ^3	9.33	4.10^6	0.97	0.0005	0.02	0.98	1.0e-06	5.6e-11	9.7e-11	6.0e-10
ξ^1	10	6.10^6	0.62	0.51	0.01	0.48	1.5e-10	1.4e-10	5.2e-11	6.2e-11
ξ^3	10	6.10^6	0.92	0.0004	0.01	0.99	1.4e-07	1.5e-11	1.8e-11	6.8e-11

Table 5: The bi-channel case. Proportion and conditional probabilities for two reaction coordinates: the norm to the initial point (ξ^1) and the abscissa (ξ^3).

Remark 5.2 (On ρ_N^{mix})**.** We observe that for both reaction coordinates the value of ρ_N^{mix} is very small: on most realizations, if at least one replica reaches B before A then all replicas reaching B go through the same channel. This effect is due to the rather small number of replicas (namely $n_{\text{rep}} = 100$). When n_{rep} increases, we observe both upper and lower paths on each realization, see for instance [10] for such experiments.

Remark 5.3 (On the efficiency of ξ^1)**.** We see in Table 5 that R_N is smaller for the best reaction coordinate ξ^1 (in terms of fluctuations) than for ξ^3. Indeed, the potential V admits a local minimum at $x^* \simeq (0, 1.54)$ with $\xi^1(x^*) \simeq 1.83$, which is very close to $z^1_{\max} = 1.9$. Notice that the influence of the local minimum is much weaker when using the abscissa since $\xi^3(x^*) = 0$ which is far from $z^3_{\max} = 0.9$ and since most of the trajectories go through the lower channel for this reaction coordinate.

Using ξ^1 as a reaction coordinate, replica X going though the upper channel is likely to get trapped around the upper local minimum, and the resampling procedure may produce trajectories which go back to A without increasing the level of the parent replica: as a consequence a higher rate of extinction is observed for ξ^1 than for ξ^3.

Moreover, it may happen that a replica satisfies $\mathrm{T}_{z_{\max}}(X) < \mathrm{T}_A(X)$ without hitting B before A; it has then no contribution in the estimator of the probability $\mathbb{P}(\tau_B < \tau_A)$. On the left plot on Figure 5, we represent the $n_{\mathrm{rep}} = 20$ replicas obtained at the end of one realization of the algorithm where $\beta = 6.33$, and using ξ^1: only 6 replicas out of 20 reach B even if all of them have a maximum level larger than the stopping level $z_{\max}$. In such a case, $P_{\mathrm{corr}} = 6/20$ (see (21)).

Remark 5.4 (On the degeneracy of the branching tree)**.** On the right plot of Figure 5, another phenomenon is illustrated: for a small number of replicas, we typically observe that the working replicas at the final iteration of the algorithm have only a few common ancestors. For example, for the realization of the algorithm represented on the right plot of Figure 5 with $\beta = 6.33$ and using ξ^3 (abscissa) as the reaction coordinate, the $n_{\mathrm{rep}} = 20$ working replicas are all issued from only two ancestors, at the end of the algorithm. Of course, the number of common ancestors increases when n_{rep} is larger (see for example [10]).

