

## CONVERGENCE ANALYSIS OF ADAPTIVE BIASING POTENTIAL METHODS FOR DIFFUSION PROCESSES\*

MICHEL BENAÏM<sup>†</sup> AND CHARLES-EDOUARD BRÉHIER<sup>‡</sup>

**Abstract.** This article is concerned with the mathematical analysis of a family of adaptive importance sampling algorithms applied to diffusion processes. These methods, referred to as Adaptive Biasing Potential methods, are designed to efficiently sample the invariant distribution of the diffusion process, thanks to the approximation of the associated free energy function (relative to a reaction coordinate). The bias which is introduced in the dynamics is computed adaptively; it depends on the past of the trajectory of the process through some time-averages.

We give a detailed and general construction of such methods. We prove the consistency of the approach (almost sure convergence of well-chosen weighted empirical probability distribution). We justify the efficiency thanks to several qualitative and quantitative additional arguments. To prove these results, we revisit and extend tools from stochastic approximation applied to self-interacting diffusions, in an original context.

**Keywords.** adaptive biasing; self-interacting diffusions; free energy computation.

**AMS subject classifications.** 60J60; 60C35.

### 1. Introduction

In many applications in physics, biology, chemistry, etc. there is a huge interest in the two following problems. First, in sampling probability distributions (denoted by  $\mu$ ), *i.e.* in constructing families of independent random variables with distribution  $\mu$ . Second, in computing averages  $\int \varphi d\mu$  of real-valued functions  $\varphi$ . These questions lead to challenging computational issues, when the support of  $\mu$  has large (possibly infinite) dimension – for instance, when  $\mu$  is the equilibrium distribution of a large system of particles, which is the typical situation in the field of molecular dynamics. The scientific literature contains many examples, as well as many approaches to construct efficient approximation procedures. We do not intend to provide an extensive review, but some relevant examples, which are connected to the methodology studied in this article, will be provided.

Many methods are based on stochastic simulation, also called MCMC methods. The idea is to run an ergodic Markov process having  $\mu$  as invariant distribution, and to use empirical averages as estimators. A standard example (but there are others) of such a process is given by the overdamped Langevin dynamics on  $\mathbb{R}^d$ ,

$$dx_t = -\nabla V(x_t)dt + \sqrt{2\beta^{-1}}dW_t,$$

whose invariant measure (under appropriate growth and regularity assumptions on the function  $V$ ) is the Boltzmann-Gibbs probability distribution

$$\mu(dx) = \mu_\star(dx) = \frac{e^{-\beta V(x)}}{Z(\beta)} dx.$$

One of the main limitations of standard MCMC approaches comes from the fact that the ergodic dynamics are metastable whenever  $\mu$  is multimodal. In the example

---

\*Received: October 6, 2017; Accepted (in revised form): October 10, 2018. Communicated by Grigorios A. Pavliotis.

<sup>†</sup>Université de Neuchâtel, Institut de Mathématiques, Rue Emile Argand 11, CH-2000 Neuchâtel, Switzerland ([michel.benaïm@unine.ch](mailto:michel.benaïm@unine.ch)). <http://members.unine.ch/michel.benaïm/perso/benaïm.html>

<sup>‡</sup>Univ Lyon, CNRS, Université Claude Bernard Lyon 1, UMR5208, Institut Camille Jordan, F-69622 Villeurbanne, France ([brehier@math.univ-lyon1.fr](mailto:brehier@math.univ-lyon1.fr)). <http://math.univ-lyon1.fr/~brehier/>

above, this happens when  $V$  has several local minima. A direct simulation is not able to efficiently and accurately sample the rare transitions between the metastable states, hence the need for advanced Monte-Carlo methods.

Many strategies have been proposed, analyzed and applied, to overcome this issue. The associated variance reduction approaches may be divided into two main families. On the one hand, importance sampling strategies are based on changing the reference probability measure. The realization of the rare events is enhanced by appropriate reweighting of  $\mu$ . In our context, such strategies require to simulate modified processes, which are constructed by biasing the dynamics. On the other hand, splitting strategies use interacting replicas, with mutation and selection procedures, without modifying the process dynamics.

The methods studied in this article are an example of adaptive biasing methods. They are based on the importance sampling strategy and the use of the so-called free energy function (which will be introduced below). When going into the details of the schemes and of the applications, there are many different versions; they all aim at flattening the free energy landscape, and to make free energy barriers disappear. We refer to the monograph [36] for an extensive review of such methods, and to [37, Section 4] for a survey on mathematical techniques. To name a few of the versions, we mention the following examples of adaptive biasing methods: the adaptive biasing force [15, 18, 27]; the Wang-Landau algorithm [40, 41]; metadynamics [1, 32]; the self-healing umbrella sampling method [39]. For related mathematical results, see for instance [14, 29, 33, 35] (adaptive biasing force); [23, 24] (Wang-Landau), [25] (self-healing umbrella sampling). This list is not exhaustive. We also refer to the recent survey paper [20] (and to references therein) for discussions and a comparison of these methods.

Our aim in this article is to give a mathematical analysis of a family of methods, independently of a comparison with the other methods mentioned above: the Adaptive Biasing Potential methods, related to [19]. In such methods, one constructs adaptive approximation of a potential energy function, instead of a mean force (see [34] for a discussion), hence the name. One of the key aspects of this work is that the dynamics is biased using quantities computed as time-averages over a single realization of the dynamics.

Let us mention the type of mathematical properties such algorithms are required to satisfy (exactly or in an approximate sense) for the estimation of averages  $\int \varphi d\mu$ . On the one hand, the consistency is the long-time convergence to this quantity, in a strong (almost sure or  $L^p$ ) sense, or in a weak sense (convergence of the expected value). On the other hand, the efficiency is generally studied in terms of the asymptotic mean-square error. It may also be considered from the point of view of the long-time behavior of occupation measures of the process.

A preliminary analysis of the Adaptive Biasing Potential methods considered in this article, has been performed in [3], in a simplified framework, without proofs. The aim of the present article is to provide the missing arguments, in a more abstract framework, and to study substantial generalizations. Below we first present the methods in the simplified framework from [3], see Section 1.1: the strategy and the results are exposed. We then present the general framework of the article, and the associated results, in Sections 1.2 and 1.3.

**1.1. Adaptive Biasing Potential method in a simplified framework.** This section is pedagogical: the ideas are introduced independently of the abstract notation which will allow us to consider many examples of diffusion processes. We are interested in sampling probability distributions on the flat  $d$ -dimensional torus  $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$ , of

the following form:

$$\mu(dx) = \mu_\star(dx) = \frac{\exp(-\beta V(x))}{Z(\beta)} dx,$$

where  $V: \mathbb{T}^d \rightarrow \mathbb{R}$  is a smooth potential energy function, and  $\beta \in (0, \infty)$  is referred to as the inverse temperature. Finally,  $dx$  is the Lebesgue measure on  $\mathbb{T}^d$ , and  $Z(\beta) = \int_{\mathbb{T}^d} e^{-\beta V(x)} dx$  is the normalization constant.

A natural choice of associated ergodic process is given by the overdamped Langevin dynamics (or Brownian dynamics)

$$dX_t^0 = -\nabla V(X_t^0) dt + \sqrt{2\beta^{-1}} dW_t, \quad X_0^0 = x_0, \quad (1.1)$$

where  $(W_t)_{t \geq 0}$  is a standard Wiener process on  $\mathbb{T}^d$ .

By ergodicity, the empirical distribution  $\bar{\mu}_t^0 = \frac{1}{t} \int_0^t \delta_{X_r^0} dr$  converges (in distribution), almost surely, towards  $\mu_\star$ , when time  $t$  goes to infinity. As already mentioned, the convergence may be slow when  $V$  has several local minima.

