



HAL
open science

Fluctuations of Rare Event Simulation with Monte Carlo Splitting in the Small Noise Asymptotics

Frédéric Cérou, Sofiane Martel, Mathias Rousset

► **To cite this version:**

Frédéric Cérou, Sofiane Martel, Mathias Rousset. Fluctuations of Rare Event Simulation with Monte Carlo Splitting in the Small Noise Asymptotics. 2022. hal-03889692

HAL Id: hal-03889692

<https://hal.science/hal-03889692>

Preprint submitted on 8 Dec 2022

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.

Fluctuations of Rare Event Simulation with Monte Carlo Splitting in the Small Noise Asymptotics

Frédéric Cérou¹, Sofiane Martel², and Mathias Rousset¹

¹IRMAR and Inria, University of Rennes, France.

²École Nationale Des Ponts et Chaussées, France.

Abstract

Diffusion processes with small noise conditioned to reach a target set are considered. The AMS algorithm is a Monte Carlo method that is used to sample such rare events by iteratively simulating clones of the process and selecting trajectories that have reached the highest value of a so-called importance function. In this paper, the large sample size relative variance of the AMS small probability estimator is considered. The main result is a large deviations logarithmic equivalent of the latter in the small noise asymptotics, which is rigorously derived. It is given as a maximisation problem explicit in terms of the quasi-potential cost function associated with the underlying small noise large deviations. Necessary and sufficient geometric conditions ensuring the vanishing of the obtained quantity ('weak' asymptotic efficiency) are provided. Interpretations and practical consequences are discussed.

Contents

1	Introduction	2
2	Summary of notations	7
3	Framework and results	8
3.1	Typical framework: The Freidlin-Wentzell theory	8
3.2	Rare event simulation: The splitting algorithms and their asymptotic variances	9
3.3	Assumptions and basic consequences	12
3.4	Main results	16
4	Interpretation and insights	18
4.1	The variance formula 1.2	18
4.2	A disclaimer about variance	20
4.3	The loss function: discussion	20
4.4	Summary	23
4.5	Interpretation as a Hamilton-Jacobi sub-solution	23
4.6	Possible practical consequences	23
4.7	Comparison to previous work	24

5	Large Deviations estimates and proof of the main result	25
5.1	Stopping times and topology	25
5.2	Continuity of the cost to reach level sets	25
5.3	Small noise asymptotics of the rare event	26
5.4	Varadhan lemmas	27
5.5	A Large Deviations estimate	28
5.6	Analysis of the loss functional	32
5.7	Small noise analysis of variance	33
5.8	AMS for two clones and at worse one iteration	35
A	More on Freidlin-Wentzell	36

1 Introduction

Let $\mathbf{X}^\varepsilon = (\mathbf{X}_t^\varepsilon)_{t \geq 0}$ denotes a diffusion process with small noise parameter $\varepsilon > 0$ and initial condition $\mathbf{X}_0 = x_0$. We are interested in this paper with the simulation of rare events of the form $\{\tau_B(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)\}$ where $\tau_B(\mathbf{x})$ generically denotes the first hitting time of a set $B \subset E$ by the trajectory \mathbf{x} . In the present work, the 'target' set B is defined as the level-set $B = \{\xi \geq l_B\} = \{x \in E \mid \xi(x) \geq l_B\}$ of a continuous function $\xi : E \rightarrow \mathbb{R}$, and the *reference* set A typically contains the attractors of the deterministic dynamics $(\mathbf{X}_t^0)_{t \geq 0}$, and, as such is a recurrent set for the process.

Such problems are of primary interest in different fields within computational physics. Notable recent examples include molecular simulation ([29, 36]), neutron transport ([30]), and climate forecast ([33, 28]). In the latter references, the Monte Carlo methods chosen to perform the rare event simulation are Importance Splitting (a.k.a. Multilevel Splitting) type methods with N clones ('fixed effort' algorithms), and are identical or minor variants of the algorithms studied in the present paper. Those algorithms can also be interpreted as Sequential Monte Carlo samplers as studied in [16], whose structure is defined by a Feynman-Kac model, leading to unbiased estimates of the rare event probability, as studied by P. Del Moral in [15] for instance.

The general idea of Importance Splitting, is to simulate N clones

$$\mathbf{X}^{(1,j)}, \dots, \mathbf{X}^{(N,j)}$$

using the dynamics of \mathbf{X}^ε in a sequential way, j denoting the iteration parameter. At each iteration, the considered algorithm discard trajectories far away from the target set $\{\xi \geq l_B\}$, and then do split (or branch/duplicate/clone) the trajectories heading closer to $\{\xi \geq l_B\}$. In order to quantify the closeness to the target set those methods critically rely on the specific choice of the *importance function* (also called *reaction coordinate*) $\xi : E \rightarrow \mathbb{R}$ on $\{\xi < l_B\}$, the rare event of interest depending only on the target set $B = \{\xi \geq l_B\}$. As one may know only very little about the typical trajectories reaching B , the specific choice of ξ is usually based on intuitions or loose qualitative knowledge. It is now accepted that this choice is the main bottleneck parameter of the efficiency of those Monte Carlo methods. An optimal theoretical choice of ξ is given by the so called *committor function* defined by

$$\xi_*^\varepsilon(x) = \varepsilon \log \mathbb{P}_x(\tau_B(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)),$$

together with $l_B = 0$. The latter choice, which is in most cases unknown and thus practically infeasible, yields an estimator of the rare event probability

with an explicit Poisson distribution and with minimal variance (see *e.g.* Section 2.2.3 in [3]). We will also use the limiting small noise committor function defined by $\xi_*(x) = \lim_{\varepsilon \rightarrow 0} \xi_*^\varepsilon(x)$, that will be equal (under our technical assumptions) to $\xi_*(x) = U(x, B)$ where U is the subsequently defined two-points quasi-potential function.

The main variant we will consider in this work is the so-called Adaptive Multilevel Splitting (AMS) algorithm. The latter can be seen as a limit of a somehow simpler variant – that we will call the Fixed Multilevel Splitting (FMS) method – which is an instance of Sequential Monte-Carlo (SMC) sampling. AMS and FMS are detailed in Section 3.2 below. In each iteration of the AMS algorithm, the least performing clone, in terms of the maximum denoted L of the importance function ξ along its trajectory, is discarded; it is then replaced by the duplicate of one of the other trajectories, chosen uniformly among the $N - 1$ survivors. The duplication is kept identical from the initial condition up until the first hitting time τ_L of the level set $\{\xi \geq L\}$, and the duplicated clone is then redrawn independently after τ_L using the dynamics of \mathbf{X}^ε .

Introducing the notation $\tau_l(\mathbf{x}) \stackrel{\text{def.}}{=} \tau_{\{\xi \geq l\}}(\mathbf{x})$, the estimator of the rare event probability associated with level l ,

$$p_l^\varepsilon \stackrel{\text{def.}}{=} \mathbb{P}[\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)],$$

is given by

$$p_{l,\text{ams}}^N \stackrel{\text{def.}}{=} (1 - 1/N)^{I_l^N}$$

where I_l^N is the random number of iterations required so that all clones have reached the target set $\{\xi \geq l\}$. The estimator $p_{l,\text{ams}}^N$ (as well as other non-normalized estimators) is unbiased $\mathbb{E}[p_{l,\text{ams}}^N] = p_l^\varepsilon$ (see [7, 2]). The empirical distribution of clones at iteration I_l^N

$$\eta_l^{N,\text{path}} = \frac{1}{N} \sum_{n=1}^N \delta_{\mathbf{x}^{(n, I_l^N)}},$$

consistently estimates the conditional distribution

$$\eta_l^{N,\text{path}} \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} \eta_l^{\varepsilon,\text{path}} \stackrel{\text{def.}}{=} \text{Law}(\mathbf{X}^\varepsilon \mid \tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)).$$

The product estimator $p_{l,\text{ams}}^N \eta_l^{N,\text{path}}$ is also unbiased. The convergence and asymptotic normality of all the latter estimators, when the number of clone N goes to infinity, $\varepsilon > 0$ being fixed, was studied in [7] with an explicit expression of the asymptotic variance (see Section 3.2). The latter is minimal when $\xi = \xi_*^\varepsilon$ in which case it is given by $-(p_l^\varepsilon)^2 \ln p_l^\varepsilon$ so that the relative asymptotic variance is only logarithmic with respect to the rare event probability.

A standard quantity assessing the efficiency of such Monte Carlo algorithms is given by the relative variance times the average computational cost, here at a logarithmic scale:

$$\mathcal{E}(p_{l,\text{ams}}^{\varepsilon,N}) \stackrel{\text{def.}}{=} \varepsilon \log \left(\mathbb{E} \left[\text{Cost}_{l,\text{ams}}^{\varepsilon,N} \right] \text{Var} \left[\frac{p_{l,\text{ams}}^{\varepsilon,N}}{p_l^\varepsilon} \right] \right).$$

The efficiency¹ \mathcal{E} is a simple variance-based criteria properly normalized, in order to be invariant by i) averaging over new independent runs of the full

¹Unlike in [4], the efficiency here is renormalized by the probability of the rare event p_l^ε .

algorithm; ii) multiplication of the estimator by a constant. As pointed out in [26], a good feature of an algorithm would be to have a bounded relative variance at fixed computational cost, when the true probability goes to 0: $p_l^\varepsilon \rightarrow_{\varepsilon \rightarrow 0} 0$. But this is virtually always out of reach in practical applications, and one will only seek, to the very best, a sub-exponential behavior. For this purpose we resort to the criteria \mathcal{E} above (adapted the small noise setting thanks to the use of a logarithmic scale).

This efficiency criterion has to be compared to i) the behavior given by the crude Monte Carlo estimator obtained by direct simulation of N independent trajectories, for which $\mathcal{E}(p_{l,\text{crude}}^{\varepsilon,N}) \sim_{\varepsilon \rightarrow 0} -\varepsilon \log p^\varepsilon$, and ii) the *best* behavior, usually referred to as *asymptotically efficient behavior*, for which the relative variance is sub-exponential at fixed cost, that is $\lim_{\varepsilon \rightarrow 0} \mathcal{E}(p_l^{\varepsilon,N}) = 0$, which happens for the AMS algorithm if (but not only if, as will be proved in this work) $\xi = \xi_*$.

The present work is dedicated to the study when $\varepsilon \rightarrow 0$ of $\lim_{N \rightarrow +\infty} \mathcal{E}(p_{l,\text{ams}}^{\varepsilon,N})$, where we stress that the limit $N \rightarrow +\infty$ is taken first. This choice considerably simplify the still intricate analysis. We mention that it is an open problem to study the small ε limit of $\mathcal{E}(p_{l,\text{ams}}^{\varepsilon,N})$ at fixed N . The optimal efficiency obtained after taking first the limit $N \rightarrow +\infty$ and then $\varepsilon \rightarrow 0$, that is $\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow +\infty} \mathcal{E}(p_l^{\varepsilon,N}) = 0$, will be called here *weak asymptotic efficiency*. We stress that both the order $\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow +\infty}$ and $\lim_{N \rightarrow +\infty} \lim_{\varepsilon \rightarrow 0}$ are relevant for practical applications. The case studied here $\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow +\infty}$ is well suited to mild cases where the Monte Carlo algorithm is able to sample the neighbourhood of the least unlikely trajectory defined by the rare event. Converse cases are more difficult.

In the specific case of the AMS (or FMC) algorithm, a single trajectory of \mathbf{X} is refreshed at each iteration, so that it is fair to set the computational cost equal to the total number of algorithmic iterations $\text{Cost}_{l_B}^{\varepsilon,N} = I_{l_B}^{\varepsilon,N}$ (hence the appellation *fixed effort* algorithm). When the algorithm is convergent (which happens under mild assumptions), one has $\lim_{N \rightarrow +\infty} I_l^{\varepsilon,N}/N = -\log p_l^\varepsilon$ in probability, and since $\log \log p^\varepsilon \ll \log p^\varepsilon$, one obtains that the cost is sub-exponential with respect to ε and can be removed from the definition of efficiency:

$$\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow +\infty} \mathcal{E}(p_{l,\text{ams}}^{\varepsilon,N}) = \lim_{\varepsilon \rightarrow 0} \varepsilon \log \left[\lim_{N \rightarrow +\infty} N \text{Var}(p_{l,\text{ams}}^{\varepsilon,N}/p_l^\varepsilon) \right]. \quad (1.1)$$

The main result of this paper is the rigorous evaluation and interpretation of (1.1) under some mild technical assumptions, most prominently a Freidlin-Wentzell type uniform large deviations principle on \mathbf{X}^ε when $\varepsilon \rightarrow 0$. The obtained logarithmic equivalent will be briefly summarized using (1.2)-(1.3) below. The result is based on the explicit formula for the variance of $p_{l,\text{ams}}^{\varepsilon,N}$ when $N \rightarrow +\infty$ recalled in Section 3.2, which becomes simpler in the large deviations picture.

Introducing the notation $q_l^\varepsilon(x) \stackrel{\text{def.}}{=} \mathbb{P}_{\mathbf{X}_0^\varepsilon=x}[\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)]$ for the probability of the rare event associated with a given level l and initial condition x , as well as $\eta_l^\varepsilon \stackrel{\text{def.}}{=} \text{Law}\left(\mathbf{X}_{\tau_l(\mathbf{X}^\varepsilon)}^\varepsilon \mid \tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)\right)$, the associated distribution of the first hitting place of $\{\xi \geq l\}$ conditioned on occurring before the hitting time of A . We will first remark that, for any given initial condition $\mathbf{X}_0^\varepsilon = x_0$

and target level $l_B \geq \xi(x_0)$:

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \left[\lim_{N \rightarrow +\infty} N \text{Var}(p_{l_B, \text{ams}}^{\varepsilon, N}) \right] = \sup_{l \in [\xi(x_0), l_B]} \lim_{\varepsilon \rightarrow 0} \varepsilon \log \text{Var}_{\eta_l^\varepsilon} (q_{l_B}^\varepsilon p_l^\varepsilon / p_l^\varepsilon), \quad (1.2)$$

which exactly states that, on large deviations scales, the relative variance of the AMS estimator of interest is given by the largest – obtained for l spanning the interval $[\xi(x_0), l_B]$ – relative variance of an unbiased (theoretical) estimators of p_l^ε . This unbiased theoretical estimator is given by $p_{l_*}^\varepsilon q_\varepsilon(X_{l_*}^\varepsilon)$ where $X_{l_*}^\varepsilon \sim \eta_{l_*}^\varepsilon$. We will show that critical levels l_* do exist and belong to the open interval $]\xi(x_0), l_B[$.

The formula (1.2) is interesting in order to interpret the relative variance of the AMS algorithm. We will explain in Section 4.1 that one of the main contributions in the variance formula (1.2) is due to those few trajectories who have been 'lucky' when reaching $\{\xi \geq l_*\}$ for the first time because they have a relatively large remaining probability $q_{l_B}^\varepsilon$ to reach the final level set B before the reference set A . Since the empirical distribution of clones at the first hitting time of $\{\xi \geq l_*\}$ is approximately $\eta_{l_*}^\varepsilon$, the 'lucky' clones do mainly contribute to the $\eta_{l_*}^\varepsilon(q_\varepsilon^2)$ quantity in the variance term.

Our main result is more precise and detailed than this preliminary remark, and consists in proving that

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \left[\lim_{N \rightarrow +\infty} N \text{Var}(p_{l, \text{ams}}^{\varepsilon, N} / p_l^\varepsilon) \right] = \sup_{l' \in [\xi(x_0), l]} \text{Loss}(l') \quad (1.3)$$

where the loss function $\text{Loss}(l) \geq 0$ is a non-negative upper semi-continuous function that quantifies the failure from weak asymptotic efficiency. It is explicitly given by the formula

$$\text{Loss}(l) \stackrel{\text{def.}}{=} 2U(x_0, B) - \inf_{\{\xi=l\}} \left[U^{(l)}(x_0, \cdot) + 2U(\cdot, B) \right] - U(x_0, \{\xi=l\}). \quad (1.4)$$

In the above, the notation $U(C_1, C_2) = \inf_{x_1 \in C_1, x_2 \in C_2} U(x_1, x_2)$ is systematically used, and one has introduced the so-called quasi-potential two-points continuous function $U : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$; a cost function satisfying the triangle inequality $U(x, z) \leq U(x, y) + U(y, z)$ for all x, y, z and defined by

$$U(x, y) \stackrel{\text{def.}}{=} \inf_{\substack{T \geq 0 \\ (\mathbf{x}_0, \mathbf{x}_T) = (x, y) \\ \mathbf{x} \notin A}} I_{[0, T]}(\mathbf{x}),$$

where $I_{[0, T]}(x)$ is the rate function associated with the large deviation principle satisfied by $(\mathbf{X}^\varepsilon)_{\varepsilon > 0}$ on each time interval $[0, T]$. In the same way, $U^{(l)}(x, y) \stackrel{\text{def.}}{=} \inf_{\substack{(\mathbf{x}_0, \mathbf{x}_{\tau_l(\mathbf{x})}) = (x, y) \\ \mathbf{x} \notin A}} I_{[0, \tau_l(\mathbf{x})]}(\mathbf{x})$, is a variant where the minimizing set of trajectories is constrained to take values in $\{\xi \leq l\}$.