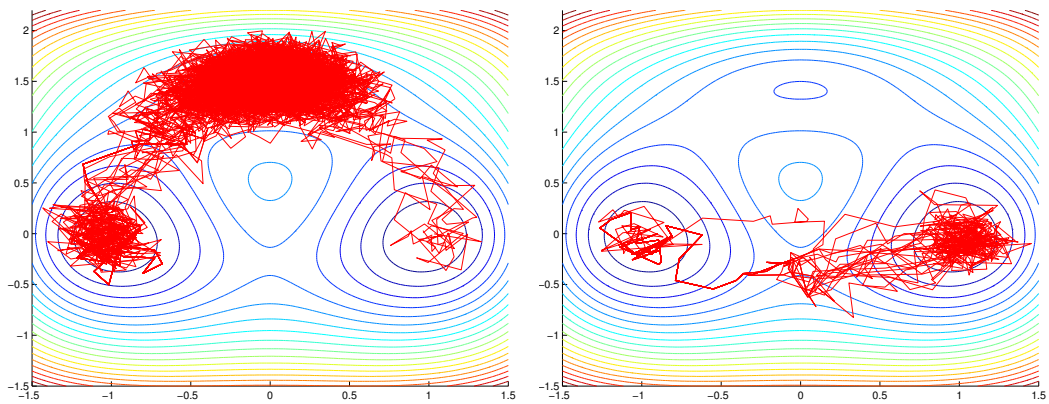

Figure 5: Trajectories obtained at the end of a realization of the algorithm for which $\beta = 6.33$, $n_{\mathrm{rep}} = 20$. Left: the reaction coordinate is ξ^1 and only 6 of the working replicas have reached B ($P_{\mathrm{corr}} = 6/20$), see Remark 5.3. Right: the reaction coordinate is ξ^3 and the 20 replicas are all issued from only two ancestors, see Remark 5.4.

Remark 5.5 (On the limit of R_N, ρ_N^{upper}, ρ_N^{lower}, ρ_N^{mix} when $N \to \infty$)**.** The fact that the limits when N goes to infinity of the proportions R_N, ρ_N^{upper}, ρ_N^{lower}, ρ_N^{mix} or the conditional probabilities $\tilde{p}_N^{\mathrm{upper}}$, $\tilde{p}_N^{\mathrm{mix}}$ or $\tilde{p}_N^{\mathrm{lower}}$ depend on ξ is not in contradiction with the unbiasedness result of Theorem 4.1. Indeed these quantities are not estimators of the form (19). In particular, there are not associated with a functional of the dynamics of the form (9). For example, $\tilde{p}_N^{\mathrm{upper}}$ should not be confused with the empirical estimator $\overline{p_N^{\mathrm{upper}}} = \frac{1}{N}\sum_{m=1}^{N} \hat{p}_m^{\mathrm{upper}}$, where

(for a fixed realization m)

$$\hat{p}^{\text{upper}} = \sum_{n \in I_{\text{on}}^{(Q_{\text{iter}})}} G^{(n,Q_{\text{iter}})} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\text{iter}})}) < \mathrm{T}_A(X^{(n,Q_{\text{iter}})})} \mathbb{1}_{X^{(n,Q_{\text{iter}})} \in \text{Upper}}$$
$$= \frac{n_{\text{rep}} - K^{(Q_{\text{iter}})}}{n_{\text{rep}}} \cdots \frac{n_{\text{rep}} - K^{(1)}}{n_{\text{rep}}} \left(\frac{1}{n_{\text{rep}}} \sum_{n \in I_{\text{on}}^{(Q_{\text{iter}})}} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\text{iter}})}) < \mathrm{T}_A(X^{(n,Q_{\text{iter}})})} \mathbb{1}_{X^{(n,Q_{\text{iter}})} \in \text{Upper}} \right)$$

which is an unbiased estimator of $\mathbb{P}\left(\mathrm{T}_B(X) < \mathrm{T}_A(X) \text{ and } X \in \text{Upper}\right)$. The (unbiased) estimator $\hat{p}^{\text{lower}}$ of $\mathbb{P}\left(\mathrm{T}_B(X) < \mathrm{T}_A(X) \text{ and } X \in \text{Lower}\right)$ is defined similarly, as well as the empirical average $\overline{p_N^{\text{lower}}} = \frac{1}{N} \sum_{m=1}^{N} \hat{p}_m^{\text{lower}}$.

Table 6 contains the values of $\overline{p_N^{\text{upper}}}$ and $\overline{p_N^{\text{lower}}}$ computing using the same realizations as above. We include the values of the width δ_N^{upper} and δ_N^{lower} of the confidence intervals associated with the Monte-Carlo procedure (see (39)).