To accelerate convergence to equilibrium, other stochastic processes need to be used. In this article, the dynamics is modified with an adaptive change of the potential energy function: the function  $V$  is replaced with a time-dependent function  $V_t$  – hence the terminology *Adaptive Biasing Potential* (ABP) method,

$$dX_t = -\nabla V_t(X_t) dt + \sqrt{2\beta^{-1}} dW_t.$$

Compared with other methods mentioned above, one of the specificities of the method considered in this article is the structure of the time-dependent potential energy function  $V_t$ . It is constructed as  $V_t = V - A_t \circ \xi$ , where  $\xi: (x_1, \dots, x_d) \in \mathbb{T}^d \mapsto (x_1, \dots, x_m) \in \mathbb{T}^m$ , with  $m \in \{1, \dots, d-1\}$ , is an auxiliary function, referred to as the reaction coordinate, and  $A_t: \mathbb{T}^m \rightarrow \mathbb{R}$  is an approximation (in the regime  $t \rightarrow \infty$ ) of the so-called free energy function. In applications, the dimension  $m$  is chosen much smaller than  $d$ , and typically  $m \in \{1, 2, 3\}$ . It thus remains to explain how the function  $A_t$  is constructed adaptively. This is done in terms of the values  $(X_r)_{0 \leq r < t}$  of the process  $X$  up to time  $t$ .

Precisely, the dynamics of the ABP method, in the simplified framework considered in the current section (for the generalized version, see Equation (3.2)), is given by the following system:

$$\begin{cases} dX_t = -\nabla(V - A_t \circ \xi)(X_t) dt + \sqrt{2\beta^{-1}} dW(t) \\ \bar{\mu}_t = \frac{\bar{\mu}_0 + \int_0^t \exp(-\beta A_r \circ \xi(X_r)) \delta_{X_r} dr}{1 + \int_0^t \exp(-\beta A_r \circ \xi(X_r)) dr} \\ \exp(-\beta A_t(z)) = \int_{\mathbb{T}^d} K(z, \xi(x)) \bar{\mu}_t(dx), \quad \forall z \in \mathbb{T}^m, \end{cases} \quad (1.2)$$

where a smooth kernel function  $K: \mathbb{T}^m \times \mathbb{T}^m \rightarrow (0, +\infty)$ , such that  $\int_{\mathbb{T}^m} K(z, \zeta) dz = 1, \forall \zeta \in \mathbb{T}^m$ , is introduced. The diffusion equation (first line of (1.2)) depends on the gradient  $\nabla A_t$  of the real-valued function  $A_t$ , whereas the probability distribution  $\bar{\mu}_t$  on  $\mathbb{T}^d$  does not have a density (second line of (1.2)), hence the need for a smooth kernel.

For a practical implementation of the method, an additional time discretization of the continuous-time dynamics is required. However, in this article, we do not discuss this question, and we only perform the analysis at the continuous-time level.

The expression in the third line in (1.2) is motivated by the definition of the free energy function:

$$\exp(-\beta A_\star(z)) = \int_{\mathbb{T}^{d-m}} \frac{\exp(-\beta V(z, x_{m+1}, \dots, x_d))}{Z(\beta)} dx_{m+1} \dots dx_d. \quad (1.3)$$

To simplify notation, we omit the dependence of  $A_\star$  with respect to the parameter  $\beta$ .

The unknowns in (1.2) are the stochastic processes  $t \mapsto X_t \in \mathbb{T}^d$ ,  $t \mapsto \bar{\mu}_t \in \mathcal{P}(\mathbb{T}^d)$  (the set of Borel probability distributions on  $\mathbb{T}^d$ , endowed with the usual topology of weak convergence of probability distributions), and  $t \mapsto A_t \in \mathcal{C}^\infty(\mathbb{T}^m)$  (the set of infinitely differentiable functions on  $\mathbb{T}^m$ ). Initial conditions  $X_{t=0} = x_0$  and  $\bar{\mu}_{t=0} = \bar{\mu}_0$  are prescribed.

An important observation is that the third equation in (1.2) introduces a coupling between the evolutions of the diffusion process  $X_t$  and of the probability distribution  $\bar{\mu}_t$ . Thus the system defines a type of self-interacting diffusion process. However, a comparison with [4] and subsequent articles [5, 6] and [7], reveals a different form of coupling. One of the aims of this article is to study the new arguments which are required for the study of the system (1.2).

The most important quantity in (1.2) is the random, time-dependent, probability distribution  $\bar{\mu}_t$ . Observe that its construction requires two successive operations to be performed. First, a weighted occupation measure  $\mu_t = \bar{\mu}_0 + \int_0^t \exp(-\beta A_r \circ \xi(X_r)) \delta_{X_r} dr$  is computed. Second, this measure is normalized to define a probability distribution,  $\bar{\mu}_t = \frac{\mu_t}{\int_{x \in \mathbb{T}^d} \mu_t(dx)}$ . The weights  $\exp(-\beta A_r \circ \xi(X_r))$  in the definition of  $\bar{\mu}_t$  are chosen so as to obtain the following consistency result for the estimation of  $\mu_\star$ .

**THEOREM 1.1.** *Almost surely,  $\bar{\mu}_t$  converges to  $\mu_\star$ , in  $\mathcal{P}(\mathbb{T}^d)$ .*

*Moreover, define the function  $A_\infty$ , such that  $\exp(-\beta A_\infty(\cdot)) = \int K(\cdot, \xi(x)) \mu_\star(dx)$ . Then, almost surely,  $A_t$  converges to  $A_\infty$ , in  $\mathcal{C}^k(\mathbb{T}^m)$ ,  $\forall k \in \mathbb{N}$ .*

The strategy of proof of this result is described in our previous work [3]. The present article, provides the technical details in a more general context.

Let us explain the role that the weights play in this result. On the one hand, they are fundamental for the consistency of the method. This observation is not surprising, indeed it is a standard feature of importance sampling approaches. Indeed, Theorem 1.1 is easily seen to be valid for non-adaptive versions of (1.2) (see (2.2) and the convergence result (2.7)), where a bias is initially given and not modified ( $A_t = A$  for all  $t \geq 0$  in the definitions of  $X_t$  and  $\bar{\mu}_t$  in (1.2)). On the other hand, the convergence of  $A_t$  to  $A_\infty$  comes from the way the evolutions of  $X_t$  and  $\bar{\mu}_t$  are coupled, in the third equation of (1.2). The convergence of  $A_t$  to  $A_\infty$  reveals the efficiency of the method: indeed,  $A_\infty$  is an approximation of the free energy function  $A_\star$ , defined by (1.3). Note that, by construction,  $\exp(-\beta A_\star(z)) dz$  is a probability distribution on  $\mathbb{T}^m$ , which is the image of  $\mu_\star$  by the reaction coordinate  $\xi$ . As will be explained below (see Section 2.4), biasing the dynamics (in a non-adaptive way) using the function  $A_\star$  (which is not known in practice) is a natural choice. The adaptive method can thus naturally be seen as a stochastic approximation algorithm, with a parameter being learnt on-the-fly, see e.g. [2, 8, 21, 31].

**1.2. General framework.** The observations made in Section 1.1, concerning the system (1.2) and the consistency result, Theorem 1.1, can be generalized as follows.

First, contrary to (1.1), the state space of the dynamics may not be compact. It may also be of infinite dimension.

The most important generalization concerns the type of diffusion processes which are considered: our general framework also encompasses the following examples (this

list is not exhaustive): (hypoelliptic) Langevin dynamics with position and momenta variables, extended dynamics – where an auxiliary variable is associated with the mapping  $\xi$ , see [38] – and Stochastic Partial Differential Equations (SPDEs) – which are infinite dimensional diffusion processes. It may also be possible to study diffusions on smooth manifolds, however to simplify the presentation this situation is not treated.

Abstract notation and analysis allow us to treat simultaneously these examples in a general framework. However note that the SPDE example is studied separately, in Section 7 to simplify the exposition.

The reaction coordinate  $\xi$  can be an arbitrary, smooth function, with values in a  $m$ -dimensional compact manifold. The associated free energy is then defined in terms of a Radon-Nikodym derivative of the image of the invariant distribution  $\mu_\star$  by  $\xi$ , with respect to a reference measure.

Note that the general framework and the associated abstract notation are constructed to emphasize the most important assumptions, made on the models and on the algorithm, which are required for the well-posedness and the consistency of the approach.