In Section 4.3, the formula (1.4) will be interpreted and discussed in details, using a decomposition into an 'underestimation' part and an 'overestimation' part (this nomenclature will be explained); each part being, i) defined by an independent minimization problem involving the intermediate level sets of ξ , ii) non-negative. As a consequence, both are identically 0 if and only if weak asymptotic efficiency is achieved. We will also give in Section 4.3 a geometric interpretation of the loss function. In particular, a *sufficient* condition for weak asymptotic efficiency is given by the following condition: for each l the level

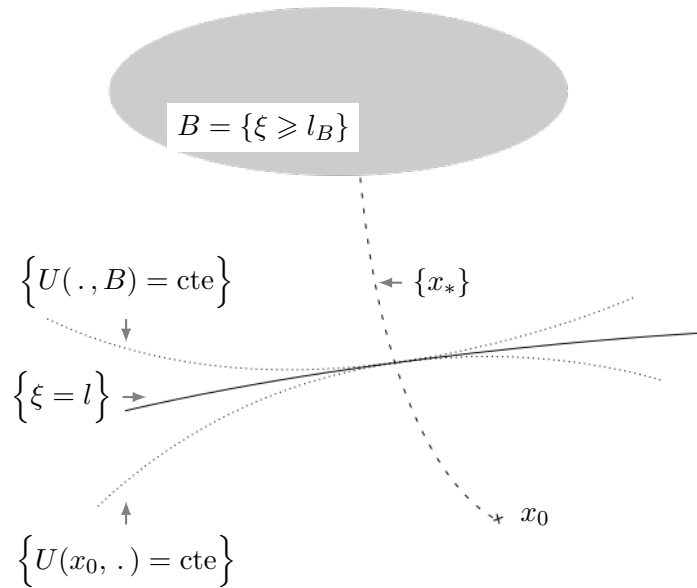


Figure 1: A sufficient condition for weak asymptotic efficiency when satisfied for all $l \in [\xi(x_0), l_B]$. Note that the level sets of the importance function ξ is 'in between' the level sets of the quasi-potential cost i) from the initial condition, and ii) up to the final set B . $\{x_*\}$ represents a least unlikely (optimal) trajectory reaching the rare event (the *instanton*, see Section 4.3).

sets of the importance function ξ are always 'in between' two tangential level sets ('initial' and 'final') of the quasi-potential U . The first 'initial' level set is defined from the initial condition $\{U(x_0, \cdot) = \text{constant}\}$, the second 'final' level set is defined up to the final rare event set $\{U(\cdot, B) = \text{constant}\}$. This condition is equivalent to the existence of a state in $\{\xi = l\}$ that simultaneously minimizes $U(x_0, \cdot)$ and $U(\cdot, B)$. By construction, this state does belong to an optimal trajectory as defined by large deviations of reaching B before A . This sufficient condition for weak asymptotic efficiency is depicted in Figure 1. We will show in Section 4 that a *necessary* condition is given by the weaker demand that only the minimizer of the quasi-potential *from the initial condition* belongs to an optimal path from the initial condition to B (an 'instanton').

The sufficient condition in Figure 1 can be related to sub-solutions of the Hamilton-Jacobi equation associated with the quasi-potential cost function U (see Section 4.5). It is equivalent to the existence of a strictly increasing real valued function F such that $F \circ \xi$ satisfies a certain weaker notion of sub-solution; this weaker notion of sub-solution is given by the usual global notion, i.e. $F \circ \xi(y) - F \circ \xi(x) \leq U(x, y)$, $\forall x, y$, but restricted to either x given by the initial condition $x = x_0$, or y taking values in the target set $y \in B$. In particular, this is a less demanding condition than being the limiting committor function $\xi = \xi_*$, or just $F \circ \xi$ being a sub-solution (which are conditions independent of the initial condition). More comparisons with known results in the literature can be found in Section 4.7.

Finally, some prospects related to practical applications are discussed in Section 4.6. The general idea is that it may be possible to approximate, at least in the large deviations picture, the loss function $\text{Loss}(l)$ in (1.2) or the

variance formula (1.2) using the AMS algorithm or some other *ad hoc* Monte Carlo algorithms. When the latter evaluation can be achieved for various ξ but with a given Monte Carlo procedure constructed with a reference ξ_0 , one can then try to improve ξ_0 by minimizing the obtained quantity over the available choices of ξ .

The paper is organized as follows. In Section 2, the main notations are summarized. Section 3 provides the mathematical context, the considered assumptions, and the rigorously states our main results. More precisely, Freidlin-Wentzell large deviations theory is recalled in Section 3.1, while the description and asymptotic normality for large sample size of the considered Monte Carlo algorithms is given in Section 3.2. Assumptions and results are detailed in respectively Section 3.3 and 3.4. Section 4 is dedicated to the interpretation of the results, with comments on few related works. Finally, Section 5 contains the mathematical proofs, in particular the main large deviations analysis and its consequences.

2 Summary of notations

E denotes the main Polish state space.

Level sets are denoted *e.g.* $\{\xi \leq l\} = \{x \in E \mid \xi(x) \leq l\}$, the minimum (resp. maximum) with \wedge (resp. \vee).

Continuous trajectories in state space E are denoted with bold lower case. \mathbf{x} denotes a generic trajectory, \mathbf{X}^ε is a random trajectory with small noise $\varepsilon > 0$ satisfying a LDP in $C([0, T], E)$ for each T with rate function $I_{[0, T]}$.

We denote entrance times in S with $\tau_S(\mathbf{x}) \stackrel{\text{def.}}{=} \inf\{t \geq 0 : \mathbf{x}_t \in S\}$, with the usual convention $\inf \emptyset = +\infty$. When the topology of S is important in order to obtain upper or lower bound, we will use the notation $\tau_S^+ \stackrel{\text{def.}}{=} \tau_{\overset{\circ}{S}}$ and $\tau_S^- \stackrel{\text{def.}}{=} \tau_{\bar{S}}$. ξ denotes the continuous importance function of interest and we also use the shorthand notation.

Quasi-potential two-points function U is defined by considering rate function minimizing trajectories avoiding A , and with fixed end points, that is $U(x, y) \stackrel{\text{def.}}{=} \inf_{\mathbf{x}, T} I_{[0, T]}(\mathbf{x})$ under the conditions $\tau_A(\mathbf{x}) > T$, and $\mathbf{x}_0 = x, \mathbf{x}_T = y$. In the same way, $U^{(l)}$ denotes the variant where trajectories are also constrained in $\{\xi \leq l\}$, that is, $U^{(l)}(x, y) \stackrel{\text{def.}}{=} \inf_{\mathbf{x}, T} I_{[0, T]}(\mathbf{x})$ under the condition $T < \tau_l^+(\mathbf{x}) \wedge \tau_A(\mathbf{x})$, and $\mathbf{x}_0 = x, \mathbf{x}_T = y$. If C is a subset of E , we will denote $U(C, y) := \inf_{x \in C} U(x, y)$, $U(y, C) := \inf_{z \in C} U(y, z)$, and similarly for $U^{(l)}$.

When $\varepsilon > 0$ plays no role, or when considering objects associated with the Monte Carlo algorithm, we may drop in notations the dependence on ε . In the latter case the dependence on N will be used instead.

The probability that \mathbf{X} reaches level l before A for the considered initial condition x_0 is denoted $p_l \stackrel{\text{def.}}{=} \mathbb{P}[\tau_l(\mathbf{X}) < \tau_A(\mathbf{X})]$; when we want to stress the dependence in the initial condition we will rather denote $q_l(x) \stackrel{\text{def.}}{=} \mathbb{P}_{\mathbf{X}_0=x}(\tau_l(\mathbf{X}) < \tau_A(\mathbf{X}))$ so that $q_l(x_0) = p_l$. The conditional distribution at the first hitting time of level l is denoted $\eta_l(\varphi) \stackrel{\text{def.}}{=} \mathbb{E}[\varphi(\mathbf{X}_{\tau_l(\mathbf{X})}) \mid \tau_l(\mathbf{X}) < \tau_A(\mathbf{X})]$, while its pathwise generalization is denoted $\eta_l^{\text{path}}(\psi) \stackrel{\text{def.}}{=} \mathbb{E}[\psi(\mathbf{X}) \mid \tau_l(\mathbf{X}) < \tau_A(\mathbf{X})]$; φ (resp. ψ) denoting a generic bounded measurable test function on E (resp. path space $C(\mathbb{R}_+, E)$).

3 Framework and results

3.1 Typical framework: The Freidlin-Wentzell theory

The classical small noise problem (see Section 6 of [37], Section 5.6 of [18], [23]) is studied for small noise diffusion processes given by the \mathbb{R}^d -valued (strong) solution $\mathbf{X}^\varepsilon = (\mathbf{X}_t^\varepsilon)_{t \geq 0}$ to the Stochastic Differential Equation (SDE)

$$\begin{cases} d\mathbf{X}_t^\varepsilon = b(\mathbf{X}_t^\varepsilon) dt + \sqrt{\varepsilon} \sigma(\mathbf{X}_t^\varepsilon) dW_t, \\ \mathbf{X}_0^\varepsilon = x_0 \end{cases} \quad (3.1)$$

where, as usual, $(W_t)_{t \geq 0}$ is an \mathbb{R}^m -valued Brownian motion; and the drift function $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ as well as the diffusion coefficients $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ are Lipschitz continuous. The process is parametrised by the noise amplitude $\varepsilon > 0$.

Let $T > 0$ be a given horizon. It is well-known that the SDE (3.1) satisfies a Large Deviations Principle in $C([0, T], \mathbb{R}^d)$ with good rate function (in the case of non-degenerate noise) given by

$$I_{[0, T]}(\mathbf{x}) = \frac{1}{2} \int_0^T |\dot{\mathbf{x}}_t - b(\mathbf{x}_t)|_{g(\mathbf{x}_t)}^2 dt.$$

The latter rate function is as usual a lower semi-continuous functional on $C([0, T], \mathbb{R}^d)$, finite-valued on a Sobolev sub-space. In the above, $|\cdot|_g$ denotes the (Riemannian) l^2 -norm associated with the metric

$$g \stackrel{\text{def.}}{=} (\sigma \sigma^T)^{-1}.$$

Clearly, the rate function has an additive structure in the sense that for any $T' \geq T$, one has

$$I_{[0, T']} = I_{[0, T]} + I_{[T, T]}, \quad (3.2)$$

where in the above the two functions of the right hand side have being trivially extended to lower semi-continuous functions on $C([0, T'], \mathbb{R}^d)$.

In our context, the *reference* set A will typically contain the attractor set associated with the limiting ordinary differential equation

$$\dot{\mathbf{x}}_t = b(\mathbf{x}_t)$$

and the considered initial condition x_0 in the sense that $\mathbf{x}_t \in A$ for all t large enough.

Given a reference set A , and an initial and final points x, y it is possible to define the (*quasi-potential*) *cost function* $U : (\mathbb{R}^d)^2 \rightarrow \mathbb{R}_+$ as the minimum of the rate function:

$$U(x, y) = U^A(x, y) \stackrel{\text{def.}}{=} \inf_{T > 0} \inf_{\substack{\mathbf{x} \in C([0, T], \mathbb{R}^d) \setminus A \\ \mathbf{x}_0 = x \\ \mathbf{x}_T = y}} I_{[0, T]}(\mathbf{x});$$

the latter quantifies *how unlikely* is a trajectory deviating from the zero-noise solution $\dot{x}_t = b(x_t)$ in order to go from x to y . The quasi-potential has a geometric, time-free² expression:

$$U(x, y) = \inf_{\substack{\mathbf{x} \in C([0, T], \mathbb{R}^d) \\ \mathbf{x}_0 = x, \mathbf{x}_1 = y \\ \mathbf{x}_\theta \notin A}} \int_0^1 \left(\left| \frac{d\mathbf{x}_\theta}{d\theta} \right|_{g(\mathbf{x}_\theta)} \left| b(\mathbf{x}_\theta) \right|_{g(\mathbf{x}_\theta)} - \left\langle \frac{d\mathbf{x}_\theta}{d\theta}, b(\mathbf{x}_\theta) \right\rangle_{g(\mathbf{x}_\theta)} \right) d\theta,$$

²It can be derived by minimizing over an arbitrary time change. A simple calculation enables to double-check that the expression is independent of the path parametrization $\theta \mapsto \mathbf{x}_\theta$.

and thus can be interpreted as an oriented type of length hence satisfying the triangle inequality

$$U(x, z) \leq U(x, y) + U(y, z) \quad \forall x, y, z \in \mathbb{R}^d \setminus \dot{A}.$$

Classically (see [24]), one assumes that the ODE (3.1) with $\varepsilon = 0$, $\dot{x} = b(x)$, as a unique attractor $x_A = \lim_{t \rightarrow +\infty} x_t$ and the quasi-potential is sometimes defined by the function $U(x) = U(x_A, x)$. The latter defines the iso-likely exit levels from the attractor defined by x_A . In that specific context one usually chooses A to be a small neighborhood of x_A . In the present work, the quasi-potential will rather refer to the two points function U .

Upon some assumptions that will be detailed in Section 3.3, the quasi-potential provides the asymptotic behavior of the probability of hitting a specific set B before the attractor A :

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}_{x_0} [\tau_B(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)] = - \inf_{y \in B} U(x_0, y).$$

3.2 Rare event simulation: The splitting algorithms and their asymptotic variances

Adaptive Multilevel Splitting (AMS)

The AMS algorithm is described assuming one can simulate the underlying time-continuous diffusion process $(\mathbf{X}_t)_{t \geq 0}$, but of course in practice the latter has to be discretized and the AMS algorithm needs to be slightly adjusted in consequence.

The AMS algorithm can be succinctly but rigorously described as follows.

Initially, N clones (a.k.a. *particles*) are simulated independently using the underlying Markov dynamics until they reach the *reference* set A . They are denoted $(\mathbf{X}^{(1,0)}, \dots, \mathbf{X}^{(N,0)})$.

First i), the iteration index of the algorithm is denoted by the index $i \geq 1$ and is used to enumerate the finite number of levels $L_1 < \dots < L_i < \dots < L_{I_l^N}$ at which a branching event (that is the killing and splitting of well-chosen clones) occurs. I_l^N denotes the first iteration at which all clones reach the target set $\{\xi \geq l\}$ for some given l . At each iteration i , the sample size

Second ii), the level L_i is computed as the k -th order statistics of the *scores* associated with each clones; the 'score' being given by the maximum of the importance function ξ over the clones' trajectories, that is: $\sup_{t \leq \tau_A(\mathbf{X}^{(n,i-1)})} \xi(\mathbf{X}_t^{(n,i-1)})$. Then the k clones with lowest scores are killed, and k new clones are uniformly³ randomly picked among the $N - k$ survivors (*selection step*). When $k = 1$ the AMS algorithm is called the *last particle algorithm*.

Third iii), each newly created clone is modified (*mutation step*) using independent simulations of the underlying Markov dynamics, *with initial condition the first hitting time of $\{\xi \geq L_j\}$* , and until the reference set A is reached.

This algorithm yields as an output two main estimators. First,

$$p_{l,\text{ams}}^N \stackrel{\text{def.}}{=} \left[\frac{N - k}{N} \right]^{I_l^N} \quad (3.3)$$

³*E.g.* with a multinomial or other permutation invariant distribution

estimates (without bias) the rare event probability $\mathbb{P}(\tau_l(\mathbf{X}) < \tau_A(\mathbf{X}))$; second the empirical distribution of the clones' trajectories at iteration I_l^N denoted

$$\eta_{l,\text{ams}}^{N,\text{path}} \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{n=1}^N \delta_{\mathbf{X}^{(n, I_l^N)}} \quad (3.4)$$

estimates the conditional distribution $\text{Law}(\mathbf{X} \mid \tau_l(\mathbf{X}) < \tau_A(\mathbf{X}))$. It is also known (see [7]), as is also true with any Sequential Monte Carlo algorithm, that non-normalized quantities are unbiased, that is $\mathbb{E} \left[p_{l,\text{ams}}^N \eta_{l,\text{ams}}^{N,\text{path}}(\Psi) \right] = \mathbb{E} \left[\Psi(\mathbf{X}) \mathbf{1}_{\tau_l(\mathbf{X}) < \tau_A(\mathbf{X})} \right]$ for any pathwise measurable bounded function Ψ .

Under some mild (non-minimal) assumptions (called Assumptions 1, 2 and 3 in [7]), the above estimators are asymptotically normal.

Theorem 3.1 (Asymptotic normality of AMS, [7]). *Let $k = 1$. Assume that:*

- $(\mathbf{X}_t)_{t \geq 0}$ is a Feller Markov process taking values in a locally compact space Polish E , with $\xi(\mathbf{X}_0) \geq l_0 > -\infty$ almost surely.
- For all $x \in \{\xi \in [l_0, l_B]\}$, entrance times of closures and interiors of sets of interest are the same, that is $\mathbb{P}_x \left[\tau_{\xi(x)}^-(\mathbf{X}) = \tau_{\xi(x)}^+(\mathbf{X}) \right] = \mathbb{P}_x \left[\tau_A^-(\mathbf{X}) = \tau_A^+(\mathbf{X}) \right] = 1$,
- $\mathbb{P}_x [\tau_{l_B}(\mathbf{X}) < \tau_A(\mathbf{X})]$ is uniformly bounded away from 0 for $x \in \{\xi \geq l_0\}$.

Then the estimators $p_{l,\text{ams}}^N$ and $\eta_{l,\text{ams}}^{N,\text{path}}$ are asymptotically normal when $N \rightarrow +\infty$ with $O(1/N)$ variance. Moreover, the asymptotic variance of $\sqrt{N} p_{l,\text{ams}}^N$ is given by

$$\sigma_{l,\text{ams}}^2 \stackrel{\text{def.}}{=} -(p_l)^2 \ln p_l + 2 \int_{l_0}^l \text{Var}_{\eta_{l'}}(q_{l'}) p_{l'} d(-p_{l'}). \quad (3.5)$$

In [7], a similar formula is given for the large sample size variance of all estimators, see Corollary 2.8 and Theorem 2.13. The extension to the case $k > 1$ under the same assumptions, where k is fixed and $N \rightarrow +\infty$ can be obtained using the results of [10].

Note that the considered assumptions, although quite mild, are probably not minimal. In particular the Feller assumption may not be necessary and the third assumption may be replaced by the strict positivity $\mathbb{P}_{x_0} [\tau_{l_B}(\mathbf{X}) < \tau_A(\mathbf{X})] > 0$, see Appendix F in [7].