	β	N	$\overline{p_N^{\text{lower}}}$	δ_N^{lower}	$\overline{p_N^{\text{upper}}}$	δ_N^{upper}	$\overline{p}_N$
ξ^1	8.67	2.10^6	5.2e-10	5.1e-11	1.2e-09	4.4e-11	1.7e-09
ξ^3	8.67	2.10^6	5.4e-10	5.0e-11	1.9e-09	3.5e-09	2.4e-09
ξ^1	9.33	4.10^6	9.1e-11	9.4e-12	2.3e-10	7.5e-12	3.2e-10
ξ^3	9.33	4.10^6	8.3e-11	7.9e-12	5.2e-10	7.7e-10	6.0e-10
ξ^1	10	6.10^6	1.6e-11	3.0e-12	4.6e-11	1.8e-12	6.2e-11
ξ^3	10	6.10^6	1.8e-11	2.2e-12	5.0e-11	7.0e-11	6.8e-11

Table 6: The bi-channel case. Estimation of upper and lower channels probabilities for two reaction coordinates: the norm to the initial point (ξ^1) and the abscissa (ξ^3).

We observe good agreement between the values computed with the two reaction coordinates, as predicted by the unbiasedness result of Theorem 4.1. The estimates of the probability $\mathbb{P}\left(\mathrm{T}_B(X) < \mathrm{T}_A(X) \text{ and } X \in \text{Lower}\right)$ and the corresponding empirical variances obtained with both reaction coordinates are very close. However, when using the abscissa ξ^3 we see that the confidence interval $[\overline{p_N^{\text{upper}}} - \delta_N^{\text{upper}}/2; \overline{p_N^{\text{upper}}} + \delta_N^{\text{upper}}/2]$ is much larger than the one when using ξ^1; the variance is larger because taking the upper channel is less likely when using ξ^3. Moreover, if we plot the equivalent of Figure 4 for $\overline{p_N^{\text{upper}}}$, we observe the same kind of behavior for the evolution of the confidence interval related with $\overline{p_N^{\text{upper}}}$ as function of the number of realizations N: $\overline{p}_N$ and $\overline{p_N^{\text{upper}}}$ jump at the same values of N.

5.3 The second two-dimensional example

Let us finally consider another example in dimension two, already used in [3]. Our aim is to show that the very large fluctuations observed with some reaction coordinates in the first two-dimensional example are related to the existence of multiple pathways from A to B. This is actually very much related to some discussions in the paper [14] about the origins of the so-called "apparent bias" which is observed with splitting algorithms.

In the example considered in this section, we will indeed observe that as a parameter is changed in the model, one goes from a situation where the estimator has the same statistical behavior whatever the reaction coordinates (when there is only one pathway from A to B)

to a situation where the estimation using one of the reaction coordinate deteriorates for too small Monte-Carlo sample sizes (when two pathways link A to B), even though the estimated probability is approximately the same in both situations.

5.3.1 The model

The second example is inspired by the space discretization of the Allen-Cahn equation (see [3]). The dynamics is again the overdamped Langevin equation (41) discretized using the Euler-Maruyama method. The potential function $\mathcal{E}_\gamma$ depends on a parameter $\gamma > 0$ and is given by

$$\mathcal{E}_\gamma(x, y) = \gamma(x - y)^2 + \frac{1}{2}\left(V(x) + V(y)\right),$$

where $V(z) = \frac{z^4}{4} - \frac{z^2}{2}$ is a double-well potential.