**1.3. Organization of the paper.** In Section 2, the abstract framework is introduced, with emphasis on the following objects: the diffusion process dynamics (Section 2.1), the main examples (Section 2.2), the invariant probability distribution  $\mu$  (Section 2.3), and the free energy function (Section 2.4).

The construction of generalized versions of the Adaptive Biasing Potential method, given by (1.2), is provided in Section 3. In particular, well-posedness results and important estimates are stated precisely there.

Section 4 contains the main results of this article, in finite dimensional cases, concerning the long-time behavior of the method. On the one hand, the consistency of the approach, *i.e.* the almost sure convergence (in distribution) of  $\bar{\mu}_t$  to  $\mu$ , is given in Theorem 4.1 and Corollary 4.1. On the other hand, the efficiency is analyzed first in terms of the convergence of the approximation  $A_t$  of the free energy function, Corollary 4.2, and of occupation measures, Corollary 4.3; second, in terms of the asymptotic mean-square error, Proposition 4.1.

Section 5 is devoted to the proof of the consistency. We first describe (see Section 5.1) how to eliminate the weights in the definition of  $\bar{\mu}_t$ , thanks to a random change of time variable. As explained in [3], the new system may then be treated using the ODE method from stochastic approximation (see [2, 8, 21, 31]), thanks to an asymptotic time scale separation into slow (occupation measure) and fast (diffusion process) evolutions, like in [4]. Section 5.2 then contains all the technical details for a direct proof of the consistency result. Auxiliary properties for solutions of Poisson equations are provided.

Section 6 is devoted to the analysis of the asymptotic mean-square error. As expected, the behavior of the variance for the adaptive system is asymptotically the same as for a non-adaptive system where the bias is chosen as the limit  $A_\infty$  of the adaptive bias  $A_t$ .

Finally, in Section 7 we consider infinite dimensional diffusion processes, which are solutions of SPDEs. This formally fits in the general framework, but a rigorous analysis needs to be performed separately.

## 2. Framework

**2.1. Dynamics and abstract notation.** To simplify notation, without loss of generality, from now on the parameter  $\beta$  is set equal to 1.

**2.1.1. Unbiased dynamics.** The unbiased (or original) dynamics is a diffusion process  $(X_t^0)_{t \geq 0}$ , with values on a state space denoted by  $\mathcal{S}$ . The process is solution of a stochastic differential equation (SDE), when the dimension of  $\mathcal{S}$  is finite; or of a stochastic partial differential equation (SPDE), when the dimension of  $\mathcal{S}$  is infinite. The SDE or the SPDE is written in the following form

$$dX_t^0 = \mathcal{D}(V)(X_t^0)dt + \sqrt{2}\Sigma dW_t, \quad X_0^0 = x_0, \quad (2.1)$$

where  $(W_t)_{t \in \mathbb{R}_+}$  is a standard Wiener process on  $\mathcal{S}$ , and  $\Sigma$  is a linear mapping which is specified in each example below.

In (2.1), the initial condition  $x_0 \in \mathcal{S}$  is arbitrary, and is assumed to be deterministic for simplicity. The convergence results may be extended to a random initial condition (independent of the Wiener noise) by a standard conditioning argument. The value of  $x_0$  plays no role in the analysis below.

The drift coefficient  $\mathcal{D}(V)$  in (2.1) depends on the potential energy function  $V: E_d \rightarrow \mathbb{R}$ , defined on a set  $E_d$ , where  $E_d = \mathbb{T}^d$  (periodic, compact case) or  $E_d = \mathbb{R}^d$  (non compact case). Note that in general  $\mathcal{S} \neq E_d$ . The functions  $V$  and  $\mathcal{D}(V)$  are assumed of class  $\mathcal{C}^\infty$ , also the results can be adapted to deal with the situation when  $V$  and  $\mathcal{D}(V)$  are merely of class  $\mathcal{C}^n$  for sufficiently large  $n \in \mathbb{N}$ . In the non compact case  $E_d = \mathbb{R}^d$ , growth conditions, that will be described for each example, are required.

**2.1.2. Non-adaptively biased dynamics.** Having introduced the unbiased dynamics (2.1), we now describe the family of biased dynamics which we consider in this article, first in a non-adaptive context. The drift coefficient  $\mathcal{D}(V)$  in (2.1) is modified, being replaced by  $\mathcal{D}(V, A)$ , where the function  $A$  depends only on a small number of degrees of freedom of the system. In the current section, the bias is non-adaptive: the function  $A$  is deterministic and does not depend on time.

We now make precise how  $\mathcal{D}(V, A)$  is defined. It depends on the mapping  $A \circ \xi$ , where

- $\xi: E_d \rightarrow \mathbb{M}_m$  is a fixed smooth function, where  $\mathbb{M}_m$  is a  $m$ -dimensional smooth, compact, manifold, and  $m \in \{1, \dots, d-1\}$ .
- $A: \mathbb{M}_m \rightarrow \mathbb{R}$  is a smooth function.

The mapping  $\xi$  is called the reaction coordinate (following the terminology from molecular dynamics applications) and the variables  $z = \xi(x)$  are often called collective variables. The functions  $\xi$  and  $A \circ \xi$  are defined on  $E_d$ , like the potential energy function  $V$ .

Finally, an extension

$$\xi_{\mathcal{S}}: \mathcal{S} \rightarrow \mathbb{M}_m$$

of the reaction coordinate  $\xi$ , is also defined on the state space  $\mathcal{S}$ , with a procedure depending on the example of diffusion process.

In the non-compact case  $E_d = \mathbb{R}^d$ , all the derivatives of  $\xi$  are assumed to be bounded.

As explained in the introduction, in practice one chooses  $m$  much smaller than  $d$ , and typically for concrete applications  $m \in \{1, 2, 3\}$ .

Note that the compactness assumption on  $\mathbb{M}_m$  is crucial in this article. In particular, it allows us to establish some stability estimates and the well-posedness of the ABP system. In some cases, it might be possible to remove this restriction (and consider for instance  $\mathbb{M}_m = \mathbb{R}^m$ ), proving appropriate estimates. We leave this non trivial technical issue for future works. The assumption that  $\mathbb{M}_m$  is a smooth manifold is required to define potential energy functions  $V - A \circ \xi$  with nice regularity properties.

To simplify the discussion, from now on  $\mathbb{M}_m = \mathbb{T}^m$  is the flat  $m$ -dimensional torus. However, we use the abstract notation and conditions to suggest possible straightforward generalizations.

We are now in a position to define the biased dynamics  $(X_t^A)_{t \geq 0}$ , for any given  $A: \mathbb{M}_m \rightarrow \mathbb{R}$  of class  $\mathcal{C}^\infty$ :

$$dX_t^A = \mathcal{D}(V, A)(X_t^A)dt + \sqrt{2}\Sigma dW_t, \quad X_0^A = x_0. \quad (2.2)$$

Consistently, we will have  $\mathcal{D}(V, 0) = \mathcal{D}(V)$ : in the absence of bias, the biased dynamics (2.2) is simply the unbiased dynamics (2.1).

**2.2. Examples of diffusions processes.** In this section, we present the three main examples of diffusion processes to be studied. We postpone the study of a fourth example, given by infinite dimensional diffusion processes (SPDE), to Section 7.

From now on, except in Section 7, the state space  $\mathcal{S}$  is finite dimensional.

### 2.2.1. Brownian dynamics.

- **State space:**  $\mathcal{S} = E_d$ .
- **Reaction coordinate:**  $\xi_{\mathcal{S}} = \xi$ .
- **Drift coefficient:**  $\mathcal{D}(V, A) = \mathcal{D}(V - A \circ \xi) = -\nabla(V - A \circ \xi)$ . **Diffusion operator:**  $\Sigma = I$ , where  $I$  denotes the identity matrix.

In the Brownian case, the dynamics (2.2) is written as

$$dx_t^A = -\nabla(V - A \circ \xi)(x_t^A)dt + \sqrt{2}dW_t.$$

In the non compact case,  $E_d = \mathbb{R}^d$ , the potential energy function  $V$  is assumed to satisfy the conditions below.