Although the AMS algorithm has been originally presented as an *adaptive* Sequential Monte Carlo method, it is more convenient, in order to understand its unbiasedness structure (typical of non-adaptive SMC methods) and to compute formally the variance formula (3.5), to recast it as a *classical, non-adaptive, time continuous Sequential Monte Carlo model*. In order to do so the role of *time* must be played by the continuum of possible levels in $[l_0, l_B]$, see Section 3 in [7]. The AMS algorithm can then be interpreted as a Fleming-Viot process which possesses the Feynman-Kac structure promoted in the work of P. Del Moral (see [17, 15]). In the next section, we present a fixed levels version of the AMS algorithm which enables to formally justify the latter ideas. A short review of those variants of fixed effort splitting algorithms is provided in [9].

Fixed Multilevel Splitting

Consider now a fixed number J of levels $\ell_1 < \dots < \ell_J = l_B$, with final level l_B . Those levels are deterministic and chosen beforehand. The Fixed Multilevel Splitting algorithm is a standard Sequential Monte Carlo method, with a Del Moral-Feynman-Kac structure as in [15] section 12.2, or see also [6].

This algorithm can be succinctly described as follows. Initially, N clones (a.k.a. *particles*) are simulated independently using the underlying Markov dynamics until they reach the *reference set* A . They are denoted $(\mathbf{X}^{(1,i)}, \dots, \mathbf{X}^{(N,i)})$ for $i = 0$.

First, i) the iteration index is given by the index j of the considered level in the ladder $\ell_1 < \dots < \ell_j < \dots < \ell_J = l_B$.

Second ii), the *selection step* is made using the following 0 or 1 weights: the K_j clones whose *score* given the maximum of the importance function ξ over the clones' trajectories – that is: $\sup_{t \leq \tau_A(\mathbf{X}^{(n,j-1)})} \xi(\mathbf{X}_t^{(n;j-1)})$ – fail to reach level ℓ_j are killed (weight 0). K_j new clones are then randomly picked (e.g. multinomial distribution⁴) among survivors. If $K_j = N$, the algorithm is stopped, and the rare event probability is estimated by 0.

Third iii), each new created clone is modified (*mutation step*) by simulating independently the underlying Markov dynamics, *with initial condition the first hitting point of the level set* $\{\xi \geq \ell_j\}$, up until reaching the *reference set* A .

This Fixed Level algorithm yields as an output similar estimators as the Adaptive algorithm; the estimator of the small probability p_{ℓ_j} is obtained for instance *mutatis mutandis* by

$$p_{\ell_j, \text{fms}}^N \stackrel{\text{def.}}{=} \prod_{j'=1}^j \frac{N - K_{j'}}{N};$$

and the empirical distribution of the clones' trajectories at iteration j , that is

$$\eta_{\ell_j, \text{fms}}^{N, \text{path}} \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{n=1}^N \delta_{\mathbf{X}^{(n,j)}}$$

estimates, as for AMS, the conditional distribution Law $(\mathbf{X} \mid \tau_{\ell_j}(\mathbf{X}) < \tau_A(\mathbf{X}))$. It is also well known and easy to check –as is always true with Sequential Monte Carlo strategies – that non-normalized quantities are unbiased, that is $\mathbb{E} \left[p_{\ell_j, \text{fms}}^N \eta_{\ell_j, \text{fms}}^{N, \text{path}}(\Psi) \right] = \mathbb{E} \left[\Psi(\mathbf{X}) \mathbf{1}_{\tau_{\ell_j}(\mathbf{X}) < \tau_A(\mathbf{X})} \right]$ for any Ψ bounded measurable pathwise test function.

The AMS algorithm can then be obtained as a *limit of the Fixed Level algorithm* when $J \rightarrow +\infty$ with $\max_j \ell_{j+1} - \ell_j \rightarrow 0$; at least in a slightly formal way. To do so, consider in the Fixed Multilevel Splitting algorithm, the random sequence of levels

$$L_1 < \dots < L_i < \dots < L_{I_{\text{iter}}}$$

defined as the subsequence in the sequence $\{\ell_1 < \dots < \ell_J\}$ for which at least one killing event occur. The Fixed level algorithm can then be equivalently reformulated by iterating on the index $i = 1 \dots I_{l_B}$ instead of $j = 1 \dots J$ with $\ell_J = l_B$; I_{l_B} denoting the total number of killing (or branching) events required so that all clones have reached the level $\{\xi \geq \ell_J = l_B\}$. When $\max_j \ell_{j+1} - \ell_j \rightarrow 0$, then $K_{j'} \in \{0, 1\}$ with probability tending to 1. The Adaptive Multilevel Splitting for $k = 1$ (the 'last particle' case) is thus simply obtained by taking the limit $J \rightarrow +\infty$ with $\max_j \ell_{j+1} - \ell_j \rightarrow 0$, or even more simply, by formally removing the constraints that $L_i \in \{\ell_1, \dots, \ell_J\}$.

The case $k > 1$ can be formulated in a similar fashion. One only needs to modify the Fixed Level algorithm above by triggering duplications of clones

⁴for a discussion on the different resampling options in that case see [32]

only when the number of surviving clones gets below $N - k$. The Adaptive Multilevel Splitting for any $k \geq 1$ is again simply obtained by taking the limit $J \rightarrow +\infty$ with $\max_j \ell_{j+1} - \ell_j \rightarrow 0$.

The asymptotic normality of the estimators of the Fixed Multilevel algorithm follows from the classical results in Section 9 of [15].

Theorem 3.2 (Asymptotic normality of Fixed Multilevel Splitting, [15]). *Let J and $\ell_1 < \dots < \ell_J$ be given, assume $\xi(\mathbf{X}_0^\varepsilon) \geq l_0$ almost surely, and assume that uniformly in the initial condition $x_0 \in \{\xi \in [l_0, \ell_J]\}$, the probability that \mathbf{X} reaches $\{\xi \geq \ell_J\}$ before A is bounded away from 0.*

Then the estimators $p_{l,\text{fms}}^N$, and $\eta_{l,\text{fms}}^{N,\text{path}}(f)$, for any test function f , are asymptotically normal when $N \rightarrow +\infty$ with $O(1/N)$ variance. Moreover, the asymptotic variance of $\sqrt{N} p_{\ell_J,\text{fms}}^N$ is given by

$$\sigma_{\ell_J,\text{fms}}^2 \stackrel{\text{def.}}{=} \sum_{j=1}^{J-1} \frac{p_{\ell_j}}{p_{\ell_{j-1}}} \left((p_{\ell_{j-1}})^2 - (p_{\ell_j})^2 \right) \text{Var}_{\eta_{\ell_j}}(q_{\ell_j}) + (p_{\ell_J})^2 \sum_{j=1}^J \left(\frac{p_{\ell_{j-1}}}{p_{\ell_j}} - 1 \right). \quad (3.6)$$

The above result should hold for $k > 1$ fixed, although a rigorous extension is not provided explicitly in the literature up to our knowledge. Note that the variance formula (3.6) is not provided explicitly in Section 9.4.2 of [15], see [9] for more comments.

It is then possible to derive formula (3.5) from (3.6) as follows:

Lemma 3.3. *Assume that the decreasing function $l \mapsto p_l$ is continuous, and that $\ell_J = l_B$ is fixed. Then one has*

$$\lim_{\substack{J \rightarrow +\infty \\ \max_j \ell_{j+1} - \ell_j \rightarrow 0}} \sigma_{\ell_B,\text{fms}}^2 = \sigma_{\ell_B,\text{ams}}^2.$$

3.3 Assumptions and basic consequences

Our results can be stated for a family of pathwise continuous time homogeneous Markov processes

$$\{(\mathbf{X}_t^\varepsilon)_{t \geq 0} : \varepsilon > 0\},$$

taking value in a Polish state space E . A denotes the *reference* set, and ξ the continuous importance function of interest. l_B denotes the level of the target set $B \stackrel{\text{def.}}{=} \{\xi \geq l_B\}$, the rare event of interest being $\{\tau_B < \tau_A\}$.

Remark 3.4. *Although not necessary, one can assume with a negligible loss of generality that for any initial condition and level l , the hitting time of the interior or closure of A or $\{\xi \leq l\}$ are the same $\tau_A^+(\mathbf{X}^\varepsilon) = \tau_A^-(\mathbf{X}^\varepsilon)$, and $\tau_l^+(\mathbf{X}^\varepsilon) = \tau_l^-(\mathbf{X}^\varepsilon)$, almost surely – avoiding any ambiguity in the precise definition of the stopping times.*

We will also need that \mathbf{X}^ε satisfies the strong Markov property (for its natural filtration) with respect to the stopping times $\tau_l^-(\mathbf{X}^\varepsilon)$, $l \in \mathbb{R}$.

For simplicity, we assume that the main rare event of interest is defined for a given deterministic initial condition with level greater than a reference l_0

$$\mathbf{X}_0^\varepsilon = x_0, \quad \xi(x_0) > l_0.$$

Remark 3.5. *Our setting and results can be easily generalized mutatis mutandis to a general initial distribution η_0^ε satisfying a Large Deviation Principle on the Polish space E . Such a generalization can be obtained by adding the rate function associated to the initial distribution to the rate function $I_{[0,T]}$ of the process on the time interval $[0, T]$ in the small noise asymptotics.*

The process \mathbf{X}^ε is assumed to satisfy on each time interval $[0, T]$ as $\varepsilon \rightarrow 0$ a large deviation principle on the Polish space $C([0, T], E)$ (endowed with uniform convergence) for some good rate function $I_{[0,T]}$. The large deviations estimates are (classically) assumed to be true uniformly (in a local sense) with respect to the initial condition. This is the content of our first assumption.

Assumption 1. *For each final time T and initial condition $x_0 \in E$, the family of processes $\{\mathbf{X}^\varepsilon\}_{\varepsilon \geq 0}$ satisfies a LDP in $C([0, T], E)$ with good rate function $I_{[0,T]}$. The LDP is locally uniform with respect to the initial condition, that is, for the upper bound:*

$$\limsup_{(x,\varepsilon) \rightarrow (x_0,0)} \varepsilon \ln \mathbb{P}_{\mathbf{x}_0^\varepsilon = x} [\mathbf{X}^\varepsilon \in C] \leq - \inf_C I_{[0,T]}$$

for any closed set $C \subset C([0, T], E)$; and similarly for the lower bound:

$$\liminf_{(x,\varepsilon) \rightarrow (x_0,0)} \varepsilon \ln \mathbb{P}_{\mathbf{x}_0^\varepsilon = x} [\mathbf{X}^\varepsilon \in O] \geq - \inf_O I_{[0,T]}$$

for any open set $O \subset C([0, T], E)$.

Classically, the uniform large deviations principle ensures that the Markovian property of the underlying process translates into the additivity property (3.2) of the rate function (using for instance the extended Varadhan lemmas detailed in Section 5.4):

Lemma 3.6. *Under Assumption 1, the family of rate functions $(I_{[T,T']})_{0 \leq T \leq T'}$ parametrized by time intervals satisfies the additivity property (3.2).*

We need now to define the quasi-potential avoiding A .

Definition 3.7. *Let us denote for each $x, y \in E$ and $A \subset E$*

$$U(x, y) = U^A(x, y) \stackrel{\text{def.}}{=} \inf_{\substack{T > 0, \mathbf{x} \in C([0, T], E \setminus A) \\ (\mathbf{x}_0, \mathbf{x}_T) = (x, y)}} I_{[0, T]}[\mathbf{x}],$$

the quasi-potential or likelihood-cost function to go from x to y while avoiding A .

Next, the following assumption is a very mild technical simplification that prevents degenerate cases in which the boundary of A may play a role in the definition of optimal trajectories.

Assumption 2. *For any $\delta > 0$, $T > 0$, and any $\mathbf{x} \in C([0, T], E \setminus \mathring{A})$ with $\mathbf{x}_0, \mathbf{x}_T \notin \partial A$, there exist $\mathbf{x}^\delta \in C([0, T], E \setminus \bar{A})$ with $\mathbf{x}_0 = \mathbf{x}_0^\delta$ and $\mathbf{x}_T = \mathbf{x}_T^\delta$ such that $I_{[0, T]}[\mathbf{x}^\delta] \leq I_{[0, T]}[\mathbf{x}] + \delta$.*

In short, Assumption 2 ensures that trajectories avoiding \mathring{A} can be modified to avoid \bar{A} at arbitrarily small cost. The role of this assumption is to make the definition of optimal costs using \bar{A} or \mathring{A} equivalent. Indeed, one immediately gets:

Remark 3.8. *Under Assumption 2, one has*

$$U^{\bar{A}} = U^{\mathring{A}}.$$

The present work will also resort to variants of the quasi-potential in which trajectories are restricted to lower level-sets $\{\xi \leq l\}$.

Definition 3.9. Let $\xi : E \rightarrow \mathbb{R}$ be continuous, and denote for each $l \in \mathbb{R}$, $x \in \{\xi \leq l\}$, $x, y \in \{\xi = l\}$ and $A \subset E$

$$U^{(l)}(x, y) \stackrel{\text{def.}}{=} U^{(l),A}(x, y) \stackrel{\text{def.}}{=} \inf_{\substack{T>0, \mathbf{x} \in C([0,T], \{\xi \leq l\} \setminus A) \\ (\mathbf{x}_0, \mathbf{x}_T) = (x, y)}} I_{[0,T]}[\mathbf{x}],$$

the quasi-potential or likelihood-cost function to go from x to y in $\{\xi \geq l\}$ while avoiding A . Note that:

$$U^{(l)} \geq U.$$

Remark 3.10. Under Assumption 2, one has

$$U^{(l),\bar{A}} = U^{(l),\dot{A}}.$$

It is also useful to remark that costs to reach level sets defined by U or $U^{(l)}$ are the same.

Lemma 3.11. For each level l and initial condition X_0 , one has by definition and additivity of the rate function

$$U(x_0, \{\xi \leq l\}) = U^{(l)}(x_0, \{\xi \leq l\})$$

Proof. Minimizers of $U(\cdot, \{\xi \leq l\})$ can be stopped at the first hitting time of $\{xi = l\}$ to obtain minimizers of $U^{(l)}(\cdot, \{\xi \leq l\})$. \square

We then need a finiteness and continuity assumption on the cost to reach level sets; broadly speaking, ensures that the cost to reach level sets is finite and the cost to 'infinitesimally increase' levels is zero.

Assumption 3. For any initial condition $x \in \{l_0 \leq \xi \leq l_B\}$, the cost $U(x, \{\xi \leq l_B\}) < +\infty$ is finite and the cost to immediately enter the open set $\{\xi > \xi(x)\}$ is zero. Formally:

$$\inf_{\substack{T>0, \mathbf{x}: \mathbf{x}_0=x, \\ \tau_{\xi(x)}^+(\mathbf{x})=0}} I_{[0,T]}[\mathbf{x}] = 0.$$

In particular, this assumption is the most important assumption required to obtain the continuity (with respect to level) of the cost to enter a level-set; as is stated in Lemma 3.14 below.

The most demanding assumption is the following. It is similar to Condition (16.22) in [4]. It implies in particular that the reference set A contains all the possible attractors of the dynamics the deterministic dynamics $\mathbf{X}^{\varepsilon=0}$ with initial condition x_0 .

Assumption 4. The process \mathbf{X}^ε with initial condition x_0 reaches the interior of the reference set \dot{A} with a probability exponentially close to 1 when $\varepsilon \rightarrow 0$; the associated rate being arbitrary for large enough times. Rigorously:

$$\limsup_{T \rightarrow +\infty} \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}_{\mathbf{X}_0^\varepsilon = x_0} [\tau_{\dot{A}}(\mathbf{X}^\varepsilon) > T] = -\infty.$$

Remark 3.12. Assumption 4 is not satisfied rigorously in many practical situations because practitioners usually do not include in A all the attractors, or even critical points, of the deterministic dynamics $\mathbf{X}^{\varepsilon=0}$; but only consider those 'close' to the initial condition. However one should remark that:

- When $\varepsilon \rightarrow 0$ attractors outside of A will considerably slow down the splitting algorithms, since some trajectories may be stuck in one of the latter for a very large time before reaching A . It is a practical argument that shows that Assumption 4, although perhaps not minimal, is not a superfluous assumption.
- It is possible to consider the formal limit $A \rightarrow \emptyset$ in the present work, the various minimization problems of interest – for instance defining the functions U , $U^{(l)}$, or Loss, still being well defined (and even sometimes continuous) in this limit.

The most important consequence of the above assumptions is the following lemma, which interprets the quasi-potential as the rate of vanishing of the rare event probability in the small noise limit.

Lemma 3.13. *Let Assumptions 1, 2, 3 and 4 hold true. Let $x \in \{\xi \geq l_0\}$, $l \in [l_0, l_B]$ be given and define*

$$q_l^\varepsilon(x) = \mathbb{P}_{\mathbf{X}_0^\varepsilon=x}[\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)],$$

that is the probability to reach the level l before A starting from x . Then:

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ x \rightarrow x_0}} -\varepsilon \log q_l^\varepsilon(x) = U(x_0, \{\xi = l\}) \stackrel{\text{def.}}{=} \inf_{\{y: \xi(y)=l\}} U(x_0, y).$$

This result is classical. We give a self-contained, warm-up proof adapted to the setting of this work in Section 5.3. The latter proof justifies the role of the proposed set of assumptions.

Finally an already mentioned consequence of the above assumptions is the continuity of costs to reach level sets. The proof is also postponed to Section 5.3.

Lemma 3.14. *Let Assumptions 1, 2, 3 and 4 hold true. For all $x \in \{l_0 \leq \xi \leq l_B\}$, the map $l \mapsto U(x, \{\xi \geq l\})$ is continuous on $[l_0, l_B]$.*

We can now explicitly state a simple example of conditions on ξ and on the finite dimensional SDE (3.1) under which the latter assumptions hold true.