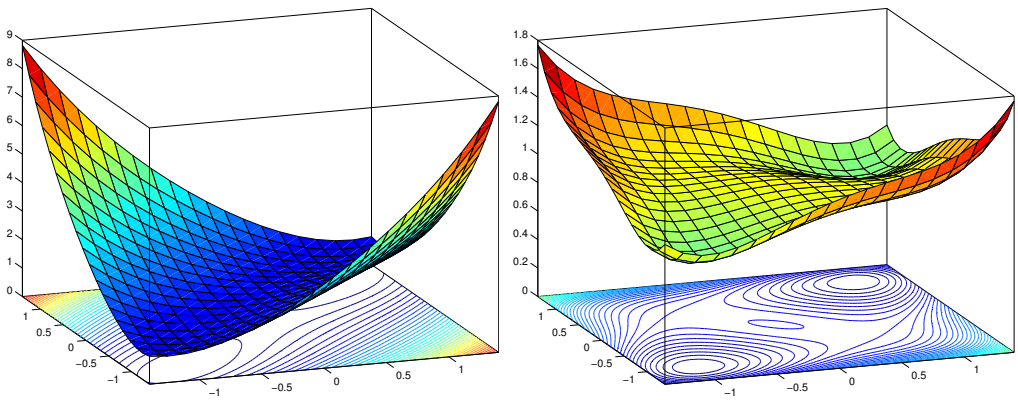

Figure 6: Plot of the potential function for the discretized Allen-Cahn problem. Left: $\gamma = 1$; right: $\gamma = 0.1$.

For any value of $\gamma > 0$, there are two global minima $m_A = (x_A, y_A) = (-1, -1)$ and $m_B = (x_B, y_B) = (1, 1)$. Moreover, this model exhibits bifurcations with respect to the parameter γ. When $\gamma > 1/8$, $(0, 0)$ is the only saddle point, whereas for $\gamma < 1/8$ the latter point degenerates into a local maximum and additional saddle points appear (as well as local minima if γ is further decreased).

The AMS algorithm is tested with four reaction coordinates (the first three ones being the same as in Section 5.2):

1. the *norm to the initial point* m_A*:* $\xi^1(x, y) = \sqrt{(x - x_A)^2 + (y - y_A)^2}$,
2. the *norm to the final point* m_B*:* $\xi^2(x, y) = \xi^1(x_B, y_B) - \sqrt{(x - x_B)^2 + (y - y_B)^2}$,
3. the *abscissa:* $\xi^3(x, y) = x$,
4. the *magnetization:* $\xi^4(x, y) = (x + y)/2$.

The fourth reaction coordinate is called the magnetization because of its interpretation in the original Allen-Cahn problem. In the figures below, we associate a color to each reaction coordinate: green for ξ^1, red for ξ^2, blue for ξ^3 and cyan for ξ^4.

For the reaction coordinate ξ^i, the maximum level used in the stopping criterion of the algorithm is denoted by $z^i_{\max}$. In the simulations, the following values are used: $z^1_{\max} = z^2_{\max} = \sqrt{7.6}$ and $z^3_{\max} = z^4_{\max} = 0.9$.

As above,

$$\begin{cases} A = \mathcal{B}(m_A, \rho) = \left\{(x,y) \in \mathbb{R}^2 : \sqrt{(x-x_A)^2 + (y-y_A)^2} < \rho\right\} \\ B = \mathcal{B}(m_B, \rho) = \left\{(x,y) \in \mathbb{R}^2 : \sqrt{(x-x_B)^2 + (y-y_B)^2} < \rho\right\}, \end{cases}$$

with $\rho = 0.05$. Notice that for $i \in \{1,2,3,4\}$, we have $B \subset (\xi^i)^{-1}(]z^i_{\max}, +\infty[)$ (see (10)).

The deterministic initial condition is $X_0 = x_0 = (-0.9, -0.9)$, and we always take $k = 1$. By default, the number of replicas is taken equal to $n_{\text{rep}} = 100$, and the empirical averages are computed with $N = 10^6$ independent realizations. When $\gamma = 0.1$, we also take $n_{\text{rep}} = 10$ (with $N = 6.10^6$) and $n_{\text{rep}} = 1000$ (with $N = 10^5$). Notice that we have also tested the algorithm in the latter case when $k > 1$: since we observe the same kind of behavior as for $k = 1$, we do not present the results of these numerical simulations.

5.3.2 Simulations for $\gamma = 1$ and $\beta \in \{10, 20, 40, 80\}$

Let us first consider the case $\gamma = 1$. In this situation there is only one reactive path to go from A to B, going through the saddle point $(0,0)$. Let us consider the following values for the inverse temperature $\beta \in \{10, 20, 40, 80\}$. We plot on Figure 7 the evolution of the confidence interval for the estimator $\hat{p}$ as a function of the number of independent realizations, for $n_{\text{rep}} = 100$. We observe that the confidence intervals overlap, and that the statistical fluctuations are very similar whatever the reaction coordinate.