ASSUMPTION 2.1. *When  $E_d = \mathbb{R}^d$ , there exist  $\alpha_V \in (0, \infty)$  and  $C_V \in \mathbb{R}$ , such that for all  $x \in E_d$ ,*

$$\langle x, \nabla V(x) \rangle \geq \alpha_V |x|^2 - C_V.$$

*Moreover,  $V$  is semi-convex:  $V = V_1 + V_2$  where  $V_1$  is a smooth bounded function, with bounded derivatives, and  $V_2$  is a smooth convex function.*

*For all  $k \in \mathbb{N}$ ,  $\int_{\mathbb{R}^d} |x|^k e^{-V(x)} dx < \infty$ .*

*Finally, there exists  $c_V \in (0, \infty)$ ,  $\varpi \in \mathbb{N}$ , such that for all  $x \in \mathbb{R}^d$ ,*

$$|V(x)| \leq c_V(1 + |x|^\varpi).$$

Assumption 2.1 is satisfied for instance for smooth potential functions  $V$  which behave like  $|\cdot|^\varpi$  at infinity.

REMARK 2.1. Assume  $E_d = \mathbb{R}^d$ . The Brownian dynamics defined above is reversible, but the framework also encompasses non-reversible situations. For instance, let  $J \neq 0$  be a  $d \times d$  skew-symmetric matrix. Then one may also define the drift coefficient as  $\mathcal{D}(V, A) = -(I + J)\nabla(V - A \circ \xi)$ .

### 2.2.2. Langevin dynamics.

- **State space:**  $\mathcal{S} = E_d \times \mathbb{R}^d$  (which is not compact). Elements of  $\mathcal{S}$  are denoted by  $(q, p)$ .
- **Reaction coordinate:**  $\xi_{\mathcal{S}}(q, p) = \xi(q)$ .

- **Drift coefficient:**  $\mathcal{D}(V, A)(q, p) = \begin{pmatrix} -\nabla(V - A \circ \xi)(q) - \gamma p \\ 0 \end{pmatrix}$ . **Diffusion operator:**  $\Sigma = \sqrt{\gamma} \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}$ , for  $\gamma \in (0, \infty)$  a damping parameter.

In the Langevin case, the dynamics (2.2) is written

$$\begin{cases} dq_t^A = p_t^A dt, & q_0^A = q_0, \\ dp_t^A = -\nabla(V - A \circ \xi)(q_t^A) dt - \gamma p_t^A dt + \sqrt{2\gamma} d\tilde{W}_t, & p_0^A = p_0, \end{cases}$$

where  $(\tilde{W}_t)_{t \geq 0}$  is a standard Wiener process on  $\mathbb{R}^d$ . In applications, the variable  $q$  represents positions of particles, whereas the variable  $p$  represents their momenta.

The value of the damping parameter  $\gamma$  plays no role in the analysis below. We recall that in the limit  $\gamma \rightarrow \infty$ , one recovers (up to a rescaling of the time variable) the Brownian dynamics of Section 2.2.1, which is thus often referred to as the overdamped Langevin dynamics. Recall also that the analysis of these two cases is different, since the Langevin diffusion is hypoelliptic, whereas the Brownian dynamics is elliptic.

In the non-compact case,  $E_d = \mathbb{R}^d$ , the potential energy function  $V$  is assumed to satisfy the conditions below.

**ASSUMPTION 2.2.** *When  $E_d = \mathbb{R}^d$ , there exists  $\kappa_V \in (0, \infty)$  and  $V_- \in \mathbb{R}$  such that  $V(q) \geq \kappa_V |q|^2 + V_-$  for all  $q \in \mathbb{R}^d$ .*

*There exist  $A_V, B_V, \kappa_V \in (0, \infty)$  and  $C_V \in \mathbb{R}$ , such that for all  $q \in \mathbb{R}^d$ ,*

$$\langle q, \nabla V(q) \rangle \geq A_V V(q) + B_V |q|^2 + C_V.$$

*Moreover,  $V$  is semi-convex:  $V = V_1 + V_2$  where  $V_1$  is a smooth bounded function, with bounded derivatives, and  $V_2$  is a smooth convex function.*

*Finally, there exists  $c_V \in (0, \infty)$ ,  $\varpi \in \mathbb{N}$ , such that for all  $q \in \mathbb{R}^d$ ,*

$$|V(q)| \leq c_V (1 + |q|^\varpi).$$

**2.2.3. Extended dynamics.** This example is a modification of the Brownian dynamics from Section 2.2.1. It is straightforward to build a similar modification of the Langevin dynamics of Section 2.2.2, the details are left to the reader.

- **State space:**  $\mathcal{S} = E_d \times \mathbb{M}_m$ . Elements of  $\mathcal{S}$  are denoted by  $(x, z)$ .
- **Reaction coordinate:**  $\xi_{\mathcal{S}}(x, z) = z$ .
- **Drift coefficient:**  $\mathcal{D}(V, A)(x, z) = \begin{pmatrix} -\nabla_x U_A(x, z) \\ -\nabla_z U_A(x, z) \end{pmatrix}$  where  $U_A(x, z) = U(x, z) - A(z)$ ,  $U(x, z) = V(x) + \frac{1}{2\epsilon} V_{\text{ext}}(\xi(x), z)$  is the extended potential energy function. It depends on a smooth function  $V_{\text{ext}}: \mathbb{M}_m \times \mathbb{M}_m \rightarrow \mathbb{R}$ , and on  $\epsilon \in (0, \infty)$ . **Diffusion operator:**  $\Sigma$  is the identity.

In the case  $\mathbb{M}_m = \mathbb{T}^m$  considered here, one may choose  $V_{\text{ext}}(\xi(x), z) = (\xi(x) - z)^2$ . Then, in the (Brownian) extended case, the dynamics (2.2) is written as

$$\begin{cases} dX_t^A = -\nabla V(X_t^A) dt - \frac{1}{\epsilon} \langle \nabla \xi(X_t^A), \xi(X_t^A) - Z_t^A \rangle dt + \sqrt{2} dW_t^x, & X_0^A = x_0, \\ dZ_t^A = -\frac{1}{\epsilon} (Z_t^A - \xi(X_t^A)) dt + \nabla A(Z_t^A) dt + \sqrt{2} dW_t^z, & Z_0^A = z_0, \end{cases}$$

for some arbitrary initial condition  $z_0 \in \mathbb{M}_m$ , and where  $(W_t^x)_{t \geq 0}$  and  $(W_t^z)_{t \geq 0}$  are independent standard Wiener processes, on  $E_d$  and on  $\mathbb{M}_m$  respectively. The dynamics



is thus obtained by considering the Brownian dynamics on  $E_d \times \mathbb{M}_m$ , with potential energy function  $U$ .

As a consequence, one could directly write the extended dynamics in the framework of Section 2.2.1. Our choice emphasizes the role of the extended dynamics as an algorithmic tool, which is available for the practitioners.

We will explain below why the extended dynamics is relevant, in the limit  $\epsilon \rightarrow 0$ , for the problem of sampling the initial distribution, and thus why it may be sufficient to deal with the extended dynamics case. Then, since  $\xi_{\mathcal{S}}(x, z) = z$  in this example, it would not be restrictive to consider reaction coordinates of the form  $\xi(x_1, \dots, x_d) = (x_1, \dots, x_m)$ .

In the non compact case,  $E_d = \mathbb{R}^d$ , it is assumed that  $V$  satisfies Assumption 2.1 (or Assumption 2.2 if one starts from the Langevin dynamics). Then the extended potential energy function  $U$  also satisfies a similar condition (recall that  $\mathbb{M}_m$  is compact) on the extended state space  $E_d \times \mathbb{M}_m$ .