Lemma 3.15. *Consider the SDE (3.1) taking values in \mathbb{R}^d . Assume $b = -\nabla V$ and that $\sigma = \sqrt{2} \text{Id}$.*

- 1) *Assume that ∇V is globally Lipschitz continuous, then Assumption 1 holds true.*
- 2) *Assume that $A = \{\xi \leq l_A\}$ with $l_A < l_0$, ξ is smooth and has no critical point that is $\nabla \xi \neq 0$ on the set $\{\xi \in [l_A - \delta, l_A + \delta]\}$ for some δ . Then Assumption 2 holds true.*
- 3) *Assume that ξ is smooth and has no critical point that is $\nabla \xi \neq 0$ on the set $\{\xi \in [l_0, l_B]\}$, then Assumption 3 holds true.*
- 4) *Assume that $|\nabla V|$ is bounded away from 0 on $\mathbb{R}^d \setminus A$. Then Assumption 4 holds true.*

Proof. Item 1) is the classical Freidlin-Wentzell LDP, see [18, 37].

Item 2) can be proved by using a diffeomorphism in the neighbourhood of ∂A that approximates the identity. Indeed, since ξ is smooth and has no critical point, A is a smooth domain, and one can locally in the neighbourhood of ∂A consider a smooth set of new coordinates of \mathbb{R}^d the form $(y, r) \in \partial A \times \mathbb{R}$ where $x = y + r \nabla_y \xi$. One can then set $\chi_\delta(x) = y + k_\delta(r) \nabla_y \xi$ where k_δ is smooth and strictly increasing, $k_\delta(0) > 0$, $k_\delta(r) = r$ outside $[-\delta, +\delta]$ and $\|k'_\delta\|_\infty < \delta$. Hence

χ_δ converges to the identity in C^1 . By construction if \mathbf{x} does not intersect \mathring{A} , then $\chi_\delta(\mathbf{x})$ does not intersect A ; yet on the other hand $I[\chi_\delta(\mathbf{x})] \rightarrow_{\delta \rightarrow 0} I[\mathbf{x}]$ and the result follows.

Item 3) is somehow similar to Item 2), yet much simpler. Indeed, just consider the trajectory $\mathbf{x}_t = t\nabla_{x_0}\xi$ and the result follows.

Item 4) By assumption $|\nabla V| \geq \kappa > 0$ on set $\mathbb{R}^d \setminus A$. The classical Freidlin-Wentzell formula for the rate function can be rewritten in the form:

$$\begin{aligned} I_{[0,T]}[\mathbf{x}] &= \frac{1}{4} \int_0^T |\dot{\mathbf{x}}|^2 dt + \frac{1}{4} \int_0^T |\nabla V(\mathbf{x}_t)|^2 dt + \frac{1}{2} (V(\mathbf{x}_t) - V(\mathbf{x}_0)) \\ &\geq \frac{1}{4} \kappa^2 T - \frac{1}{2} V(\mathbf{x}_0) \xrightarrow{T \rightarrow +\infty} +\infty \end{aligned}$$

□

Remark 3.16. *Although it is not done in practice, it might be interesting to include in the reference set A the subset $\{|\nabla V| \leq \kappa\}$ for a well-chosen small κ , see Remark 3.12.*

3.4 Main results

We can now state rigorously the main results of the present work, whose proof is postponed to Section 5. The first (and most prominent) result provides the small noise asymptotics of the (large sample size asymptotic) variance formula $(\sigma_{l_B, \text{ams}}^\varepsilon)^2$ of the AMS rare event probability estimator $p_{l_B, \text{ams}}^{\varepsilon, N}$, as defined by (3.5). We have recalled in Section 3.2 that, under mild assumptions, one has indeed $(\sigma_{l_B, \text{ams}}^\varepsilon)^2 = \lim_{N \rightarrow +\infty} \frac{1}{N} \text{Var}(p_{l_B, \text{ams}}^{\varepsilon, N})$.

Theorem 3.17 (Small-noise asymptotics of AMS fluctuations). *Let the variance of an AMS probability estimator $\sigma_{l_B, \text{ams}}^\varepsilon$ be defined by (3.5). Under Assumptions 1, 2, 3 and 4 the following holds true:*

$$\begin{aligned} &\lim_{\varepsilon \rightarrow 0} \varepsilon \log \left[(\sigma_{l_B, \text{ams}}^\varepsilon)^2 / p_{l_B}^2 \right] \\ &= \sup_{l \in [\xi(x_0), l_B]} \lim_{\varepsilon \rightarrow 0} \varepsilon \log \text{Var}_{\eta_{\tilde{l}}} (q_{l_B}^\varepsilon p_l^\varepsilon / p_{l_B}^\varepsilon) = \sup_{l \in [\xi(x_0), l_B]} \text{Loss}(l), \end{aligned}$$

where the loss functions is defined by:

$$\text{Loss}(l) \stackrel{\text{def.}}{=} 2U(x_0, \{\xi = l_B\}) - \inf_{\{\xi=l\}} \left[U^{(l)}(x_0, \cdot) + 2U(\cdot, \{\xi = l_B\}) \right] - U(x_0, \{\xi = l\}). \quad (3.7)$$

Moreover there exist at least one critical level $l_* \in]\xi(x_0), l_B[$ such that $\sup_{l \in [\xi(x_0), l_B]} \text{Loss}(l) = \text{Loss}(l_*)$.

The most important result associated with the above theorem is the following sufficient condition for *weak asymptotic efficiency*, which is definition the vanishing of the logarithmic equivalent stated in Theorem 3.17.

Theorem 3.18. *For all $l \in [\xi(x_0), l_B]$, the loss function (3.7) is non-negative: $\text{Loss}(l) \geq 0$. If for some $l \in [\xi(x_0), l_B]$, the initial minimal cost $U(x_0, \{\xi = l\})$ and the final minimal cost $U(\{\xi = l\}, \{\xi = l_B\})$ are attained by a same state $x_*(l) \in \{\xi = l\}$, that is*

$$U(x_0, x_*(l)) = U(x_0, \{\xi = l\})$$

and

$$U(x_*(l), \{\xi = l_B\}) = U(\{\xi = l\}, \{\xi = l_B\}),$$

then the loss function vanishes $Loss(l) = 0$.

The above theorem gives a sufficient criterion ensuring that the AMS algorithm is weakly asymptotically efficient, in the sense that $Loss(l) = 0$ for all $l \in [\xi(x_0), l_B]$. The interpretation and the geometric visualisation of the loss function and of conditions ensuring weak asymptotic efficiency will be discussed in Section 4. In particular, it will be shown that a *necessary* condition for weak asymptotic efficiency is that the minimum of the quasi-potential from the initial condition $U(x_0, \cdot)$ coincide with a state of the optimal Freidlin-Wentzell path from x_0 to B (the instanton).

Proof of Theorem 3.18. Since $U(x_0, \{\xi = l\}) = U^{(l)}(x_0, \{\xi = l\})$, one has on $\{\xi = l\}$:

$$U^{(l)}(x_0, \cdot) \leq 2U^{(l)}(x_0, \cdot) - U(x_0, \{\xi = l\}).$$

One can then remark that

$$\inf_{\{\xi=l\}} \left[2U^{(l)}(x_0, \cdot) + 2U(\cdot, \{\xi = l_B\}) \right] = 2U(x_0, \{\xi = l_B\}),$$

and combining the two last equations we get the positivity of the loss function.

Now, the existence of $x_*(l)$ ensures that for all $x \in \{\xi = l\}$, first i) $U^{(l)}(x_0, x) \geq U(x_0, x_*(l))$ and second ii) $U(x, \{\xi = l_B\}) \geq U(x_*(l), \{\xi = l_B\})$. This shows that the minimization in the definition of the loss functions is attained for $x = x_*(l)$, leading to $Loss(l) = 0$. \square

We also obtain a similar result for the fixed level variant. In what follows, we will use the abuse of notation

$$U(x, \ell) = U(x, \{\xi = \ell\})$$

which is clearer in that context to keep track of the different levels.

Theorem 3.19 (Small-noise asymptotics of FMS fluctuations). *Let ℓ_1, \dots, ℓ_J denotes a fixed sequence of levels with ℓ_J . Let the variance of a FMS probability estimator $(\sigma_{\ell_J, \text{fms}}^\varepsilon)^2$ be defined by (3.6). Under Assumptions 1, 2, 3 and 4, the following holds true:*

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \left[(\sigma_{\ell_J, \text{SMC}}^\varepsilon)^2 / p_{\ell_J, \varepsilon}^2 \right] = \max(C_1, C_2) \geq 0,$$

where

$$C_1 \stackrel{\text{def.}}{=} 2U(x_0, \ell_J) - \min_{1 \leq j \leq J-1} \left[\inf_{\{\xi=\ell_j\}} \left(U^{(\ell_j)}(x_0, \cdot) + 2U(\cdot, \ell_J) \right) + U(x_0, \ell_{j-1}) \right],$$

and

$$C_2 \stackrel{\text{def.}}{=} \max_{1 \leq j \leq J} [U(x_0, \ell_j) - U(x_0, \ell_{j-1})] \geq 0.$$

It is worth noticing the following:

Lemma 3.20. *The quantity C_1 in Theorem 3.19 satisfies:*

$$\begin{aligned} 0 &\leq \min_{1 \leq j \leq J-1} U(x_0, \ell_j) - U(x_0, \ell_{j-1}) \\ &\leq C_1 - \min_{1 \leq j \leq J-1} Loss(\ell_j) \leq \max_{1 \leq j \leq J-1} U(x_0, \ell_j) - U(x_0, \ell_{j-1}) \end{aligned}$$

Proof. The lower bound comes from

$$\begin{aligned} & \min_{1 \leq j \leq J-1} \left[\inf_{\{\xi = \ell_j\}} \left(U^{(\ell_j)}(x_0, \cdot) + 2U(\cdot, \ell_j) \right) + U(x_0, \ell_{j-1}) \right] + \min_{1 \leq j \leq J-1} [U(x_0, \ell_j) - U(x_0, \ell_{j-1})] \\ & \leq \min_{1 \leq j \leq J-1} \left[\inf_{\{\xi = \ell_j\}} \left(U^{(\ell_j)}(x_0, \cdot) + 2U(\cdot, \ell_j) \right) + U(x_0, \ell_j) \right] \end{aligned}$$

while the upper bound comes from

$$\begin{aligned} & \min_{1 \leq j \leq J-1} \left[\inf_{\{\xi = \ell_j\}} \left(U^{(\ell_j)}(x_0, \cdot) + 2U(\cdot, \ell_j) \right) + U(x_0, \ell_{j-1}) \right] \\ & \geq \min_{1 \leq j \leq J-1} \left[\inf_{\{\xi = \ell_j\}} \left(U^{(\ell_j)}(x_0, \cdot) + 2U(\cdot, \ell_j) \right) + U(x_0, \ell_j) \right] - \max_{1 \leq j \leq J-1} [U(x_0, \ell_j) - U(x_0, \ell_{j-1})] \end{aligned}$$

□

Note that in the above lemma the equality case $C_1 = \min_{1 \leq j \leq J-1} \text{Loss}(\ell_j)$ is satisfied at least if the difference of initial cost between two levels is constant, that is $U(x_0, \ell_j) - U(x_0, \ell_{j-1})$ is independent of j for $j = 1 \dots J$, and thus equal C_2 .

The previous lemma thus shows that the surplus of loss in the FMS case as compared to the AMS case exactly comes from the differences $U(x_0, \ell_j) - U(x_0, \ell_{j-1})$. The AMS and the FMS small noise asymptotic variance will be similar only if the latter are small as compared to the loss function. This requires to choose sufficiently many levels ℓ_j in the FMS algorithms.

As a corollary we obtain equality between the small noise asymptotic variance of the adaptive AMS and fixed level FMS algorithms, when the number of levels J tends to infinity. This shows that under the different assumptions used in this work (Assumptions 1, 2, 3 and 4 and the assumptions of Theorem 3.1), one can commute the $J \rightarrow +\infty$ and the $\varepsilon \rightarrow 0$ limit.

Corollary 3.21. *Assume $\ell_J = l_B$ is fixed and $J \rightarrow +\infty$ with $\max_j(\ell_j - \ell_{j-1}) \rightarrow 0$. Then under Assumptions 1, 2, 3 and 4*

$$\lim_{J \rightarrow +\infty} \lim_{\varepsilon \rightarrow 0} -\varepsilon \log \left((\sigma_{l_B, \text{FMS}}^\varepsilon)^2 \right) = \lim_{\varepsilon \rightarrow 0} -\varepsilon \log \left((\sigma_{l_B, \text{AMS}}^\varepsilon)^2 \right).$$

4 Interpretation and insights

This section is dedicated to the interpretation of main result of this work, namely the small-noise large-sample-size variance formulas (1.2) to (1.4) for the AMS algorithm presented in the introduction, and then stated rigorously in Theorem 3.17 and Theorem 3.18.

Throughout this section we will use the notation $B = \{\xi \geq l_B\}$.

4.1 The variance formula 1.2

A result already mentioned in the introduction (formula (1.2)) says that the relative variance of the AMS estimator of the probability of interest $p_{l_B}^\varepsilon$ is equivalent – at large deviations regime and for the worst possible level $l = l_* \in]\xi(x_0), l_B[$ – to the variance of the following unbiased (theoretical) estimator of $p_{l_B}^\varepsilon$:

$$\text{Var} \left[p_l^\varepsilon q_{l_B}^\varepsilon (X_l^\varepsilon) \right],$$

where $X_l^\varepsilon \sim \eta_l^\varepsilon$ is distributed according to η_l^ε , the distribution of the first hitting place of $\{\xi = l\}$ by a trajectory (conditioned to happen before reaching A).

The large deviations estimates obtained in this paper suggests a decomposition of the above variance into the product of p_l^ε on the one hand, and $\gamma_l^\varepsilon ((q_{l_B}^\varepsilon)^2)$ on the other hand – where $\gamma_l^\varepsilon = p_l^\varepsilon \times \eta_l^\varepsilon$ is the non-normalized version of the conditional distribution η_l^ε . This will be discussed in Section 4.3. We will rather now comment, quite informally, the behavior of the conditional distribution η_l^ε , of the probability p_l^ε and their relations to the variance formula above.

The conditional distribution η_l^ε is by definition concentrated in areas of $\{\xi = l\}$ that are the most likely to be reached by trajectories (before A). Unfortunately, the importance function ξ usually *misleads* (so to speak) trajectories, in the sense that those likely areas of $\{\xi = l\}$ may have a very small remaining probability to hit B . In that scenario, we can informally decompose trajectories into two types of events: i) those *typical but unuseful trajectories* that hit $\{\xi = l\}$ in the most likely areas but have a very small remaining probability $q_{l_B}^\varepsilon$ to reach the final level B (before A), and ii) rare *lucky trajectories* that are outliers with a relatively large remaining probability $q_{l_B}^\varepsilon$ to reach the final level B (before A).

By definition, the 'typical but unuseful trajectories' of η_l^ε are involved in *underestimation* of the final probability

$$\eta_l^\varepsilon(q_{l_B}^\varepsilon \mid \text{typical}) \ll \frac{p_{l_B}^\varepsilon}{p_l^\varepsilon};$$

while the 'lucky trajectories' ii) are involved in *overestimation*

$$\eta_l^\varepsilon(q_{l_B}^\varepsilon \mid \text{lucky}) \gg \frac{p_{l_B}^\varepsilon}{p_l^\varepsilon}.$$

Note that in an AMS algorithm, clones sampling 'typical but unuseful trajectories' will have little or no offspring, while clones sampling 'lucky' will have many offsprings and will chiefly contribute to the final estimation.

In that scenario, the quantity $\text{var}_{\eta_l^\varepsilon}(q_{l_B}^\varepsilon)$ which quantifies the fluctuations of the function $q_{l_B}^\varepsilon$ is dominated by 'lucky trajectories' since a large remaining probability will *mainly* contribute to variance through the average square $\eta_l^\varepsilon((q_{l_B}^\varepsilon)^2)$.

On the other hand, the total mass of 'typical but unuseful trajectories' is related to p_l^ε : indeed, the larger p_l^ε is, the easier it is to reach $\{\xi = l\}$, and the more 'typical but unuseful trajectories' will happen. This idea will be made rigorous using the 'underestimation' part of the loss function in Section 4.3 below.

This informally described phenomenon is somehow similar to what happens with a naive i.i.d. Monte-Carlo. Let p denote the small target probability to be estimated. Assume one estimates $\varepsilon \ll p$ (with probability $1 - p_\varepsilon$), or success 1 (very rarely, with probability p_ε); where $p_\varepsilon = (p - \varepsilon)/(1 - \varepsilon) \simeq p$. The overall relative variance is only driven by the rare but highly overestimating value 1.

It should also be noted that in those considerations, the precise value of the underestimation does not influence variance. To fix ideas in the above simple i.i.d. example, ε does *not* impact the order of the variance $\simeq p(1 - p)$.

4.2 A disclaimer about variance

As said before, it is well-known that an unbiased estimation of a rare event probability usually leads to a typical systematic *underestimation* (an 'apparent bias'), as well as to a variance driven by rare *overestimations* (where some erroneously large values of the estimator contribute strongly to the variance while being nonetheless rarely seen by the algorithm). This has been noted for instance in [12, 35, 25]. This can be seen as a limitations of the present analysis for in some practical cases, if for instance one is only interested in understanding the *typical underestimation of an AMS algorithm* (as said in the end of the last section this information is lost in our variance analysis) when using the algorithm with limited number of particles.

However, our analysis is restricted to a regime where N can be taken to infinity *before* $\varepsilon \rightarrow 0$. In that perspective, there are sufficiently many clones to obtain a Central Limit Theorem [7], and the variance is a legitimate quantity to discuss the fluctuations of the algorithm. This situation happens in practice when the importance function is sufficiently good to enable a consistent sampling of paths close enough to rate-funtion-optimal trajectories.