For comparison, the results obtained with a standard direct Monte-Carlo estimation are given in Table 7 for $\beta \in \{10, 20, 40\}$ with $N = 6.10^8$ realizations. The results are consistent with those obtained by the AMS algorithm. For $\beta = 80$, the probability is very small and we were not able to get a reliable result by standard direct Monte-Carlo simulations with a reasonable number of realizations.

β	10	20	40
$\overline{p}_N^{\text{MC}}$	2.755e-2	2.062e-3	1.582e-5
δ_N	0.003e-2	0.007e-3	0.063e-5

Table 7: Standard direct Monte-Carlo estimation with $N = 6.10^8$ realizations.

5.3.3 Simulations for $\gamma = 0.1$ and $\beta = 80$

When the parameter γ is smaller than $1/8$, the point $(0,0)$ is no longer a saddle point (but instead a local maximum). In this case, there are two saddle points on the line $\xi^4 = 0$ (they are symmetric with respect to $(0,0)$). We take $\gamma = 0.1$ and represent on Figure 8 the evolution of the confidence interval for the estimator $\hat{p}$ as a function of the number N of independent runs of the algorithm and for different values of n_{rep}: $n_{\text{rep}} = 10$ (with up to $N = 6.10^6$ realizations), $n_{\text{rep}} = 100$ (with up to $N = 10^6$ realizations) and $n_{\text{rep}} = 1000$ (with up to $N = 10^5$ realizations). The inverse temperature is fixed to $\beta = 80$.

We observe that the statistical fluctuations of the estimator for the reaction coordinate ξ^3 (abscissa) are much larger than for the three other ones. The estimator is unbiased, but a large