**2.3. Invariant probability distributions of the diffusion processes.** In all the examples presented above in Section 2.2, the diffusion processes,  $(X_t^0)_{t \geq 0}$  and  $(X_t^A)_{t \geq 0}$ , given by (2.1) and (2.2), are ergodic. The associated unique invariant distributions, defined on  $\mathcal{S}$  (equipped with the Borel  $\sigma$ -field), are denoted by  $\mu_\star^0$  and  $\mu_\star^A$ . Since the notation is consistent when  $A=0$ , we only deal with  $\mu_\star^A$ , with arbitrary  $A$ , in the remainder of this section.

The ergodicity in our context is understood in the following sense:

- there exists a unique invariant probability distribution for the Markov process  $X^A$  defined by (2.2), which is equal to  $\mu_\star^A$ ;
- for any initial condition  $x_0 \in \mathcal{S}$ , almost surely,

$$\frac{1}{t} \int_0^t \delta_{X_\tau^A} d\tau \xrightarrow[t \rightarrow \infty]{} \mu_\star^A,$$

where the notation  $\xrightarrow[t \rightarrow \infty]{} \mu_\star^A$  stands for the convergence of probability distributions on  $\mathcal{S}$ . Recall that, if  $(\mu_n)_{n \in \mathbb{N}}$  and  $\mu$  are probability distributions on  $\mathcal{S}$ , then

$$\mu_n \xrightarrow[n \rightarrow \infty]{} \mu$$

if  $\mu_n(\varphi) \rightarrow \mu(\varphi)$  for every bounded continuous function  $\varphi: \mathcal{S} \rightarrow \mathbb{R}$ .

The invariant distribution  $\mu_\star^A$  is expressed explicitly in terms of the following data:

- a reference Borel,  $\sigma$ -finite, measure  $\lambda$  on  $\mathcal{S}$ , which does not depend on  $V$  and  $A$ ;
- a total energy function  $\mathcal{E}(V, A): \mathcal{S} \rightarrow \mathbb{R}$ .

The expression of  $\mu_\star^A$  is then given by:

$$\mu_\star^A(dx) = \frac{\exp(-\mathcal{E}(V, A)(x))}{Z^A} \lambda(dx), \quad (2.3)$$

where  $Z^A = \int_{\mathcal{S}} \exp(-\mathcal{E}(V, A)(x)) \lambda(dx)$  is a normalizing constant.

Computing averages  $\mu_\star^A(\varphi) = \int_{\mathcal{S}} \varphi d\mu_\star^A$ , for instance with  $A=0$ , is typically a challenging computational task. This may be due to the large dimension of the state space, or to the multimodality of the measure. Importance sampling techniques, as considered in this article, consist in proposing choices of functions  $A$  such that it is cheaper to sample  $\mu_\star^A$  than  $\mu_\star$ .

Let us make precise the reference measure  $\lambda$  and the mapping  $(V, A) \mapsto \mathcal{E}(V, A)$  for the diffusion processes of Section 2.2. First, the total energy function satisfies the identity

$$\mathcal{E}(V, A) = \mathcal{E}(V, 0) - A \circ \xi_S. \quad (2.4)$$

It thus remains to specify the mapping  $\mathcal{E}(V) = \mathcal{E}(V, 0)$  and the measure  $\lambda$  for each example.

- Brownian dynamics. The reference measure  $\lambda$  is the Lebesgue measure on  $\mathcal{S}$ . The total energy function is  $\mathcal{E}(V) = V$ .
- Langevin dynamics. The reference measure  $\lambda$  is the Lebesgue measure on  $\mathcal{S}$ . The total energy function is the Hamiltonian function,  $\mathcal{E}(V)(q, p) = H(q, p) = V(q) + \frac{|p|^2}{2}$ . The total energy is thus the sum of potential and kinetic energies.
- Extended dynamics. The reference measure  $\lambda$  is the Lebesgue measure on  $\mathcal{S} = E_d \times \mathbb{M}_m$ . The total energy function is  $\mathcal{E}(V)(x, z) = U(x, z) = V(x) + \frac{1}{2\epsilon} (\xi(x) - z)^2$ .

The ergodicity of the dynamics (2.1), resp. (2.2), with unique invariant probability distribution  $\mu_\star^0$ , resp.  $\mu_\star^A$ , is well-known. We refer to Appendix 8.1. In addition (see Proposition 8.1), the convergence is exponentially fast, with a rate which can be controlled, uniformly with respect to  $A$ , thanks to the following auxiliary result.

PROPOSITION 2.1. *Let  $m, M, M^{(1)}, M^{(2)}, \dots \in \mathbb{R}$  denote real numbers, and*

$$\mathcal{A} \subset \left\{ A \in C^\infty(\mathbb{M}_m, \mathbb{R}) ; \min A \geq m, \max A \leq M, \max |\partial^k A| \leq M^{(k)}, \forall k \geq 1 \right\},$$

where  $\partial^k$  denotes the derivative of order  $k$ .

Then

- if  $V$  satisfies Assumption 2.1, there exists  $\alpha_{V, \mathcal{A}} \in (0, \infty)$  and  $C_{V, \mathcal{A}} \in (0, \infty)$  such that for every  $A \in \mathcal{A}$  and every  $x \in E_d = \mathbb{R}^d$ ,

$$\langle x, \nabla(V - A \circ \xi)(x) \rangle \geq \alpha_{V, \mathcal{A}} |x|^2 - C_{V, \mathcal{A}}.$$

- if  $V$  satisfies Assumption 2.2, there exists  $A_{V, \mathcal{A}}, B_{V, \mathcal{A}} \in (0, \infty)$  and  $C_{V, \mathcal{A}} \in (0, \infty)$  such that for every  $A \in \mathcal{A}$  and every  $q \in E_d = \mathbb{R}^d$ ,

$$\langle q, \nabla(V - A \circ \xi)(q) \rangle \geq A_{V, \mathcal{A}} V(q) + B_{V, \mathcal{A}} |q|^2 + C_{V, \mathcal{A}}.$$

- For every  $k \geq 1$ ,

$$\sup_{A \in \mathcal{A}} \int_{\mathcal{S}} |x|^k \mu_\star^A(dx) < \infty. \quad (2.5)$$

The distribution of interest, in practice, is  $\mu_\star = \mu_\star^0$ . However, sampling the process  $X^A$  provides an approximation of  $\mu_\star^A$ . The following expression provides a way to compute an average  $\mu_\star(\varphi) = \int_{\mathcal{S}} \varphi(x) \mu_\star(dx)$  in terms of averages with respect to  $\mu_\star^A$ : for bounded and continuous functions  $\varphi: \mathcal{S} \rightarrow \mathbb{R}$ ,

$$\mu_\star(\varphi) = \frac{\mu_\star^A(\varphi \exp(-A \circ \xi_S))}{\mu_\star^A(\exp(-A \circ \xi_S))}. \quad (2.6)$$

Using (2.4), averages with respect to  $\mu_\star$  may therefore be approximated by temporal averages along the biased dynamics (2.2): indeed, the quantity  $\bar{\mu}_t^A$  defined by the left-hand side below satisfies

$$\bar{\mu}_t^A(\varphi) := \frac{1 + \int_0^t e^{-A(\xi_S(X_r^A))} \varphi(X_r^A) dr}{1 + \int_0^t e^{-A(\xi_S(X_r^A))} dr} \xrightarrow{t \rightarrow +\infty} \frac{\mu_\star^A(\varphi \exp(-A \circ \xi_S))}{\mu_\star^A(\exp(-A \circ \xi_S))} = \mu_\star(\varphi). \quad (2.7)$$

This expression serves as the guideline for the construction of the Adaptive Biasing Potential methods (1.2), and (3.2) in the general case: the empirical distributions  $\bar{\mu}_t$  are weighted, to ensure consistency. For well chosen functions  $A$ , the convergence is expected to be faster than when  $A=0$ . In Section 2.4 below, we identify such a function  $A$ , the so-called free energy function.