4.3 The loss function: discussion

The loss function can be interpreted using a decomposition into an *underestimation* part and a *overestimation* part, in the spirit of the discussion of Section 4.1. The underestimation part, denoted Loss_U , is associated with the large deviation equivalent of the factor $p_{l_*}^\varepsilon$ which satisfies

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \ln p_{l_*}^\varepsilon = -U(x_0, \{\xi = l\}).$$

The overestimation part, denoted Loss_O , is associated with the large deviation equivalent of the factor $\gamma_l^\varepsilon ((q_{l_B}^\varepsilon)^2)$ which satisfies

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \ln \gamma_l^\varepsilon ((q_{l_B}^\varepsilon)^2) = - \inf_{\{\xi=l\}} \left[U^{(l)}(x_0, \cdot) + 2U(\cdot, B) \right].$$

This decomposition is also motivated by the following to facts:

- Each part is non-negative and identically 0 when weak asymptotic efficiency is achieved.
- The conditions ensuring that each part is 0 are mostly independent in terms of the importance function ξ , as both are defined by two different minimization problems.

In order to be more precise, we can consider for each l a state $x_*(l) \in \{\xi = l\}$ such that

$$U(x_0, B) = U(x_0, x_*(l)) + U(x_*(l), B).$$

A continuous level-indexed path $l \mapsto x_*(l)$ satisfying the above condition is called an *instanton* in physics literature (*e.g.* [1]). Then one can consider the decomposition

$$\text{Loss}(l) = \text{Loss}_U(l) + \text{Loss}_O(l),$$

where we define

$$\text{Loss}_U(l) \stackrel{\text{def.}}{=} U(x_0, x_*(l)) - U(x_0, \{\xi = l\}) \geq 0,$$

as well as

$$\text{Loss}_O(l) \stackrel{\text{def.}}{=} U(x_0, x_*(l)) + 2U(x_*(l), B) - \inf_{\{\xi=l\}} \left[U^{(l)}(x_0, \cdot) + 2U(\cdot, B) \right] \geq 0.$$

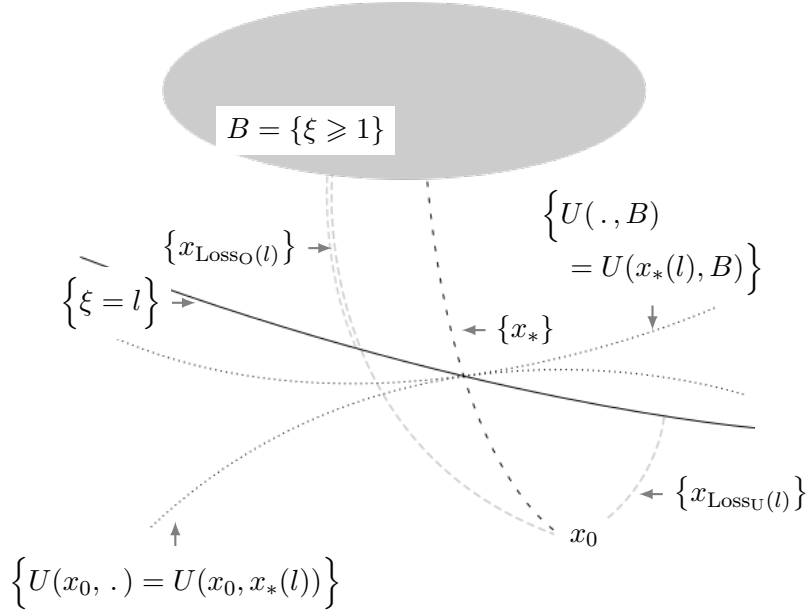


Figure 2: Graphical interpretation of the main small noise variance formula (1.4). The path $\{x_*\}$ is the minimizer defining the cost $U(x_0, B)$. The path $\{x_{\text{Loss}_U(l)}\}$ represents 'typical but unuseful' trajectories, and is the minimizer of the cost $U(x_0, \{\xi = l\})$, that defines the underestimation part of the loss function. The path $\{x_{\text{Loss}_O(l)}\}$ represents 'lucky' trajectories, and is the minimizer of the cost $\inf_{\{\xi=l\}} U(x_0, \cdot) + 2U(\cdot, B)$ that defines the overestimation part of the loss function.

The underestimation part of the loss function $\text{Loss}_U(l)$ for a given l is obtained by the minimizing trajectories from x_0 to the set $\{\xi = l\}$ which yield the cost $U(x_0, \{\xi \geq l\})$. An example of such a trajectory (assuming it exists for simplicity) is denoted $\mathbf{x}_{\text{Loss}_U(l)}$ and is depicted in Figure 2 and 3. A first result is that $\text{Loss}_U(l) = 0$ if and only if $U(x_0, \{\xi \geq l\}) = U(x_0, x_*(l))$, or equivalently, if and only if the level set $\xi = l$ is *above* the level set $\{U(x_0, \cdot) = U(x_0, x_*(l))\}$, see Figure 3.

The overestimation part of the loss $\text{Loss}_O(l)$ for a given l is characterized by trajectories minimizing the sum of the rate function from x_0 to the first hitting time of the level $\{\xi = l\}$ plus *twice* the rate function from the associated entrance point up to $\{\xi \geq l_B\}$. An example of such a trajectory denoted $\mathbf{x}_{\text{Loss}_O(l)}$ is depicted in Figure 2 and Figure 3. Note that $\text{Loss}_O(l) = 0$ if and only if the composed cost above is attained at the point $x_*(l)$, see again Figure 2. A *sufficient* condition (but not necessary) ensuring $\text{Loss}_O(l) = 0$ is that the level set $\xi = l$ is *below* the level set $\{U(\cdot, B) = U(x_*(l), B)\}$, see Figure 3. This condition is not necessary because contrary to the underestimation part, this overestimation part of the loss involves a competition between an initial cost from x_0 and a final cost up to B .

Note that the obtained sufficient condition involving $x_*(l)$ for weak asymptotic efficiency is exactly the one given in theorem 3.18. Under that condition, minimizer of either $\text{Loss}_U(l)$ or $\text{Loss}_O(l)$ can be identified with instantons minimizing the rate function among trajectories reaching B before A .

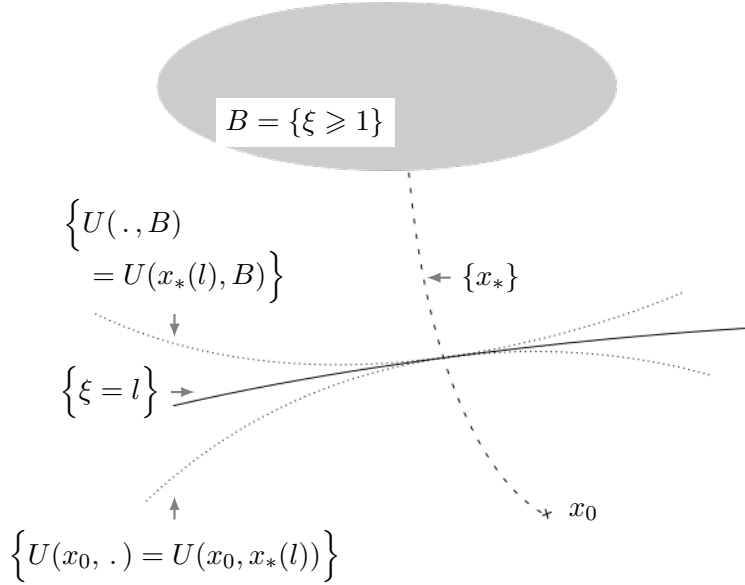


Figure 3: A sufficient condition for weak asymptotic efficiency, given by the simultaneous two conditions $\text{Loss}_U(l) = \text{Loss}_O(l) = 0$ for all $l \in [\xi(x_0), l_{\max}]$. Note that the level sets of the importance function ξ is *between* the level sets of the quasi-potential cost i) from the initial condition, and ii) up to the final set B . Moreover, $\{x_{\text{Loss}_O(l)}\} = \{x_{\text{Loss}_U(l)}\} = \{x_*\}$ is a (non necessarily unique) instanton.

Note also that the geometric conditions ensuring $\text{Loss}_O = \text{Loss}_U \equiv 0$ stated above are much weaker than the restriction that ξ is defined by the limiting committor function $\xi = \xi_*$. Note that, contrary to the latter, our conditions depends on the initial condition x_0 , so that a choice of weakly asymptotic efficient ξ for a given initial condition may not be so for a different initial condition. A related relaxed class of *optimal* ξ is given by *sub-solutions of the Hamilton-Jacobi equation* that underlies the rate function of the LDP satisfied by $(\mathbf{X}^\varepsilon)_{\varepsilon \geq 0}$, see [13, 14, 5, 4]. This will be discussed in a section below.

The probabilistic and algorithmic interpretation of $\text{Loss}_U(l)$ for a given l is the following. $\text{Loss}_U(l)$ will be large when the minimizers associated with the optimal cost $U(x_0, \{\xi = l\})$ do not correspond to $x_*(l)$ where x_* is an instanton (a minimizer associated with the global cost $U(x_0, \{\xi = l_B\})$). This means that the conditional distribution η_l^ε which is concentrated towards the minimizers associated with $U(x_0, \{\xi = l\})$ (by a standard Gibbs conditioning argument in large deviations theory). $\text{Loss}_U(l)$ thus quantifies the likelihood of 'typical but unuseful trajectories', as discussed in Section 4.1. The large deviation picture is depicted in Figure 3. Those trajectories are eventually associated with underestimation. We stress that the quantity Loss_U is not related to the specific value of this underestimation (the latter is rather encoded by the function $U(\cdot, B)$ evaluated at the minimizer associated with $U(x_0, \{\xi = l\})$, which does not appear in the definition of the loss function). In an AMS algorithm, $\text{Loss}_U(l)$ can be associated with the proportion of clones that will be quickly killed after having reached the level l .

The algorithmic interpretation of $\text{Loss}_O(l)$ for a given l is complementary. We have seen that it is associated with large values of the quantity $\gamma_l^\varepsilon ((q_{i_B}^\varepsilon)^2)$. As discussed in Section 4.1, this term can be associated with specific *lucky*

trajectories which turn out to be very contributive to variance in the end because they can reach the rare event set $\{\xi \geq l_B\}$ with a relatively large probability $q_{l_B}^\varepsilon$. This is quantified in the loss function by the variational problem:

$$\inf_{\{\xi=l\}} \left[U^{(l)}(x_0, \cdot) + 2U(\cdot, B) \right]$$

in which the final cost $U(\cdot, B)$ counts *twice* as compared to

$$\inf_{\{\xi=l\}} \left[U^{(l)}(x_0, \cdot) + U(\cdot, B) \right] = U(x_0, B)$$

which is minimized by global minimizers x_* . This implies that states with lower final cost $U(\cdot, B)$ are much preferred. These states defines the 'lucky outliers' discussed in Section 4.1. A typical lucky trajectory is depicted in Fig.3. In an AMS algorithm, Loss_U is associated with the overestimation by the small fraction of clones that are the most likely to reach B after having reached the level l .

4.4 Summary

To summarize our main results, we proved that there exists a *critical level* l_* that will contribute mostly to variance. This contribution is described by the distribution $\eta_{l_*}^\varepsilon$ of trajectories at the first hitting times of l_* and the associated probability $p_{l_*}^\varepsilon$ to reach level l_* (all before A). The logarithmic equivalent of the relative variance can also be decomposed into two independent non-negative terms. The first term, $\text{Loss}_U(l_*)$, quantifies the likelihood to have 'typical but unuseful' states in the distribution $\eta_{l_*}^\varepsilon$. The second term, $\text{Loss}_O(l_*) \geq 0$, quantifies the overestimation by outliers in $\eta_{l_*}^\varepsilon$ that are likely to eventually reach B . We also provide a simple geometric sufficient (resp. necessary and sufficient) condition on ξ depicted in Figures 2 and 3 such that $\text{Loss}_O \equiv 0$ or $\text{Loss}_U \equiv 0$.

4.5 Interpretation as a Hamilton-Jacobi sub-solution

We have thus obtained a simple geometric sufficient criteria for weak asymptotic efficiency ($\sup_l \text{Loss}(l) = 0$): for each l the level set $\xi = l$ lies in between the iso-cost set from x_0 , as depicted in Figure 3. Formally, this amounts to the the existence of an increasing function F (realized by the cost along x_*) such that

$$\begin{cases} F(\xi(x_0)) - F(\xi(y)) \leq U(x_0, y), & \forall y \in \{\xi \in [\xi(x_0), l_B]\} \\ F(\xi(x)) - F(\xi(y)) \leq U(x, y), & \forall (x, y) \in \{\xi \in [\xi(x_0), l_B]\} \times \{\xi \geq l_B\} \end{cases} \quad (4.1)$$

This expression is related to *sub-solutions of the Hamilton-Jacobi equation* that underlies the rate function of the LDP satisfied by $(\mathbf{X}^\varepsilon)_{\varepsilon \geq 0}$ (see *e.g.* [13, 14, 5, 4]). A sub-solution f is a function that satisfies the inequality $f(x) - f(y) \leq U(x, y)$ for all $x, y \in \{\xi \leq l_B\}$. This is however a much more demanding condition than (4.1) because: i) we do not need to compute the reparametrization F which be given by $F(\xi(x_*(l))) = U(x_*(l), B) + \text{cte}$; and ii) the sub-solution inequality need to be true *only* for initial points in the initial condition of \mathbf{X}^ε , and final points in B . See also Section 4.7 for comments on related work.

4.6 Possible practical consequences

Finally, there are several consequences of our results for practical purposes. We list them below. All are left for future work.

The first question is about estimating the variance (1.2) or equivalently the loss (1.4) which are the main result of this paper.

- An idea is to try to estimate a non-asymptotic (with respect to ε) form of the variance formula (1.2) after one realisation of the AMS algorithm. Variance estimation has recently been studied in [11, 27, 19]. The present work suggests that variance estimation can be considerably simplified at the cost of being accurate only asymptotically for large N and small ε . For instance, one can first estimate on the states x visited by the clones the probability $q_\varepsilon(x)$ to reach the final set $\{\xi \geq l_B\}$. This can be done using the genealogy of the clones and the formula (3.4). One can then proceed using the formula (1.2) by averaging those estimations over η_l^ε for each l , and then by minimizing on the level l .
- We will also remark in Section 5.8 that the loss function can be expressed as a new, simulable, rare event probability. The latter is the probability that an AMS algorithm with $N = 2$ clones succeeds in at most one iteration. This fact may be used to estimate $\sup_l \text{Loss}(l)$ using a secondary Monte Carlo rare event algorithm, more appropriate than the first one since it will purposely simulate the rare clones involved in the overestimation (that drives the variance) of the final probability.
- The variance estimation proposed in the first item above can in fact estimate the variance obtained with various importance function ξ , using, say, an AMS algorithm performed with a given reference ξ_0 . Indeed, the only quantity depending on ξ is the first hitting place η_l^ε associated with set $\{\xi \geq l\}$, but this can be estimated using the full genealogical estimator (3.4).
- There is currently a lot of effort in practical applications aimed at optimizing the importance function ξ in order to obtain reliable results, see for instance the references in the review for molecular simulation applications [34], or the paper [31] that uses a data-driven approach. Our analysis provides insights on the minimal conditions an importance function must satisfy in order to provide efficiency. In particular, instead of trying to exactly compute the committor function ξ_* , one may try, after a rare event simulation, to update ξ by trying to minimize the rough estimation of the variance as discussed in the previous item.

4.7 Comparison to previous work

The idea to analyse rare event (multi-level) splitting Monte Carlo simulation algorithms in a large deviation setting has been mainly developed by P. Dupuis and his co-authors.

In [13] a fixed multilevel splitting method with varying number of clones is studied. The main difference from our study is that the clones do not interact with each other through the splitting mechanism and the rate of splitting is given by the variations of the values of the level function. The authors then show that a sufficient condition to obtain asymptotic efficiency in a large deviation small noise limit is that the level function must be a sub-solution of the Hamilton-Jacobi problem associated with the Lagrangian formulation of the rate function (see Section ...). They also argue that this condition should be necessary. This type of condition on the level function (or importance function) is reminiscent to the type of condition required on the importance function in importance sampling in order to achieve asymptotic efficiency (see *e.g.* [21, 22, 20]); note that the importance function must in addition be a smooth sub-solution

which highlights the likely generic lack of robustness of importance sampling as compared to importance splitting.

As compared to our work, those result are more precise in the sens that the number of clones (although randomly varying) is finite, whereas our analysis is restricted to asymptotic (in terms of clones sample size) variance. However, our work suggest three (related with each other) improvements enabled by the adaptivity of levels and the fixed number of clones:

- Weak asymptotic efficiency can be achieved if *some parametrization* of the level function satisfies a certain weaker condition related to sub-solutions of the considered Hamilton-Jacobi problem. This condition is sufficient but not necessary.
- The notion of sub-solution is weaker: it only has to be one with respect to the support of the initial condition and the final target set (and not for every pair of points in space).
- We do not face the problem of explosion or implosion of the total number of clones. In [13] the splitting rate has to be tuned carefully – close to the inverse of probability of transitions between levels – to avoid such population size issues, even if the population size can still grow polynomially.

Similarly [14] studies a variant called RESTART which enables to reduce the trajectory length of most of the clones. The analysis is also improved (the notion of sub-solution is defined variationally instead of as a viscosity solution of a PDE) and then recapitulated in [4]. In [5], the authors nonetheless studies a splitting algorithm with fixed number of clones, but in dimension one only (asymptotic efficiency is then conditionless).

5 Large Deviations estimates and proof of the main result

This Section is devoted to the the large deviations estimates that eventually lead to the main result theorem 3.17.

5.1 Stopping times and topology

We start that a key technical remark on the semi-continuity of stopping times.

Lemma 5.1. *Let E be a Polish space, $T > 0$ be given, and $A, B \subset E$. The map $\mathbf{x} \mapsto \tau_A^-(\mathbf{x})$ (resp. τ_A^+) is lower (resp. upper) semi-continuous as a function of $C([0, T], E)$ with the topology of uniform convergence. In particular,*

$$\{\mathbf{x} \mid \tau_A^+(\mathbf{x}) < \tau_B^-(\mathbf{x})\}$$

is an open subset of $C([0, T], E)$.