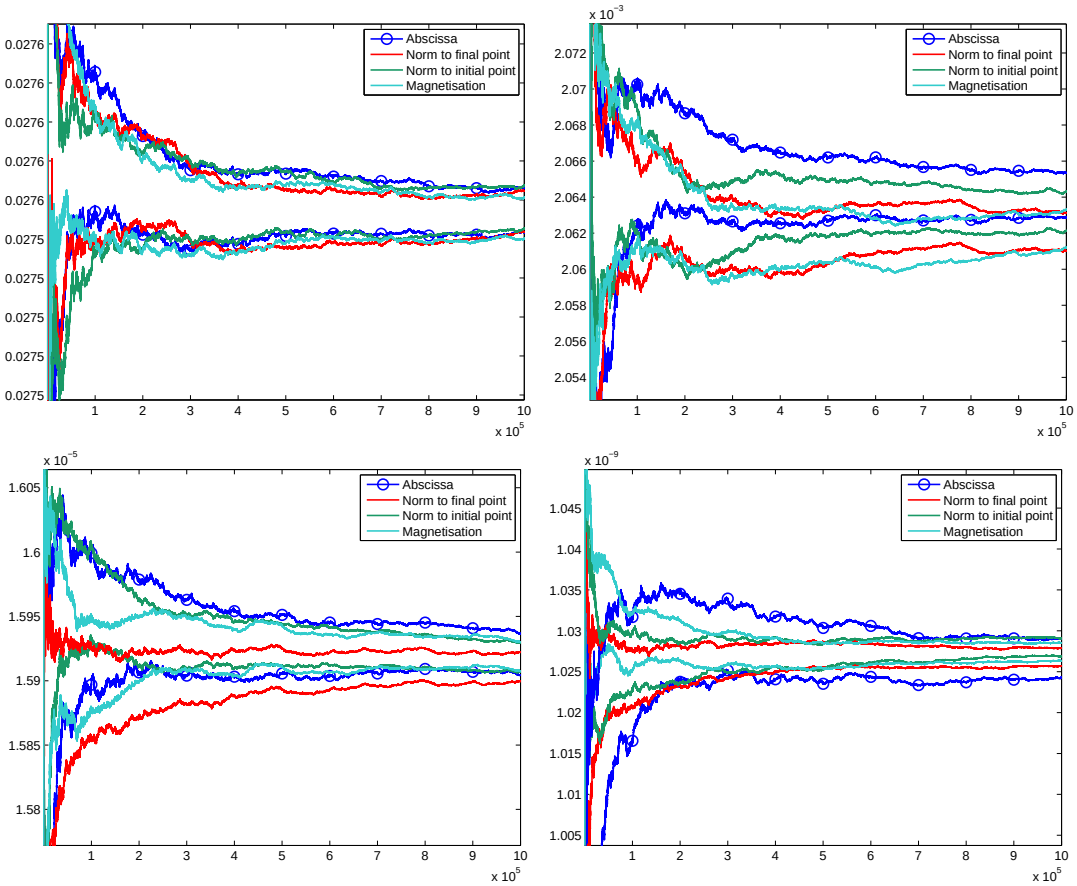

Figure 7: Evolution as a function of N of the empirical mean $\overline{p}_N$ and of the associated 95% confidence intervals $[\overline{p}_n - \delta_N/2, \overline{p}_n + \delta_N/2]$ with $\gamma = 1$. Upper left: $\beta = 10$, upper right: $\beta = 20$, lower left: $\beta = 40$, lower right $\beta = 80$.

number of realizations is required to get a significant confidence interval. By comparing with the results obtained in the Section 5.2, we thus conclude that the origin of these fluctuations is related to the existence of two pathways from A to B, rather than to the smallness of the estimated probability for example. This is in agreement with some discussions in the paper [14] about the origins of the so-called "apparent bias" which is observed with splitting algorithms.