**2.4. The Free Energy function.** In this section, we introduce one of the key quantities in our study: the free energy function  $A_\star: \mathbb{M}_m \rightarrow \mathbb{R}$ . We explain why this function is a quantity of interest for the computational problem we are interested in, and why it is expected that choosing  $A=A_\star$  in the biased dynamics (2.2) leads to efficient sampling. This property is indeed the guideline of the Adaptive Biasing Potential approach of this article: we construct an adaptive version which is both consistent and designed such that  $A_t$  converges to an approximation  $A_\infty$  of  $A_\star$  when  $t \rightarrow \infty$ .

The definition of the free energy function depends on the choice of a reference probability distribution  $\pi$  on  $\mathbb{M}_m$ . In this article, since  $\mathbb{M}_m = \mathbb{T}^m$ , it is natural to choose the Lebesgue measure, but abstract notation suggests other possible choices, see Remark 2.2 below.

For every smooth  $A: \mathbb{M}_m \rightarrow \mathbb{R}$ , let  $\pi_\star^A$  denote the image by  $\xi_S: \mathcal{S} \rightarrow \mathbb{M}_m$  of the probability distribution  $\mu_\star^A$  on  $\mathcal{S}$ . Recall that this means that for any bounded, continuous function  $\phi: \mathbb{M}_m \rightarrow \mathbb{R}$ ,

$$\int_{\mathbb{M}_m} \phi(z) \pi_\star^A(dz) = \int_{\mathcal{S}} \phi(\xi_S(x)) \mu_\star^A(dx).$$

The following assumption is required.

**ASSUMPTION 2.3.** *The measures  $\pi_\star^0$  and  $\pi$  are equivalent:  $\pi_\star^0$  (resp.  $\pi$ ) is absolutely continuous with respect to  $\pi$  (resp.  $\pi_\star^0$ ).*

When Assumption 2.3 holds true, then  $\pi_\star^A$  is equivalent to  $\pi$ , for all smooth functions  $A: \mathbb{M}_m \rightarrow \mathbb{R}$ . Thanks to the smoothness conditions on  $\xi$ , and to growth conditions on  $V$ , Assumption 2.3 is satisfied in all the examples presented above, when  $\pi$  is the Lebesgue measure on  $\mathbb{M}_m = \mathbb{T}^m$ .

**REMARK 2.2.** Another natural choice, in the periodic case  $E_d = \mathbb{T}^d$ , for finite dimensional dynamics, is as follows:  $\pi$  is defined as the image by  $\xi: E_d \rightarrow \mathbb{M}_m$  of the Lebesgue measure on  $\mathbb{E}_d$ . With this definition,  $\pi$  depends on  $\xi$ . In the non-compact case, for instance one may define (for instance)  $\pi$  as the image by  $\xi$  of the standard Gaussian distribution on  $\mathbb{R}^d$ .

With these examples, Assumption 2.3 is satisfied by construction of  $\pi$ .

We are now in a position to define the free energy function  $A_\star$ .

**DEFINITION 2.1.** *The free energy function  $A_\star: \mathbb{M}_m \rightarrow \mathbb{R}$  is defined by the following property:  $\exp(-A_\star(\cdot))$  is the Radon-Nikodym derivative of  $\pi_\star^0$  with respect to  $\pi$ .*

This means that for every bounded measurable function  $\phi: \mathbb{M}_m \rightarrow \mathbb{R}$ ,

$$\int_{\mathbb{M}_m} \phi(z) e^{-A_*(z)} \pi(dz) = \int_{\mathcal{S}} \phi(\xi_{\mathcal{S}}(x)) \mu_*(dx).$$

Observe that, thanks to Assumption 2.3,  $A_*$  takes values in  $(-\infty, \infty)$ . Moreover,  $e^{-A_*(z)} \pi(dz)$  is by construction a probability distribution on  $\mathbb{M}_m$ , thus no normalizing constant appears on the left-hand side.

It is then straightforward to check that the Radon-Nikodym derivative of  $\pi_*^A$  with respect to  $\pi$  is equal to  $\exp(-A_* + A)$ , thanks to the condition (2.4).

The function  $A_*$  may be interpreted as an effective potential energy function, for the unbiased dynamics, depending on the variable  $z = \xi_{\mathcal{S}}(x)$  only. Indeed, note that for any sufficiently smooth, bounded, function  $\phi: \mathbb{M}_m \rightarrow \mathbb{R}$ , by ergodicity of the unbiased dynamics (2.1), with respect to  $\mu_*$ , almost surely

$$\frac{1}{t} \int_0^t \phi(\xi_{\mathcal{S}}(X_r^0)) dr \xrightarrow[t \rightarrow \infty]{} \int_{\mathcal{S}} \phi \circ \xi_{\mathcal{S}} d\mu_*^0 = \int_{\mathbb{M}_m} \phi d\pi_*^0 = \int_{\mathbb{M}_m} \phi(z) e^{-A_*(z)} d\pi(z).$$

Similarly, when considering the biased dynamics,

$$\frac{1}{t} \int_0^t \phi(\xi_{\mathcal{S}}(X_r^A)) dr \xrightarrow[t \rightarrow \infty]{} \int_{\mathbb{M}_m} \phi(z) e^{-A_*(z) + A(z)} d\pi(z).$$

We now give an interpretation of the qualitative properties of the free energy function  $A_*$ . Assume that  $\pi$  is the Lebesgue measure on  $\mathbb{M}_m$ , and that  $A_*$  admits several local minima: then the distribution  $\pi_*^0$  is multimodal, and the convergence to equilibrium, when using the unbiased dynamics, is slow. Indeed, the process must visit regions near all the local minima of  $A_*$ , and transitions between these metastable states are rare events. Thus  $A_*$  encodes the metastability of the dynamics along the variable  $z = \xi(z) \in \mathbb{M}_m$ .

On the contrary, if the biased dynamics with  $A = A_*$  is used, the associated ergodicity result indicates that convergence is expected to be faster – at least if the convergence in the other variables is not slow due to metastability. Indeed, the repartition of the values of  $\xi_{\mathcal{S}}(X_t^A)$  tends to be uniform when  $t \rightarrow \infty$ ; this is the flat-histogram property which is the guideline of the strategies mentioned in Section 1.

Note also that, in many applications (for instance in molecular dynamics), computing free energy differences, *i.e.*  $A_*(z_1) - A_*(z_2)$ , may be the ultimate goal of the simulation, instead of computing averages  $\int \varphi d\mu_*$ . The Adaptive Biasing Potential methods of this article can also be seen as efficient free energy computation algorithms.

Since in general the free energy function is not known, the associated biased dynamics with  $A = A_*$  cannot be simulated in practice; the guideline of the adaptive version proposed and analyzed below is to (approximately) reproduce the nice flat-histogram property for variable  $z = \xi(x)$  in the asymptotic regime  $t \rightarrow \infty$ , without a priori knowing the free energy function  $A_*$ ; moreover an estimation  $A_*$  is also computed.

### 3. ABP: construction and well-posedness

In this section, the construction of the Adaptive Biasing Potential (ABP) system is performed in the general framework of Section 2. The rigorous construction of the process, and the statement of appropriate assumptions, is one of the contributions of this paper. The ABP system is built starting from the unbiased dynamics (2.1), with an adaptive bias  $A = A_t$  (random and depending on time  $t$ ) introduced in the biased

dynamics (2.2). The construction is a generalization of (1.2), considered in Section 1.1 in a simplified setting.

In an abstract framework, the coupling of the evolutions of the diffusion process  $X_t$  and of the bias  $A_t$  requires the introduction of several auxiliary tools, with details provided below.