Proof. See Lemma A.4 in [8]. □

5.2 Continuity of the cost to reach level sets

Let us prove Lemma 3.14. .

Right continuity. Let l_* be a given level. By definition of U , there is for each $\delta > 0$ a trajectory \mathbf{x}^δ with $U(x_0, \{\xi \geq l_*\}) + \delta \geq I_{[0, \tau_{l_*}^-(\mathbf{x}^\delta)]}[\mathbf{x}^\delta]$. By Assumption 3, it is possible to strictly extend \mathbf{x}^δ so that $\tau_{l_*}^-(\mathbf{x}^\delta) = \tau_{l_*}^+(\mathbf{x}^\delta)$

and $U(x_0, \{\xi \geq l_*\}) + 2\delta \geq I_{[0,T]}[\mathbf{x}^\delta]$. This implies that there exists a $l = \max_t \xi(\mathbf{x}_t^\delta) > l_*$ with $U(x_0, \{\xi \geq l_*\}) + 2\delta \geq U(x_0, \{\xi \geq l\})$. The result follows since δ is arbitrary and the function non-decreasing.

Left continuity. Let l be an arbitrary level, l_k be increasing with k and converging to l , and $\delta > 0$ be given, arbitrary. By definition of U and by Assumption 4, there is a time horizon $T > 0$ and a sequence of paths \mathbf{x}^k such that $I_{[0,T]}[\mathbf{x}^k] \leq U(x_0, \{\xi \leq l_k\}) + \delta$. Since I is a good rate function in Assumption 1 (lower semi-continuous with compact pull-back of closed bounded above intervals), one can extract a (uniformly) converging sub-sequence of paths such that $I_{[0,T]}[\mathbf{x}^\infty] \leq \liminf_k I_{[0,T]}[\mathbf{x}^k]$. By construction $\liminf_k I_{[0,T]}[\mathbf{x}^k] \leq \lim_k U(x_0, \{\xi \leq l_k\}) + \delta$, and by continuity of ξ , $\tau_l^-(\mathbf{x}^\infty \leq T)$ which implies $I_{[0,T]}[\mathbf{x}^\infty] \geq U(x_0, \{\xi \geq l\})$, hence the result.

5.3 Small noise asymptotics of the rare event

We can now turn to the proof of Lemma 3.13. We recall that the latter states that under Assumptions 1, 2, 3 and 4, then $\lim_{\varepsilon \rightarrow 0, x \rightarrow x_0} \varepsilon \ln \mathbb{P}_x [\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)] = -U(x_0, \{\xi \geq l\})$.

Proof of Lemma 3.13.

Upper bound. Assume $T > t_*$ given, arbitrary large. We can then consider the upper bound

$$\mathbb{P}_x [\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)] \leq \mathbb{P}_x [\tau_l^-(\mathbf{X}^\varepsilon) \leq \tau_A^+(\mathbf{X}^\varepsilon) \wedge T] + \mathbb{P} [\tau_A^+(\mathbf{X}^\varepsilon) \geq T]$$

On the one hand, using Assumption 4, one has $\limsup_\varepsilon \varepsilon \ln \mathbb{P} [\tau_A^+(\mathbf{X}^\varepsilon) \geq T] \leq -C_T$, with $\lim_{T \rightarrow \infty} C_T = +\infty$. On the other hand Lemma 5.1 ensures that the set $\{\mathbf{x} : \tau_l^-(\mathbf{x}) \leq \tau_A^+(\mathbf{x}) \wedge T\}$ is closed in $C([0, T], E)$. Thus, the uniform LDP upper bound yields

$$\begin{aligned} -\limsup_{\substack{\varepsilon \rightarrow 0 \\ x \rightarrow x_0}} \varepsilon \ln \mathbb{P}_x [\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)] &\geq \min \left[\inf_{\substack{\mathbf{x} \in C([0, T], E) \\ \mathbf{x}(0) = x_0 \\ \tau_l^-(\mathbf{x}) \leq T}} I_{[0, T]}[\mathbf{x}], C_T \right] \\ &\geq \min(U(x_0, \{\xi = l\}), C_T), \end{aligned}$$

C_T being arbitrary large.

Lower bound Let $l \geq 0$ be a given level, $\delta > 0$ be given, arbitrarily small, and let $x_0 \in \{\xi \geq l_0\}$ be an initial condition. Using the continuity property of Lemma 3.14, there is a small enough $h > 0$ such that $U(x_0, \{\xi = l + h\}) \leq U(x_0, \{\xi = l\}) + \delta/2$. One can thus construct a minimizing continuous path \mathbf{x}_* with $\mathbf{x}_*(0) = x_0$ and a time $t_* \geq 0$ such that

- i) $I_{[0, t_*]}[\mathbf{x}_*] \in [U(x_0, \{\xi = l\}), U(x_0, \{\xi = l\}) + \delta/2]$,
- ii) $\xi(\mathbf{x}_*(t_*)) > l$ so that $\tau_l^+(\mathbf{x}_*) < t_*$,
- iii) $\tau_A^-(\mathbf{x}_*) > t_*$.

Extending the trajectory \mathbf{x}_* with a $\delta/2$ minimizer of the rate function $I_{[t_*, T]}[\mathbf{x}_{t_*}]$, we obtain a path which satisfies for any arbitrary large final time $T > t_*$, $I_{[0, T]}[\mathbf{x}_*] \in [U(x_0, \{\xi = l\}), U(x_0, \{\xi = l\}) + \delta]$.

We can then consider the lower bound

$$\mathbb{P}_x [\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)] \geq \mathbb{P}_x [\tau_l^+(\mathbf{X}^\varepsilon) < t_* < \tau_A^-(\mathbf{X}^\varepsilon) \wedge T].$$

Applying the uniform LDP to the open (see Lemma 5.1) set

$$\{\mathbf{x} \in C([0, T], E) \mid \tau_l^+(\mathbf{x}) < t_* < \tau_A^-(\mathbf{x})\}$$

it yields

$$\begin{aligned} -\liminf_{\substack{\varepsilon \rightarrow 0 \\ x \rightarrow x_0}} \varepsilon \ln \mathbb{P}_x [\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)] &\leq \inf_{\substack{\mathbf{x} \in C([0, T], E) \\ \mathbf{x}(0) = x_0 \\ \tau_l^+(\mathbf{x}) < t_* < \tau_A^-(\mathbf{x})}} I_{[0, T]}[\mathbf{x}] \\ &\leq I_{[0, T]}[\mathbf{x}_*] \leq U(x_0, \{\xi = l\}) + \delta, \end{aligned}$$

δ being arbitrary small. □

5.4 Varadhan lemmas

In this section, we state and prove minor variants of the classical Varadhan lemmas, in the case where the potential function is general (ε -dependent and only measurable).

We start with the easier lower bound.

Lemma 5.2. *Assume $(\mathbf{X}^\varepsilon)_{\varepsilon > 0}$ satisfies a LDP with rate function I on a Polish state space. Let $(V_\varepsilon)_{\varepsilon > 0}$ be a family of measurable functions in $]-\infty, +\infty]$. Define for each state x the upper semi-continuous envelope $V_+(x)$ of the latter by:*

$$V_+(x) \stackrel{\text{def.}}{=} \limsup_{\substack{\varepsilon \rightarrow 0 \\ y \rightarrow x}} V_\varepsilon(y). \quad (5.1)$$

Then

$$-\liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{E} \left[e^{-\frac{1}{\varepsilon} V_\varepsilon(\mathbf{X}^\varepsilon)} \right] \leq \inf (I + V_+).$$

Proof. Similar to the usual proof of Varadhan's lemma lower bound.

First note that by definition of the envelope V_+ , for each state x and each $\delta > 0$, one can find an open neighbourhood $U_{x, \delta}$ such that

$$\limsup_{\varepsilon \rightarrow 0} \sup_{U_{x, \delta}} V_\varepsilon \leq V_+(x) + \delta/2.$$

Let $\delta > 0$ be given, arbitrarily small. By definition of the infimum, there is a x_δ such that

$$I(x_\delta) + V_+(x_\delta) \leq \inf [I + V_+] + \delta/2.$$

We can then consider the lower bound

$$\begin{aligned} \mathbb{E} \left[e^{-\frac{1}{\varepsilon} V_\varepsilon(\mathbf{X}^\varepsilon)} \right] &\geq \mathbb{E} \left[\mathbf{1}_{U_{x_\delta, \delta}}(\mathbf{X}^\varepsilon) e^{-\frac{1}{\varepsilon} V_\varepsilon(\mathbf{X}^\varepsilon)} \right] \\ &\geq \mathbb{P} [\mathbf{X}^\varepsilon \in U_{x_\delta, \delta}] e^{-V_+(x_\delta)/\varepsilon - \delta/2\varepsilon}, \end{aligned}$$

and using the LDP lower bound

$$\begin{aligned} -\liminf_{\varepsilon} \varepsilon \ln \mathbb{E} \left[e^{-\frac{1}{\varepsilon} V_\varepsilon(\mathbf{X}^\varepsilon)} \right] &\leq \inf_{U_{x_\delta, \delta}} I + V_+(x_\delta) + \delta/2 \\ &\leq I(x_\delta) + V_+(x_\delta) + \delta/2 \\ &\leq \inf [I + V_+] + \delta. \end{aligned}$$

□

Lemma 5.3. Assume $(\mathbf{X}^\varepsilon)_{\varepsilon>0}$ satisfies a LDP with good rate function I on a Polish state space. Let $(V_\varepsilon)_{\varepsilon>0}$ be a family of measurable functions in $[0, +\infty]$. Define for each state x the lower semi-continuous envelope of the latter by:

$$V_-(x) \stackrel{\text{def.}}{=} \liminf_{\substack{\varepsilon \rightarrow 0 \\ y \rightarrow x}} V_\varepsilon(y). \quad (5.2)$$

Then

$$-\limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{E} \left[e^{-\frac{1}{\varepsilon} V_\varepsilon(\mathbf{X}^\varepsilon)} \right] \geq \inf (I + V_-).$$

Proof. This is the classical proof of Varadhan upperbound based on the rate function goodness.

Let $\delta > 0$ be given, arbitrarily small. For each x , by lower semi-continuity of I and definition of V_- , we can find an open neighborhood $U_{x,\delta}$ such that:

$$\inf_{U_{x,\delta}} I \geq I(x) - \delta/2,$$

as well as

$$\liminf_{\varepsilon \rightarrow 0} \inf_{U_{x,\delta}} V_\varepsilon \geq V_-(x) - \delta/2.$$

Using the compactness of the level sets of I , one can choose a finite covering $U_{x_i,\delta}$, $i = 1 \dots I$ of the level set $\{x \mid I(x) \leq v_{\max}\}$ with v_{\max} arbitrarily large, and denote $U \stackrel{\text{def.}}{=} \bigcup_{i=1 \dots I} U_{x_i,\delta}$. Consider now the main upper bound:

$$\begin{aligned} \mathbb{E} \left[e^{-\frac{1}{\varepsilon} V_\varepsilon(\mathbf{X}^\varepsilon)} \right] &\leq \sum_{i=1}^I \mathbb{E} \left[\mathbf{1}_{\mathbf{X}^\varepsilon \in U_{x_i,\delta}} e^{-V_\varepsilon(\mathbf{X}^\varepsilon)/\varepsilon} \right] + \mathbb{P}[\mathbf{X}^\varepsilon \in U^c] \\ &\leq \sum_{i=1}^I \mathbb{P}[\mathbf{X}^\varepsilon \in \overline{U_{x_i,\delta}}] e^{-\inf_{\overline{U_{x_i,\delta}}} (V_\varepsilon)/\varepsilon} + \mathbb{P}[\mathbf{X}^\varepsilon \in U^c]. \end{aligned}$$

By construction of the neighbourhood $U_{x,\delta}$ and the LDP upper bound

$$-\limsup_{\varepsilon} \varepsilon \ln \mathbb{P}[\mathbf{X}^\varepsilon \in \overline{U_{x_i,\delta}}] e^{-\inf_{\overline{U_{x_i,\delta}}} (V_\varepsilon)/\varepsilon} \geq I(x) + V_-(x) - \delta,$$

so that

$$\begin{aligned} -\limsup_{\varepsilon} \varepsilon \ln \mathbb{E} \left[e^{-\frac{1}{\varepsilon} V_\varepsilon(\mathbf{X}^\varepsilon)} \right] &\geq \min(V_-(x_1) + I(x_1), \dots, V_-(x_I) + I(x_I), \underbrace{\inf_{U^c} I}_{\geq v_{\max}}) - \delta \\ &\geq \min(\inf [V_- + I], v_{\max}) - \delta. \end{aligned}$$

□

5.5 A Large Deviations estimate

In order to analyze the (large sample size) variance of the AMS algorithm in the small noise regime, we will need precise estimates on the quantity $\varepsilon \ln \gamma_l^\varepsilon(q_\varepsilon^2)$ when $\varepsilon \rightarrow 0$. The latter will yield the 'final' part $\text{Loss}_O(l)$ of the loss function.

The goal of the present section is to detail the proof of these results. The proof is based on the extended Varadhan lemmas of the previous section.

We start by defining the set $H_l(\mathbf{x})$ of excursions in $\{\xi = z\}$ before hitting the interior $\{\xi > l\}$:

Lemma 5.4. Let $\mathbf{x} \in C([0, T], E)$ be given with $\tau_l^-(\mathbf{x}) < +\infty$ and T arbitrarily large. Denote the closed set

$$H_l(\mathbf{x}) \stackrel{\text{def.}}{=} \{\xi = l\} \cap \{\mathbf{x}_t, t \in [\tau_l^-(\mathbf{x}), \tau_l^+(\mathbf{x}) \wedge \tau_A^+(\mathbf{x}) \wedge T]\}.$$

Assume $\mathbf{x}^n \rightarrow_n \mathbf{x}$ in $C([0, +\infty[, E)$ for the uniform topology on bounded time intervals and that $H_l(\mathbf{x}) \neq \emptyset$. Let $t^n, n \geq 1$, denotes a sequence such that for all $n \geq 1$:

$$t_n \in [\tau_l^-(\mathbf{x}^n), \tau_l^+(\mathbf{x}^n) \wedge \tau_A^+(\mathbf{x}^n) \wedge T] \text{ and } \mathbf{x}_{t_n}^n \in \{\xi = l\}.$$

Then up to extraction of a sub-sequence

$$\lim_n \mathbf{x}_{t_n}^n \in H_l(\mathbf{x}).$$

Proof. Let us denote $(y, s) = \lim_n (\mathbf{x}_{t_n}^n, t^n)$, which always exists up to extraction by a compactness argument. We need to prove that $s \geq \tau_l^-(\mathbf{x})$, $s \leq \tau_l^+(\mathbf{x}) \wedge \tau_A^+(\mathbf{x}) \wedge T$ and $\mathbf{x}_s = y$ with $\xi(y) = l$ in order to complete the proof.

Trivially, $s \leq T$. Assume $\tau_l^+(\mathbf{x}) \wedge \tau_A^+(\mathbf{x}) < T$. By definition, \mathbf{x}^n which converges to \mathbf{x} hits $\{\xi > l\} \cup \dot{A}$ before $\tau_l^+(\mathbf{x}) \wedge \tau_A^+(\mathbf{x}) + \delta$ for all n large enough and δ arbitrary small; hence $s \leq \tau_z^+(\mathbf{x}) \wedge \tau_A^+(\mathbf{x})$.

Similarly, assume $s < \tau_l^-(\mathbf{x})$. Then we can find a small δ and an infinite number of \mathbf{x}^n such that $\tau_l^-(\mathbf{x}^n) < \tau_l^-(\mathbf{x}) - \delta$, which contradicts the uniform convergence.

Finally, $\mathbf{x}_{t_n}^n$ converges towards \mathbf{x}_s by uniform convergence, and since $\xi(\mathbf{x}_{t_n}^n) = l$, y also belongs to $\{\xi = l\}$ by continuity of ξ . \square

One can then consider the (measurable) potential function defined on continuous trajectories restricted to $[0, T]$:

$$V_{l,\varepsilon}^T(\mathbf{x}) \stackrel{\text{def.}}{=} -\varepsilon \ln [\mathbf{1}_{\tau_l(\mathbf{x}) < T \wedge \tau_A(\mathbf{x})} q^\varepsilon(\mathbf{x}_{\tau_l(\mathbf{x})})], \quad (5.3)$$

with the convention $\ln 0 = -\infty$. We will denote by $V_{l,\varepsilon}(\mathbf{x}) \stackrel{\text{def.}}{=} \lim_{T \rightarrow +\infty} V_{l,\varepsilon}^T(\mathbf{x})$ the trivial extension to the time interval \mathbb{R}_+ .

Before applying the extended Varadhan's lemmas, we need to estimate the lower and upper semi-continuous envelopes of $V_{l,\varepsilon}$. For this purpose, we denote by

$$\mathbf{x}^{T \wedge \tau_A^+}$$

the restriction of \mathbf{x} to the time interval $[0, T \wedge \tau_A^+(\mathbf{x})]$ and define

$$\underline{V}_l^T(\mathbf{x}) \stackrel{\text{def.}}{=} \begin{cases} \inf_{x \in H_l(\mathbf{x}^{T \wedge \tau_A^+})} U(x, \{\xi = l_B\}), & \text{if } \tau_l^-(\mathbf{x}) \leq T \wedge \tau_A^+(\mathbf{x}), \\ +\infty, & \text{else,} \end{cases}$$

as well as

$$\overline{V}_l^T(\mathbf{x}) \stackrel{\text{def.}}{=} \begin{cases} \sup_{x \in H_l(\mathbf{x})} U(x, \{\xi = l_B\}), & \text{if } \tau_l^+(\mathbf{x}) < T \wedge \tau_A^-(\mathbf{x}), \\ +\infty, & \text{else,} \end{cases}$$

and denote by $V_{l,\pm}(\mathbf{x}) \stackrel{\text{def.}}{=} \lim_{T \rightarrow +\infty} V_{l,\pm}^T(\mathbf{x})$ their natural extensions on \mathbb{R}_+ .