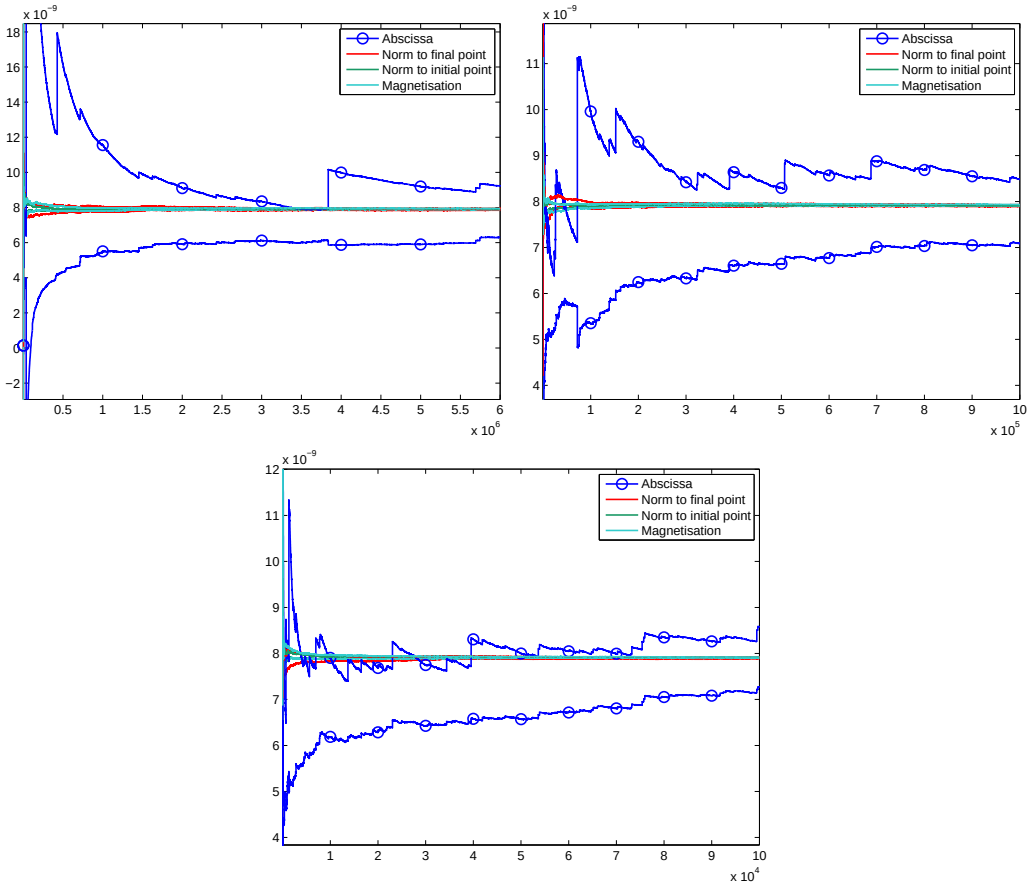

Figure 8: Evolution as a function of N of the empirical mean $\overline{p}_N$ and of the associated 95% confidence intervals $[\overline{p}_n - \delta_N/2, \overline{p}_n + \delta_N/2]$ with $\gamma = 0.1$ and $\beta = 80$. Upper left $n_{\text{rep}} = 10$, upper right $n_{\text{rep}} = 100$, lower $n_{\text{rep}} = 1000$.

5.4 Conclusions and practical recommendations

Let us summarize our findings on these numerical simulations.

- We always observe that for sufficiently large values of N (number of independent Monte Carlo simulations), the confidence intervals of the estimator $\hat{p}$ overlap, whatever n_{rep}, k or ξ. This is in accordance with our theoretical result on the unbiasedness of this estimator.

- We observe numerically in our two-dimensional simulations that the estimator $\hat{p}$ has a heavy tail so that very few realizations contribute a lot to the empirical average. This

explains why for some choices of the reaction coordinate the efficiency of the estimation may be very poor since the empirical average converges very slowly to its limit (namely N should be taken sufficiently large to obtain significant results). Actually, as explained in Remark 5.1, even in some idealized setting where the reaction coordinate is the committor function, in some small probability regime, the estimator has a log-normal distribution, and thus has heavy tails.

- In multiple channel cases (namely when multiple pathways exist from A to B), one may observe non-overlapping empirical confidence intervals of the estimator for different reaction coordinates if the number of independent realizations N is too small. This is related to the fact that very large contributions to the average of the estimator are associated with trajectories going through very unlikely (for the considered reaction coordinate and value of n_{rep}) channels. This is a known phenomenon for splitting algorithms in general, see [14], where it is referred to as "apparent bias". In particular, a good reaction coordinate in a multiple channel case is such that, conditionally to reach a certain maximum level z, the relative likelihood of the channels used by the paths to reach this maximum level does not depend too much on z. For example, a reaction coordinate close to the committor function is a good candidate to achieve this purpose. This opens the route to adaptive algorithms, where the reaction coordinate would be updated in order to get closer and closer to the committor function as long as successive AMS algorithms are launched (see [10]). We intend to investigate this direction in future works.

As a conclusion to these numerical results, we thus recommend the following in order to get reliable estimates of the probability $\mathbb{P}(\tau_B < \tau_A)$ with the AMS algorithm.

- One should be careful in the implementation of the splitting and branching steps, in particular in the treatment of replicas which have the same maximum level and in the definition of the branching point in the resampling procedure. For correct implementations, unbiased estimators can be built, and the general framework of Section 3 yields many variants for the algorithm.

- Thanks to the unbiasedness property, one should check the independence of the computed probability on the choice of the parameters: the number of replicas n_{rep}, the minimum number of resampled replicas k and the reaction coordinates ξ. In particular, we recommend to perform simulations with various reaction coordinates and to set the minimal number of independent realizations such that the empirical confidence intervals overlap.

- Thanks to the unbiasedness property, one can perform many independent realizations of the algorithm with a relatively small number of replicas, instead of using a few independent realizations with a large number of replicas. Indeed, assume that we are in a regime where the variance scales like $\frac{1}{n_{\text{rep}} \times N}$; this is the case for instance in the so-called ideal case for sufficiently large n_{rep} and N, see [5]. Since the parallelization of independent runs of the algorithm is trivial, for a fixed product $n_{\text{rep}} \times N$ (namely for a fixed CPU cost), the strategy with less replicas is thus much more interesting in terms of wall-clock time (which scales like n_{rep}) than the strategy with more replicas.