- A kernel function  $K : \mathbb{M}_m \times \mathbb{M}_m \rightarrow (0, \infty)$ , see Assumption 3.1. Then, a mapping  $\mathcal{K} : \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{C}^\infty(\mathbb{M}_m)$  is defined by

$$\mathcal{K}(\bar{\mu})(z) = \int_{\mathcal{S}} K(z, \xi_{\mathcal{S}}(x)) \bar{\mu}(dx). \quad (3.1)$$

- A normalization operator  $\mathcal{N} : \mathcal{C}^0(\mathbb{M}_m, (0, \infty)) \rightarrow \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$ , on the set of continuous functions on  $\mathbb{M}_m$  with values in  $(0, \infty)$ .
- The notation  $\bar{F}$  is defined by

$$\bar{F}(z) = \frac{F(z)}{\int_{\mathbb{M}_m} F d\pi}.$$

The ABP system in its general formulation is written as follows:

$$\begin{cases} dX_t = \mathcal{D}(V, A_t)(X_t)dt + \sqrt{2}\Sigma dW_t, \\ \bar{\mu}_t = \frac{\bar{\mu}_0 + \int_0^t F_\tau(\xi_{\mathcal{S}}(X_\tau))\delta_{X_\tau} d\tau}{1 + \int_0^t F_\tau(\xi_{\mathcal{S}}(X_\tau))d\tau}, \\ F_t = \mathcal{N}(\mathcal{K}(\bar{\mu}_t)) = \mathcal{N}\left(\int_{\mathcal{S}} K(\cdot, \xi_{\mathcal{S}}(x))\bar{\mu}_t(dx)\right), \\ A_t = -\log(\bar{F}_t), \end{cases} \quad (3.2)$$

where there are four unknown processes:  $(X_t)_{t \geq 0}$  (with values in  $\mathcal{S}$ ),  $(\bar{\mu}_t)_{t \geq 0}$  (with values in  $\mathcal{P}(\mathcal{S})$  the set of probability distributions on  $\mathcal{S}$ ),  $(F_t)_{t \geq 0}$  (with values in  $\mathcal{C}^0(\mathbb{M}_m, (0, \infty))$ ), and  $(A_t)_{t \geq 0}$  (with values in  $\mathcal{C}^\infty(\mathbb{M}_m)$ ). Note that the initial conditions  $F_0 = \mathcal{N}(\mathcal{K}(\bar{\mu}_0))$  and  $A_0 = -\log(\bar{F}_0)$  are prescribed by the initial condition  $\bar{\mu}_0$ ; we also set  $X_0 = x_0$ .

Observe that it is not necessary to consider the four unknowns in (3.2). Indeed, as will be explained below,  $F_t$  and  $\bar{F}_t = \exp(-A_t)$  only differ by a multiplicative constant (depending on  $t$ ), which is determined only by the choice of the normalization operator  $\mathcal{N}$ . Moreover, it would be possible to consider only the processes  $(X_t)_{t \geq 0}$  and  $(A_t)_{t \geq 0}$  to define the dynamics of the ABP system; however, we wish to emphasize the role of the probability distribution  $\bar{\mu}_t$ , this is why it is included explicitly in (3.2).

Important observations concerning the system (3.2) are in order.

The diffusion process is biased, following (2.2), and the bias  $A_t$  at time  $t$  is defined in terms of the values  $(X_r)_{0 \leq r \leq t}$  of the diffusion process up to time  $t$ . As a consequence, the diffusion process in (3.2) can be considered as a self-interacting diffusion on  $\mathcal{S}$ . However, the standard framework of self-interacting processes does not encompass the system (3.2), and we thus need to adapt and generalize the arguments concerning well-posedness and convergence in our setting.

The function  $A_t$  is constructed in order to be an approximation, in the regime  $t \rightarrow \infty$ , of the free energy function  $A_\star$ , introduced in Section 2.4; indeed, knowing  $A_\star$  would lead to an optimal non-adaptive biased dynamics. The adaptive system is designed to approximate  $A_\star$  both adaptively and efficiently.

As already mentioned in the introduction (see Theorem 1.1), the central object in the analysis is the probability distribution  $\bar{\mu}_t$ . Indeed, we will prove that it converges

almost surely to  $\mu_*$ , see Theorem 4.1. Note that  $\bar{\mu}_t$  is defined as a weighted empirical distribution, with weights  $F_\tau(\xi_{\mathcal{S}}(X_\tau))$ ; this choice is motivated by (2.6) (in the non-adaptive setting).

Below, we state assumptions on the kernel and on the normalization operator, which play a key role first for the well-posedness of the algorithms, second for the analysis of its asymptotic behavior. In the sequel, the assumptions on the model, stated in Section 2, are satisfied.

**3.1. Kernel.** The kernel function  $K : (z, \zeta) \in \mathbb{M}_m \times \mathbb{M}_m \mapsto K(z, \zeta) \in (0, +\infty)$  is a continuous, positive, smooth function. In the following, this function is often referred to as the regularization kernel, and it is assumed to satisfy the conditions below.

**ASSUMPTION 3.1.** *The function  $K$  is positive, of class  $\mathcal{C}^\infty$  on  $\mathbb{M}_m \times \mathbb{M}_m$ . Moreover, for all  $\zeta \in \mathbb{M}_m$ , the normalization condition  $\int_{\mathbb{M}_m} K(z, \zeta) \pi(dz) = 1$  is satisfied.*

Since  $\mathbb{M}_m$  is compact, one has  $m(K) = \min_{z, \zeta \in \mathbb{M}_m} K(z, \zeta) > 0$ , and, for all integers  $r \in \{0, 1, \dots\}$ ,  $M^{(r)}(K) = \sup_{z, \zeta \in \mathbb{M}_m} |\partial_z^r K(z, \zeta)| < +\infty$ . Moreover,  $\sup_{z \in \mathbb{M}_m} \sup_{\zeta_1, \zeta_2} \frac{K(z, \zeta_1) - K(z, \zeta_2)}{d(\zeta_1, \zeta_2)} < +\infty$  (Lipschitz continuity in the second variable, uniformly in the first variable).

The mapping  $\mathcal{K} : \bar{\mu} \in \mathcal{P}(\mathcal{S}) \mapsto \mathcal{K}(\bar{\mu}) \in \mathcal{C}^\infty(\mathbb{M}_m)$ , is then defined by (3.1) above. Note that  $\int_{\mathcal{S}} \mathcal{K}(\bar{\mu})(z) \pi(dz) = 1$ , and that the mapping  $\mathcal{K}(\bar{\mu})$  is of class  $\mathcal{C}^\infty$ , thanks to Assumption 3.1. Note also that (3.1) also makes sense if the probability distribution  $\bar{\mu}$  is replaced with a positive, finite, measure  $\mu$ .

One may consider the following example of kernel  $K$ , in the case  $\mathbb{M}_m = \mathbb{T}^m$ . Let  $k : \mathbb{R}^m \rightarrow (0, \infty)$  be an even function of class  $\mathcal{C}^\infty$ , with bounded derivatives, such that  $\int_{\mathbb{M}_m} k(z) \pi(dz) = 1$ . For  $\epsilon \in (0, 1)$ , let  $K(z, \zeta) = \frac{1}{\epsilon} k\left(\frac{z - \zeta}{\epsilon}\right)$ . In the regime  $\epsilon \rightarrow 0$ , such kernels  $K = K_\epsilon$  are smooth mollifiers. If the function  $k$  is chosen with compact support, the positivity condition on  $K$  is satisfied by choosing  $K(z, \zeta) = \frac{\alpha}{\epsilon} k\left(\frac{z - \zeta}{\epsilon}\right) + 1 - \alpha$ , with  $\alpha \in (0, 1)$ .

It may also be useful to consider kernel functions which are not homogeneous, *i.e.*  $K(z, \zeta)$  does not depend only on  $z - \zeta$ . For instance, set  $K(z, \zeta) = \sum_{n=1}^N K_n(z, \zeta) \theta_n(\zeta)$ , where  $N \in \mathbb{N}$ ,  $K_1, \dots, K_N$  are kernel functions satisfying Assumption 3.1, and  $\theta_1, \dots, \theta_N$  are smooth functions  $\mathbb{M}_m \rightarrow (0, \infty)$ , such that  $\sum_{n=1}^N \theta_n(\zeta) = 1$  for all  $\zeta \in \mathbb{M}_m$ . Such examples are useful to build a bias which takes into account local properties.