Lemma 5.5. *Let \mathbf{x} denotes a path in $C([0, T], E)$. Under Assumptions 1, 2, 3 and 4, one has:*

$$\limsup_{\substack{\mathbf{x}^n \rightarrow \mathbf{x} \\ \varepsilon \rightarrow 0}} V_{l, \varepsilon}^T(\mathbf{x}^n) \leq \overline{V}_l^T(\mathbf{x}).$$

as well as

$$\liminf_{\substack{\mathbf{x}^n \rightarrow \mathbf{x} \\ \varepsilon \rightarrow 0}} V_{l, \varepsilon}^T(\mathbf{x}^n) \geq \underline{V}_l^T(\mathbf{x}).$$

Proof. Let $\mathbf{x}^n \rightarrow_n \mathbf{x}$ be a uniformly convergent sequence on $[0, T]$.

Upper bound The condition $\tau_l^+(\mathbf{x}) < \tau_A^-(\mathbf{x}) \wedge T$ defines an open subset of $C([0, T], E)$ according to Lemma 5.1, thus there is a n_0 above which the sequence satisfies $\tau_l(\mathbf{x}^n) < T \wedge \tau_A^-(\mathbf{x})$. We can then use Lemma 5.4, together with Lemma 3.13 giving the uniform limit of q^ε , to obtain the claimed upper bound.

Lower bound The condition $\tau_l^-(\mathbf{x}) > T \wedge \tau_A^+(\mathbf{x})$, defines again an open set according to Lemma 5.1, thus there is a n_0 above which the sequence satisfies $\tau_l(\mathbf{x}^n) > T$ and thus $\underline{V}_l^T(\mathbf{x}) = +\infty$. Otherwise, if \mathbf{x} is outside that open set, we use again Lemma 5.4, together with Lemma 3.13 giving the uniform limit of q^ε , to obtain the claimed lower bound. \square

We can now proceed and estimate $\limsup / \liminf_\varepsilon \varepsilon \ln \gamma_l^\varepsilon(q_\varepsilon^2)$ using the extended Varadhan lemmas.

Lemma 5.6. *Let Assumptions 1, 2, 3 and 4 hold true. Then one has*

$$\begin{aligned} & - \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \gamma_l^\varepsilon(q_\varepsilon^2) \\ & \geq \inf_{\mathbf{x}: \tau_{l_B}^-(\mathbf{x}) < \tau_A^+(\mathbf{x})} I_{[0, \tau_l^+(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_l^+(\mathbf{x}), \tau_{l_B}^-(\mathbf{x})]}(\mathbf{x}) \left(\stackrel{def.}{=} \overline{w}(l) \right). \end{aligned}$$

Proof. Assume $T > 0$ given, arbitrary large. By definition of γ^ε , and using $q_\varepsilon \leq 1$, we get the upper bound

$$\begin{aligned} \gamma_l^\varepsilon(q_\varepsilon^2) &= \mathbb{E} \left[\mathbf{1}_{\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)} q_\varepsilon^2 \left(\mathbf{X}_{\tau_l(\mathbf{X}^\varepsilon)}^\varepsilon \right) \right] \\ &\leq \mathbb{E} \left[\mathbf{1}_{\tau_l(\mathbf{X}^\varepsilon) < T \wedge \tau_A(\mathbf{X}^\varepsilon)} q_\varepsilon^2 \left(\mathbf{X}_{\tau_l(\mathbf{X}^\varepsilon)}^\varepsilon \right) \right] + \mathbb{P} [\tau_A(\mathbf{X}^\varepsilon) \geq T] \\ &= \mathbb{E} \left[e^{-\frac{2}{\varepsilon} V_{l, \varepsilon}(\mathbf{X}^\varepsilon)} \right] + \mathbb{P} [\tau_A(\mathbf{X}^{\varepsilon, T}) \geq T]. \end{aligned}$$

Using Varadhan's upper bound (Lemma 5.3) and Lemma 5.5, with Assumption 4, we get

$$- \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \gamma_l^\varepsilon(q_\varepsilon^2) \geq \min_{\mathbf{x}} \left[\inf_{[0, T]} I_{[0, T]}[\mathbf{x}] + 2\underline{V}_l^T[\mathbf{x}], C_T \right],$$

where $\lim_{T \rightarrow +\infty} C_T = +\infty$.

Let $\delta > 0$ be given. Without loss of generality one can assume that the lower bound above is finite. As a consequence there exists $\mathbf{x}^\delta \in C([0, T], E)$ such that

$$\inf_{[0, T]} I_{[0, T]} + 2\underline{V}_l^T \geq I_{[0, T]}[\mathbf{x}^\delta] + 2\underline{V}_l^T[\mathbf{x}^\delta] - \delta.$$

By definition of \underline{V}_l^T , there exists

$$t_*^\delta \in [\tau_l^-(\mathbf{x}^\delta), \tau_l^+(\mathbf{x}^\delta) \wedge \tau_A^+(\mathbf{x}^\delta) \wedge T]$$

such that $\mathbf{x}_{t_*^\delta}^\delta \in \{\xi = l\}$ and

$$\underline{V}_l^T[\mathbf{x}^\delta] \geq U(\mathbf{x}_{t_*^\delta}^\delta, \{\xi = l_B\}) - \delta.$$

By additivity of the rate function, one also have $I_{[0,T]}[\mathbf{x}^\delta] \geq I_{[0,t_*^\delta]}[\mathbf{x}^\delta]$.

We can pick a $\mathbf{x}^\delta, t_*^\delta, T^\delta$ with $\tau_l^-(\mathbf{x}^\delta) \leq t_*^\delta < T^\delta < +\infty$ such that for any $T \geq T^\delta$ it holds

$$\inf I_{[0,T]} + 2\underline{V}_l^T \geq I_{[0,t_*^\delta]}[\mathbf{x}^\delta] + 2U(\mathbf{x}_{t_*^\delta}^\delta, \{\xi = l_B\}) - \delta.$$

We can now modify \mathbf{x}^δ , and extend it after t_*^δ by a trajectory δ -close to the optimal trajectories defining the cost $U^A(\mathbf{x}_{t_*^\delta}^\delta, \{\xi = l_B\})$, that is

$$U(\mathbf{x}_{t_*^\delta}^\delta, \{\xi = l_B\}) \geq I_{[t_*^\delta, \tau_{l_B}^-(\mathbf{x}^\delta)]}[\mathbf{x}^\delta] - \delta,$$

and such that $\tau_A^+(\mathbf{x}^\delta) > \tau_{l_B}^-(\mathbf{x}^\delta)$. Finally, one gets

$$\begin{aligned} \inf I_{[0,T]} + 2\underline{V}_l^T &\geq I_{[0,t_*^\delta]}[\mathbf{x}^\delta] + 2I_{[t_*^\delta, \tau_{l_B}^-(\mathbf{x}^\delta)]}[\mathbf{x}^\delta] - 3\delta \\ &\geq I_{[0, \tau_l^+(\mathbf{x}^\delta)]}[\mathbf{x}^\delta] + 2I_{[\tau_l^+(\mathbf{x}^\delta), \tau_{l_B}^-(\mathbf{x}^\delta)]}[\mathbf{x}^\delta] - 3\delta \\ &\geq \inf_{\mathbf{x}: \tau_{l_B}^-(\mathbf{x}) < \tau_A^+(\mathbf{x})} I_{[0, \tau_l^+(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_l^+(\mathbf{x}), \tau_{l_B}^-(\mathbf{x})]}(\mathbf{x}) - 3\delta, \end{aligned}$$

where in the second line of the above one has used again the additivity of the rate functions. The fact that δ small and T large are arbitrary yields the result. \square

Lemma 5.7. *Let Assumptions 1, 2, 3 and 4 hold true. Then one has*

$$\begin{aligned} & - \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \gamma_l^\varepsilon(q_\varepsilon^2) \\ & \leq \inf_{\mathbf{x}: \tau_{l_B}^-(\mathbf{x}) < \tau_A^+(\mathbf{x})} I_{[0, \tau_l^-(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_l^-(\mathbf{x}), \tau_{l_B}^-(\mathbf{x})]}(\mathbf{x}) \left(\stackrel{\text{def.}}{=} \bar{w}(l) \right). \end{aligned}$$

Proof. Let $T > 0$ be given. By definition of γ^ε and q^ε (and since $q^\varepsilon \leq 1$), we get the upper bound

$$\begin{aligned} \gamma_l^\varepsilon(q_\varepsilon^2) &= \mathbb{E} \left[\mathbf{1}_{\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)} q_\varepsilon^2 \left(\mathbf{X}_{\tau_l(\mathbf{X}^\varepsilon)}^\varepsilon \right) \right] \\ &\geq \mathbb{E} \left[\mathbf{1}_{\tau_l(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon) \wedge T} q_\varepsilon^2 \left(\mathbf{X}_{\tau_l(\mathbf{X}^\varepsilon)}^\varepsilon \right) \right] \\ &= \mathbb{E} \left[e^{-\frac{2}{\varepsilon} V_{l,\varepsilon}(\mathbf{X}^\varepsilon, T)} \right] \end{aligned}$$

We can thus directly use Varadhan's lower bound (Lemma 5.2) and Lemma 5.5 to get

$$- \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \gamma_l^\varepsilon(q_\varepsilon^2) \leq \inf I_{[0,T]} + 2\bar{V}_l^T.$$

One can then restrict in the infimum above to trajectories verifying $\tau_l^+(\mathbf{x}) = \tau_l^-(\mathbf{x})$ to get the upper bound:

$$\begin{aligned} \inf I_{[0,T]} + 2\bar{V}_l^T &\leq \\ \inf_{\tau_l^-(\mathbf{x})=\tau_l^+(\mathbf{x})} I_{[0,T]}[\mathbf{x}] + 2\bar{V}_l^T[\mathbf{x}] &\stackrel{\text{def.}}{=} A(l), \end{aligned}$$

and remark that by definition of \bar{V}_l^T , if $\tau_l^-(\mathbf{x}) = \tau_l^+(\mathbf{x}) < \tau_A^-(\mathbf{x})$, then $\bar{V}_l[\mathbf{x}] = U(\mathbf{x}_{\tau_l^-(\mathbf{x})}, \{\xi \geq l_B\})$. Next, Assumption 2 ensures that minimizing $I_{[\tau_l^-(\mathbf{x}),T]}(\mathbf{x})$ with the constraint that $\tau_l^-(\mathbf{x}) = \tau_l^+(\mathbf{x})$ yields 0 so that:

$$\begin{aligned} A(l) &= \inf_{\tau_l^-(\mathbf{x}) < \tau_A^-(\mathbf{x}) \wedge T} I_{[0,\tau_l^-(\mathbf{x})]}[\mathbf{x}] + 2U(\mathbf{x}_{\tau_l^-(\mathbf{x})}, \{\xi \geq l_B\}) \\ &= \inf_{\substack{\tau_l^-(\mathbf{x}) < T \\ \tau_{l_B}^-(\mathbf{x}) < \tau_A^-(\mathbf{x})}} I_{[0,\tau_l^-(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_l^-(\mathbf{x}),\tau_{l_B}^-(\mathbf{x})]}[\mathbf{x}] \end{aligned}$$

It finally remains to remark that the horizon time T is arbitrary to conclude the proof. \square

5.6 Analysis of the loss functional

Lemma 5.8. *Let Assumption 2 holds true. Let us define for $l \in [\xi(x_0), l_B]$ and $C = A, \bar{A}$ or \dot{A} :*

$$\begin{cases} \underline{w}^C(l) \stackrel{\text{def.}}{=} \inf_{\substack{\mathbf{x}: \tau_{l_B}^-(\mathbf{x}) < \tau_C(\mathbf{x}) \\ \mathbf{x}_0 = x_0}} I_{[0,\tau_l^+(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_l^+(\mathbf{x}),\tau_{l_B}^-(\mathbf{x})]}[\mathbf{x}], \\ \bar{w}^C(l) \stackrel{\text{def.}}{=} \inf_{\substack{\mathbf{x}: \tau_{l_B}^-(\mathbf{x}) < \tau_C(\mathbf{x}) \\ \mathbf{x}_0 = x_0}} I_{[0,\tau_l^-(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_l^-(\mathbf{x}),\tau_{l_B}^-(\mathbf{x})]}[\mathbf{x}]. \end{cases}$$

And we simply denote $\underline{w}(l) = \underline{w}^A(l)$ and $\bar{w}(l) = \bar{w}^A(l)$.

First, $\bar{w}^{\dot{A}} = \bar{w}^{\bar{A}}$ and $\underline{w}^{\dot{A}} = \underline{w}^{\bar{A}}$. Second, \bar{w} and \underline{w} are respectively the left- and right-continuous versions of the same decreasing function. In other words: $\bar{w}(l) = \lim_{l^-} \underline{w}$ and $\underline{w}(l) = \lim_{l^+} \bar{w}$.

Proof. First let us remark that by a direct application of Assumption 2, the definition of \underline{w} and \bar{w} are independent of the choice $C = \dot{A}$ or $C = \bar{A}$.

Let $l' < l$ be given, and let \mathbf{x} be a trajectory that reaches the final level l_B . Since $\tau_{l'}^+ < \tau_l^-$, by additivity of the (non-negative) rate functions, one has

$$I_{[0,\tau_l^-(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_l^-(\mathbf{x}),\tau_{l_B}^-(\mathbf{x})]}[\mathbf{x}] \leq I_{[0,\tau_{l'}^+(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_{l'}^+(\mathbf{x}),\tau_{l_B}^-(\mathbf{x})]}[\mathbf{x}].$$

Taking the infimum it yields $\bar{w}(l) \leq \underline{w}(l')$.

Using also additivity of rate functions, one can check that $\bar{w} \geq \underline{w}$ and that \bar{w} and \underline{w} are decreasing functions. This yields the result. \square

Lemma 5.9. *Let Assumption 2 and 3. Let x_0 and l_B be given. For $l \in [\xi(x_0), l_B]$, let us denote by $u(l) \stackrel{\text{def.}}{=} U(x_0, \{\xi \geq l\})$. Then it holds*

$$2u(l_B) - u(l) - \bar{w}(l) \leq \text{Loss}(l) \leq 2u(l_B) - u(l) - \underline{w}(l),$$

The maximization problem

$$\sup_{l \in [\xi(x_0), l_B]} \text{Loss}(l)$$

has at least a maximizer $l_* \in]\xi(x_0), l_B[$, with $\text{Loss}(0) = \text{Loss}(l_B) = 0$.

Proof. First remark that by construction:

$$\underline{w}(l) \leq \inf_{x \in \{\xi=l\}} U^{(l)}(x_0, \cdot) + 2U(\cdot, B) \leq \bar{w}(l)$$

which implies the first inequality by definition of the loss function.

Next it holds by definition of u :

$$\begin{aligned} & \bar{w}(l) + u(l) \\ &= \inf_{\mathbf{x}: \tau_{l_B}(\mathbf{x}) < \tau_A(\mathbf{x})} I_{[0, \tau_l^+(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_l^+(\mathbf{x}), \tau_{l_B}(\mathbf{x})]}[\mathbf{x}] + \inf_{\mathbf{x}': \tau_l^+(\mathbf{x}') < \tau_A(\mathbf{x}')} I_{[0, \tau_l^+(\mathbf{x}')] }[\mathbf{x}'] \\ &\leq \inf_{\mathbf{x}: \tau_{l_B}(\mathbf{x}) < \tau_A(\mathbf{x})} 2I_{[0, \tau_l^+(\mathbf{x})]}[\mathbf{x}] + 2I_{[\tau_l^+(\mathbf{x}), \tau_{l_B}(\mathbf{x})]}[\mathbf{x}] = 2u(l_B), \end{aligned}$$

which implies that $0 \leq 2u(l_B) - u(l) - \bar{w}(l)$. By construction one also has that $2u(l_B) - u(l) - \underline{w}(l) \leq 2u(l_B) < +\infty$.

Next, since by Assumption 3, u is an increasing continuous and \bar{w} is a left-continuous decreasing function, while \underline{w} is its right continuous version.

This implies that the bounded functions $u + \bar{w}$ and $u + \underline{w}$ attain their extrema, i) at the same values (either by left or right). An so it holds for Loss.

Finally, we remark that by construction $u(\xi(x_0)) = 0$ and $w(\xi(x_0)) = 2u(l_B)$, whereas $w(l_B) = 2u(l_B)$, so that $\text{Loss}(0) = \text{Loss}(l_B) = 0$. \square

5.7 Small noise analysis of variance

A key property enabling the analysis of the AMS large sample size variance is the following.

Lemma 5.10. *The map*

$$l \mapsto \gamma_l(q_{l_B}^2)$$

is increasing.