Acknowledgements

C.-E. Bréhier is grateful to INRIA Rocquencourt for having funded his postdoctoral position (September 2013-December 2014), and acknowledges support from the SNF grant 200020-149871/1. M. Gazeau is grateful to INRIA Lille Nord Europe where a part of this research has been conducted and to Labex CEMPI (ANR-11-LABX-0007-01). The work of T. Lelièvre and M. Rousset is supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement number 614492. The authors are grateful to the Labex Bezout (ANR-10-LABX-58-01) which supported this project at an early stage, during the CEMRACS 2013. Finally, the authors would like to thank F. Bouchet, F. Cérou, A. Guyader, J. Rolland and E. Simonnet for many fruitful discussions.

References

[1] P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999.

[2] C.-E. Bréhier. Large deviations principle for the adaptive multilevel splitting algorithm in an idealized setting. *Preprint*, 2015. `http://arxiv.org/abs/1502.06780`.

[3] C.-E. Bréhier, M. Gazeau, L. Goudenège, and M. Rousset. Analysis and simulations of rare events for SPDEs. *ESAIM Proceedings and Reviews*, 48:364–384, 2015.

[4] C.-E. Bréhier, L. Goudenège, and L. Tudela. Central limit theorem for adaptive multilevel splitting estimators in an idealized setting. *Preprint*, 2014. `http://arxiv.org/abs/1501.01399`.

[5] C.-E. Bréhier, T. Lelièvre, and M. Rousset. Analysis of adaptive multilevel splitting algorithms in an idealized setting. *ESAIM Probability and Statistics*, to appear, 2015.

[6] J.A. Bucklew. *Introduction to rare event simulation.* Springer Series in Statistics. Springer-Verlag, New York, 2004.

[7] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Stat. Comput.*, 22(3):795–808, 2012.

[8] F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis. *Stoch. Anal. Appl.*, 25(2):417–443, 2007.

[9] F. Cérou and A. Guyader. Fluctuations of adaptive multilevel splitting. *Preprint*, 2014. `http://arxiv.org/abs/1408.6366`.

[10] F. Cérou, A. Guyader, T. Lelièvre, and D. Pommier. A multiple replica approach to simulate reactive trajectories. *Journal of Chemical Physics*, 134(5), 2011.

[11] P. Del Moral. *Feynman-Kac Formulae.* Springer, 2004.

[12] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice.* Springer, 2001.

[13] M.J.J. Garvels, D.P. Kroese, and J.C.W. van Ommeren. On the importance function in splitting simulation. *European Transactions on Telecom- munications*, 13(4):363–371, 2002.

[14] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. A large deviations perspective on the efficiency of multilevel splitting. *IEEE Trans. Automat. Control*, 43(12):1666–1679, 1998.

[15] A. Guyader, N. Hengartner, and E. Matzner-Løber. Simulation and estimation of extreme quantiles and extreme probabilities. *Appl. Math. Optim.*, 64(2):171–196, 2011.

[16] A.M. Johansen, P. Del Moral, and A. Doucet. Sequential Monte Carlo samplers for rare events. In *Proceedings of the 6th International Workshop on Rare Event Simulation, RESIM 2006, Bamberg*, pages 256–267, 2006.

[17] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991.

[18] P. Metzner, C. Schütte, and E. Vanden-Eijnden. Illustration of transition path theory on a collection of simp le examples. *Journal of Chemical Physics*, 125(1), 2006.

[19] S. Park, M.K. Sener, D. Lu, and K. Schulten. Reaction paths based on mean first-passage times. *J. of Chem. Phys.*, 119:1313–1319, 2003.

[20] J. Rolland and E. Simonnet. Statistical behaviour of adaptive multilevel splitting algorithms in simple models. *Journal of Computational Physics*, 283:541 – 558, 2015.

[21] A.N. Shiryayev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1984. Translated from the Russian by R. P. Boas.

[22] E. Simonnet. Combinatorial analysis of the adaptive last particle method. *Statistics and Computing*, pages 1–20, 2014.