Proof. Let $l' \geq l$ be given and denote by $\{\mathcal{F}_s, s \geq 0\}$ the natural filtration of X . By definition of γ_l ,

$$\begin{aligned} \gamma_l(q_{l_B}^2) &= \mathbb{E} \left[\mathbf{1}_{\tau_l(\mathbf{X}) < \tau_A(\mathbf{X})} \mathbb{P}_{\mathbf{X}_{\tau_l(\mathbf{X})}}^2 \left(\tau_{l_B}(\mathbf{X}) < \tau_A(\mathbf{X}) \right) \right] \\ &= \mathbb{E} \left[\mathbf{1}_{\tau_l(\mathbf{X}) < \tau_A(\mathbf{X})} \left[\mathbb{E}_{\mathbf{X}_{\tau_l(\mathbf{X})}} \left(\mathbf{1}_{\tau_{l_B}(\mathbf{X}) < \tau_A(\mathbf{X})} \mid \mathcal{F}_{\tau_l(\mathbf{X}) \wedge \tau_A(\mathbf{X})} \right) \right]^2 \right] \end{aligned}$$

so that applying Jensen's inequality to the conditional expectation above, using in addition the strong Markov property:

$$\begin{aligned} & E_{\mathbf{X}_{\tau_l(\mathbf{X})}} \left(\mathbf{1}_{\tau_{l_B}(\mathbf{X}) < \tau_A(\mathbf{X})} \mid \mathcal{F}_{\tau_l(\mathbf{X}) \wedge \tau_A(\mathbf{X})} \right)^2 \\ &= E_{\mathbf{X}_{\tau_l(\mathbf{X})}} \left(E_{\mathbf{X}_{\tau_{l'}(\mathbf{X})}} \left(\mathbf{1}_{\tau_{l'}(\mathbf{X}) < \tau_A(\mathbf{X})} \mathbf{1}_{\tau_{l_B}(\mathbf{X}) < \tau_A(\mathbf{X})} \mid \mathcal{F}_{\tau_{l'}(\mathbf{X}) \wedge \tau_A(\mathbf{X})} \right) \mid \mathcal{F}_{\tau_l(\mathbf{X}) \wedge \tau_A(\mathbf{X})} \right)^2 \\ &\leq E_{\mathbf{X}_{\tau_l(\mathbf{X})}} \left(\mathbf{1}_{\tau_{l'}(\mathbf{X}) < \tau_A(\mathbf{X})} E_{\mathbf{X}_{\tau_{l'}(\mathbf{X})}} \left(\mathbf{1}_{\tau_{l_B}(\mathbf{X}) < \tau_A(\mathbf{X})} \mid \mathcal{F}_{\tau_{l'}(\mathbf{X}) \wedge \tau_A(\mathbf{X})} \right)^2 \mid \mathcal{F}_{\tau_l(\mathbf{X}) \wedge \tau_A(\mathbf{X})} \right), \end{aligned}$$

which, once put in the former expression, finally gives

$$\gamma_l(q_{l_B}^2) \leq \gamma_l(q_{l_B}^2).$$

□

We can then compute

Lemma 5.11. *Assume Let Assumptions 1, 2, 3 and 4. One has*

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \varepsilon \log \int_{\xi(x_0)}^{l_B} \gamma_l^\varepsilon \left((q_{l_B}^\varepsilon)^2 \right) d(-p_l^\varepsilon) \\ = - \inf_{l \in [\xi(x_0), l_B]} \left[U(x_0, \{\xi = l\}) + \inf_{\{\xi=l\}} \left[U^{(l)}(x_0, \cdot) + 2U(\cdot, \{\xi = l_B\}) \right] \right] \end{aligned}$$

Proof. Let us consider a finite discretization $\{l_j\}$ of the interval $[\xi(x_0), l_B]$, and let us denote for simplicity $\gamma_l^\varepsilon \left((q_{l_B}^\varepsilon)^2 \right) \equiv \gamma_l^\varepsilon$ throughout the present proof. By the monotony property of Lemma 5.10, it yields:

$$\sum_j \gamma_{l_j}^\varepsilon (p_{l_j}^\varepsilon - p_{l_{j+1}}^\varepsilon) \leq \int_{\xi(x_0)}^{l_B} \gamma_l^\varepsilon d(-p_l^\varepsilon) \leq \sum_j \gamma_{l_{j+1}}^\varepsilon (p_{l_j}^\varepsilon - p_{l_{j+1}}^\varepsilon).$$

Using the usual rule for finite sum and any quantity $*_j^\varepsilon$

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \sum_j *_j^\varepsilon = \max_j \lim_{\varepsilon \rightarrow 0} \varepsilon \log *_j^\varepsilon,$$

one can use Lemmata 3.14, 5.6 and 5.7 to obtain

$$\max_j \underline{u}_{l_j, l_{j+1}} \leq \lim_{\varepsilon \rightarrow 0} \varepsilon \log \int_{\xi(x_0)}^{l_B} \gamma_l^\varepsilon d(-p_l^\varepsilon) \leq \max_j \bar{u}_{l_{j+1}, l_{j+1}},$$

where we have used the notation

$$\underline{u}_{l_j, l_k} \stackrel{\text{def.}}{=} u(l_j) + \underline{w}(l_k),$$

taking a converging sequence of discretizations, $\max_j |l_{j+1} - l_j| \rightarrow 0$, and recalling that \underline{w} and \bar{w} are the left- and -right continuous version of the same decreasing function while u is continuous (see Lemma 5.8), we can conclude. □

With the two lemmas above, we can finally conclude the proof of Theorem 3.17.

Proof of Theorem 3.17. Recall that $\sigma_{\varepsilon, \text{ams}}^2$ denotes the large sample size variance of the estimator of p^ε in (3.5).

First, one has $\sigma_{\varepsilon, \text{ams}}^2 \leq -p_\varepsilon^2 \ln p_\varepsilon$ so that $\lim_{\varepsilon} \varepsilon \log \sigma_{\varepsilon, \text{ams}}^2 \leq -2U(x_0, \{\xi = l_B\})$.

Then a simple integration by parts then shows that:

$$\sigma_{\varepsilon, \text{ams}}^2 = p_\varepsilon^2 \ln p_\varepsilon + 2 \int_{\xi(x_0)}^{l_B} \gamma_l^\varepsilon (q_\varepsilon^2) d(-p_l^\varepsilon); \quad (5.4)$$

but since the term $p_\varepsilon^2 \ln p_\varepsilon$ is negative, we need to distinguish two cases.

First, if $\sup_l \text{Loss}(l) = 0$ that is if

$$2U(x_0, \{\xi = l_B\}) = \inf_{l \in [\xi(x_0), l_B]} \left[U(x_0, \{\xi = l\}) + \inf_{\{\xi=l\}} [U(x_0, \cdot) + 2U(\cdot, \{\xi = l_B\})] \right]$$

one can simply bound $-p_\varepsilon^2 \ln p_\varepsilon \geq \sigma_{\varepsilon, \text{ams}}^2 \geq 2 \int_{\xi(x_0)}^{l_B} \gamma_l^\varepsilon(q_\varepsilon^2) d(-p_l^\varepsilon)$ one obtains that the lower and upper bounds are logarithmically equivalent so that $\lim_\varepsilon \varepsilon \log \sigma_{\varepsilon, \text{ams}}^2 = -2U(x_0, \{\xi = l_B\})$.

Second, if $\sup_l \text{Loss}(l) > 0$, that is if

$$2U(x_0, \{\xi = l_B\}) > \inf_{l \in [\xi(x_0), l_B]} \left[U(x_0, \{\xi = l\}) + \inf_{\{\xi=l\}} [U(x_0, \cdot) + 2U(\cdot, \{\xi = l_B\})] \right]$$

then the second term in (5.4) dominates at logarithmic scales, yielding the claimed result. \square

5.8 AMS for two clones and at worse one iteration

This section discusses the proof that for an AMS algorithm with $N = 2$, one has

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P} \left[I_{\text{iter}}^{2, \varepsilon} \leq 1 \right] = - \inf_{l \in [\xi(x_0), l_B]} \left[U(x_0, \{\xi = l\}) + \inf_{\{\xi=l\}} [U(x_0, \cdot) + 2U(\cdot, \{\xi = l_B\})] \right].$$

For the sake of concision, the proof is only sketched.

By exchangeability between the two clones, $\mathbb{P} \left[I_{\text{iter}}^{2, \varepsilon} \leq 1 \right]$ is twice the probability of the same event with the additional requirement that the killed clone is the clone with index 2. Then remark by construction of the AMS algorithm that

$$\mathbb{P} \left[I_{\text{iter}}^{2, \varepsilon} = 1 \right] = 2\mathbb{E} \left[\mathbf{1}_{\tau_{\text{Max}^\varepsilon}(\mathbf{X}^\varepsilon) < \tau_A(\mathbf{X}^\varepsilon)} \left(q_{l_B}^\varepsilon(\mathbf{X}_{\tau_{\text{Max}^\varepsilon}(\mathbf{X}^\varepsilon)}^\varepsilon) \right)^2 \right],$$

where in the above \mathbf{X}^ε denotes the clone with index 1 and Max^ε is the maximum level in $[\xi(x_0), l_B]$ of the clone with index 2, independent from Max^ε . Using Lemma 3.13, Max^ε satisfies a LDP with good rate function $l \mapsto I_{\text{Max}}(l) = U(x_0, \{\xi, l\})$. By the tensorization principle $(\text{Max}^\varepsilon, \mathbf{X}^\varepsilon)$ satisfy a LDP with a good rate function so that we can apply Varadhan lemmas similarly to the main estimates of this paper with

$$V_\varepsilon(\mathbf{x}, l) = -\ln \left(\mathbf{1}_{\tau_l(\mathbf{x}) < \tau_A(\mathbf{x})} q_{l_B}^\varepsilon(\mathbf{x}_{\tau_l(\mathbf{x})}) \right),$$

the only difference being the additional dependence in l (we skip the time horizon cut-off T for clarity), in which l is replace by Max^ε . Using similar technical arguments in the application of Varadhan's lemmas, we get that

$$\begin{aligned} \varepsilon \log \mathbb{P} \left[I_{\text{iter}}^{2, \varepsilon} = 1 \right] &= \varepsilon \log 2\mathbb{E} \left[e^{-2V_\varepsilon(\mathbf{X}^\varepsilon, \text{Max}^\varepsilon)} \right] \xrightarrow{\varepsilon \rightarrow 0} \\ &- \inf_{\substack{l \in [\xi(x_0), l_B] \\ \mathbf{x}_l \in \{\xi=l\}}} \left[I_{\text{Max}}(l) + U^{(l)}(x_0, \mathbf{x}_l) + 2U(\mathbf{x}_l, \{\xi \geq l_B\}) \right] \end{aligned}$$

which is precisely the claimed result.

A More on Freidlin-Wentzell

In this Section, we recall the definitions of the rate function and the quasi-potential for a generic SDE (3.1) with possibly degenerate noise.

Define for $r \in H^1([0, T], \mathbb{R}^m)$, where H^1 denotes the Sobolev space defined by $\frac{d}{dt}r \in L^2([0, T], \mathbb{R}^m)$, the following functional

$$I_{[0, T]}(\mathbf{x}) \stackrel{\text{def.}}{=} \inf_{\substack{r \in H^1([0, T], \mathbb{R}^m) \\ \dot{\mathbf{x}} = b(\mathbf{x}) + \sigma(\mathbf{x})\dot{r}}} \frac{1}{2} \int_0^T |\dot{r}|^2 dt.$$

The latter, when finite-valued, can be written explicitly as:

$$I_{[0, T]}(\mathbf{x}) = \frac{1}{2} \int_0^T \|\dot{\mathbf{x}}_t - b(\mathbf{x}_t)\|_{(\sigma\sigma^T)^{-1}(\mathbf{x}_t)}^2 dt,$$

where in the above $(\sigma\sigma^T)^{-1}$ denotes the spectral pseudo-inverse so that:

$$\|y\|_{(\sigma\sigma^T)^{-1}}^2 \stackrel{\text{def.}}{=} \inf_{\lambda: y = \sigma\lambda} |\lambda|^2.$$

The associated quasi potential is then given by:

$$U(x, y) = \inf_{\mathbf{x}_0=x, \mathbf{x}_1=y} \int_0^1 \inf_{\lambda \in \mathbb{R}_*} \left\| \lambda \frac{d\mathbf{x}_\theta}{d\theta} - \frac{1}{\lambda} b(\mathbf{x}_\theta) \right\|_{(\sigma\sigma^T)^{-1}(\mathbf{x}_\theta)}^2 d\theta.$$

Acknowledgement

This work has been partially supported by ANR SINEQ, ANR-21-CE40-0006.

References

- [1] Freddy Bouchet, Jason Laurie, and Oleg Zaboronski. Langevin dynamics, large deviations and instantons for the quasi-geostrophic model and two-dimensional euler equations. *Journal of Statistical Physics*, 156(6):1066–1092, 2014.
- [2] C.-E. Bréhier, M. Gazeau, L. Goudenège, T. Lelièvre, and M. Rousset. Unbiasedness of some generalized adaptive multilevel splitting algorithms. *Ann. Appl. Probab.*, 26(6):3559–3601, 2016.
- [3] Charles-Edouard Bréhier, Tony Lelièvre, and Mathias Rousset. Analysis of adaptive multilevel splitting algorithms in an idealized case. *ESAIM: Probability and Statistics*, 19:361–394, 2015.
- [4] A. Budhiraja and P. Dupuis. *Analysis and approximation of rare events*, volume 94 of *Probability Theory and Stochastic Modelling*. Springer, New York, 2019. Representations and weak convergence methods.
- [5] Y. Cai and P. Dupuis. Analysis of an interacting particle method for rare event estimation. *Queueing Syst.*, 73(4):345–406, 2013.
- [6] F. Cérou, P. Del Moral, F. Le Gland, and P. Lezaud. Genetic genealogical models in rare event analysis. *ALEA Lat. Am. J. Probab. Math. Stat.*, 1:181–203, 2006.
- [7] F. Cérou, B. Delyon, A. Guyader, and M. Rousset. On the Asymptotic Normality of Adaptive Multilevel Splitting. *SIAM/ASA Journal on Uncertainty Quantification*, 7(1):1–30, 2019.

- [8] Frédéric Cérou, Bernard Delyon, Arnaud Guyader, and Mathias Rousset. A central limit theorem for Fleming–Viot particle systems. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 56(1):637–666, 2020.
- [9] Frédéric Cérou, Arnaud Guyader, and Mathias Rousset. Adaptive multi-level splitting: Historical perspective and recent results. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(4):043108, 2019.
- [10] Frédéric Cérou, Arnaud Guyader, and Mathias Rousset. On synchronized Fleming–Viot particle systems. *Theory of Probability and Mathematical Statistics*, 102:45–71, 2020.
- [11] Hock Peng Chan and Tze Leung Lai. A general theory of particle filters in hidden Markov models and some applications. *The Annals of Statistics*, 41(6):2877–2904, 2013.
- [12] Sourav Chatterjee and Persi Diaconis. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018.
- [13] T. Dean and P. Dupuis. Splitting for rare event simulation: a large deviation approach to design and analysis. *Stochastic Process. Appl.*, 119(2):562–587, 2009.
- [14] T. Dean and P. Dupuis. The design and analysis of a generalized RESTART/DPR algorithm for rare event simulation. *Ann. Oper. Res.*, 189:63–102, 2011.
- [15] P. Del Moral. *Feynman-Kac formulae, Genealogical and interacting particle systems with applications*. Springer-Verlag, New York, 2004.
- [16] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B*, 68(3):411–436, 2006.
- [17] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In Jacques Azéma, Michel Ledoux, Michel Émery, and Marc Yor, editors, *Séminaire de Probabilités XXXIV*, volume 1729 of *Lecture Notes in Mathematics*, pages 1–145. Springer Berlin Heidelberg, 2000.
- [18] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Applications of mathematics. Springer, 1998.
- [19] Qiming Du and Arnaud Guyader. Variance estimation in adaptive sequential Monte Carlo. *The Annals of Applied Probability*, 31(3):1021–1060, 2021.
- [20] P. Dupuis and H. Wang. Importance sampling, large deviations, and differential games. *Stoch. Stoch. Rep.*, 76(6):481–508, 2004.
- [21] P. Dupuis and H. Wang. Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Math. Oper. Res.*, 32(3):723–757, 2007.
- [22] P. Dupuis and H. Wang. Importance sampling for Jackson networks. *Queueing Syst.*, 62(1-2):113–157, 2009.
- [23] M.I. Freidlin, J. Sziucs, and A.D. Wentzell. *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2012.
- [24] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag, 1984.
- [25] Arnaud Guyader and Hugo Touchette. Efficient large deviation estimation based on importance sampling. *Journal of Statistical Physics*, 181(2):551–586, 2020.

- [26] P. L'Ecuyer, J. Blanchet, B. Tuffin, and P.W. Glynn. Asymptotic robustness of estimators in rare-event simulation. *ACM Transactions on Modeling and Computer Simulation*, 18(3):1269–1283, 2008.
- [27] Anthony Lee and Nick Whiteley. Variance estimation in the particle filter. *Biometrika*, 105(3):609–625, 2018.
- [28] Thibault Lestang, Francesco Ragone, Charles-Edouard Bréhier, Corentin Herbert, and Freddy Bouchet. Computing return times or return periods with rare event algorithms. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(4):043213, 2018.
- [29] L.J.S. Lopes, C.G. Mayne, C. Chipot, and T. Lelièvre. Adaptive multilevel splitting method: Isomerization of the alanine dipeptide. *arXiv preprint arXiv:1707.00950*, 2017.
- [30] Henri Louvin, Eric Dumonteil, Tony Lelièvre, Mathias Rousset, and Cheikh M Diop. Adaptive multilevel splitting for monte carlo particle transport. In *EPJ Web of Conferences*, volume 153, page 06006. EDP Sciences, 2017.
- [31] Dario Lucente, Joran Rolland, Corentin Herbert, and Freddy Bouchet. Coupling rare event algorithms with data-based learned committor functions using the analogue markov chain. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(8):083201, 2022.
- [32] Hao Ma and Henk AP Blom. Random assignment versus fixed assignment in multilevel importance splitting for estimating stochastic reach probabilities. *Methodology and Computing in Applied Probability*, pages 1–26, 2022.
- [33] Francesco Ragone, Jeroen Wouters, and Freddy Bouchet. Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, 115(1):24–29, 2018.
- [34] Jutta Rogal. Reaction coordinates in complex systems—a perspective. *The European Physical Journal B*, 94(11):1–9, 2021.
- [35] Joran Rolland and Eric Simonnet. Statistical behaviour of adaptive multilevel splitting algorithms in simple models. *Journal of Computational Physics*, 283:541–558, 2015.
- [36] I. Teo, C.G. Mayne, K. Schulten, and T. Lelièvre. Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time. *Journal of chemical theory and computation*, 12(6):2983–2989, 2016.
- [37] S. R. S. Varadhan. *Large Deviations and Applications*. Society for Industrial and Applied Mathematics, 1984.