Note that a symmetry assumption for the kernel  $-K(z, \zeta) = K(\zeta, z) -$  is not required to prove the consistency of the approach. For instance, assume that  $K(z, \zeta) = \tilde{K}(z)$  does not depend on  $\zeta$ ; in this case, one checks that  $\mathcal{K}(\bar{\mu}_t) = \mathcal{K}(\bar{\mu}_0) = \tilde{K}(\cdot)$  does not depend on  $t$ , and thus  $A_t = A_0$ : the adaptive system (3.2) reduces for this choice of kernel to the non-adaptive biased dynamics (2.2). Based on this observation, it is clear that the kernel  $K$  is the object which governs the coupling of the evolutions of  $X$  and  $A$  in the adaptive dynamics (3.2), and that its choice may be crucial in practice to define an efficient algorithm. In the sequel, we consider that a kernel function  $K$ , satisfying Assumption 3.1, is given, and do not study quantitatively the dependence with respect to  $K$  of the asymptotic results.

**3.2. Normalization.** The aim of this section is to introduce normalization operators, denoted by  $\mathcal{N} : \mathcal{C}^0(\mathbb{M}_m, (0, \infty)) \rightarrow \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$  on the set of continuous functions from  $\mathbb{M}_m$  to  $(0, \infty)$ . The compactness of  $\mathbb{M}_m$  plays a crucial role again. We provide below several natural families of normalization operators. However, the presentation remains abstract to emphasize the key assumptions which will lead to the stability estimates provided below.

We will use the following convention:  $f$  denotes an arbitrary element in  $\mathcal{C}^0(\mathbb{M}_m, (0, \infty))$ , whereas  $F = \mathcal{N}(f)$  (capital letter) denotes its normalized version.

The most important example, for which a specific notation is introduced, is when normalization is meant to construct probability distributions  $\bar{f}d\pi$  which are equivalent to the reference measure  $\pi$  on  $\mathbb{M}_m$ :

$$\bar{f}(z) = \frac{f(z)}{\int_{\mathbb{M}_m} f(\zeta)\pi(d\zeta)}.$$

In the ABP system (3.2),  $\exp(-A_t)$  is thus the density (with respect to  $\pi$ ) of a probability distribution on  $\mathbb{M}_m$ , for every  $t \geq 0$ .

More generally, the normalization operator  $\mathcal{N}$  is defined by

$$\mathcal{N}(f) = \frac{f}{n(f)},$$

where  $n: \mathcal{C}^0(\mathbb{M}_m, (0, \infty)) \rightarrow (0, \infty)$  is a function which satisfies the technical (but easy to check in practice) conditions presented below.

**ASSUMPTION 3.2.** *The operator  $n: \mathcal{C}^0(\mathbb{M}_m, (0, \infty)) \rightarrow (0, \infty)$  satisfies the following conditions.*

- *There exists a sequence  $(n^{(k)})_{k \in \mathbb{N}}$ , such that, for every  $k \in \mathbb{N}$ ,  $n^{(k)}: \mathcal{C}^0(\mathbb{M}_m, (0, \infty)) \rightarrow (0, \infty)$  is continuously differentiable, and for every  $f \in \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$ ,*

$$n^{(k)}(f) \xrightarrow{k \rightarrow \infty} n(f);$$

*moreover the convergence is assumed to be uniform on sets of the form*

$$\{f \in \mathcal{C}^0(\mathbb{M}_m, (0, \infty)) ; \min f \geq m, \max f \leq M\},$$

*for every  $0 < m \leq M < \infty$ .*

- *There exists  $\gamma_n \in (0, \infty)$  such that for all  $f \in \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$  and  $k \in \mathbb{N}^*$*

$$\frac{1}{\gamma_n} \min f \leq n^{(k)}(f) \leq \gamma_n \max f.$$

- *For all  $f \in \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$ ,  $\alpha \in (0, \infty)$  and  $k \in \mathbb{N}^*$*

$$n^{(k)}(\alpha f) = \alpha n^{(k)}(f).$$

- *There exists  $C_n \in (0, \infty)$  such that for all  $f_1, f_2 \in \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$  and  $k \in \mathbb{N}^*$*

$$|n^{(k)}(f_1) - n^{(k)}(f_2)| \leq C_n \max |f_1 - f_2|.$$

Only the continuous differentiability condition is relaxed when considering the limit  $k \rightarrow \infty$ :  $n$  is not required to satisfy this condition. The three other conditions are satisfied when  $n^{(k)}$  is replaced with  $n$ .

Let us provide some important consequences of the definition of  $\mathcal{N}$  in terms of an operator  $n$  satisfying Assumption 3.2. First, note that  $\mathcal{N} \circ \mathcal{N} = \mathcal{N}$ : the normalization operator is a projection. Moreover,  $F = \mathcal{N}(f)$  and  $f$  are equal up to a multiplicative constant; more generally, for two different normalization operators  $\mathcal{N}_1$  and  $\mathcal{N}_2$ , and

any function  $f$ , the normalized versions  $F_1 = \mathcal{N}_1(f)$  and  $F_2 = \mathcal{N}_2(f)$  are equal up to a multiplicative constant. In particular,  $F = \mathcal{N}(\bar{F})$ , and thus in the ABP system (3.2), the weights  $F_\tau(\xi_S(X_\tau))$  are not necessarily equal to  $\exp(-A_\tau(\xi_S(X_\tau)))$  like in (1.2) from the introduction; however it is important to have a fixed normalization operator, since by the second condition in Assumption 3.2 the ratio between these quantities remains bounded from below and from above by positive constants.

We conclude this section with additional examples of normalization operators.

- Let  $q \in [1, \infty)$ , and define

$$n_q(f) = \left( \int_{\mathbb{M}_m} f(z)^q \pi(dz) \right)^{\frac{1}{q}}.$$

In the case  $q = 1$ , we recover the example introduced above:  $\mathcal{N}_1(f) = \frac{f}{n_1(f)} = \bar{f}$ .

- Let  $z_0 \in \mathbb{M}_m$ , then define

$$n^{z_0}(f) = f(z_0) = \int_{\mathbb{M}_m} f(z) \delta_{z_0}(dz).$$

- Let also

$$n_{\min}(f) = \min_{z \in \mathbb{M}_m} f(z), \quad n_{\max}(f) = \max_{z \in \mathbb{M}_m} f(z).$$

For these examples, the relaxation of the continuous differentiability condition in Assumption 3.2 is essential: continuously differentiable approximations are given by

$$n_{\min}(f) = \lim_{q \rightarrow +\infty} \frac{1}{n_q(1/f)}, \quad n_{\max}(f) = \lim_{q \rightarrow +\infty} n_q(f).$$

**3.3. Well-posedness.** This section is devoted to the analysis of the well-posedness of the ABP system (3.2). First, Lemma 3.1 below, is stated and proved. Second, this result is combined with a Picard iteration scheme to establish global well-posedness of the self-interacting diffusion process (3.2), under stronger global Lipschitz continuity conditions for the drift coefficient. Finally, a localization argument implies global well-posedness under the assumptions on  $V$  stated in Section 2.

LEMMA 3.1. *Let  $m = \frac{\min(\min h_0, m(K))}{\gamma_n \max(\max h_0, M^{(0)}(K))}$  and  $M^{(k)} = \frac{\max(\max h_0, m^{(k)}(K)) \gamma_n}{\max(\min h_0, m(K))}$ , for  $k \in \{0, 1, \dots\}$ , where  $h_0 = \mathcal{K}(\bar{\mu}_0)$ ,  $m(K)$ ,  $M^{(k)}(K)$  are given by Assumption 3.1, and  $\gamma_n$  is given by Assumption 3.2*

*Let  $\tau \mapsto x_\tau \in \mathcal{S}$  and  $\tau \mapsto F_\tau \in \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$  be continuous mappings, such that  $n(F_\tau) = 1$  for all  $\tau \in \mathbb{R}^+$ . Define*

$$\mu_t = \bar{\mu}_0 + \int_0^t F_\tau(\xi_S(x_\tau)) \delta_{x_\tau} d\tau, \quad h_t = \mathcal{K}(\mu_t), \quad H_t = \mathcal{N}(h_t).$$

*Then, for all  $t \in \mathbb{R}^+$ ,  $k \in \mathbb{N}$ ,  $z \in \mathbb{M}_m$ ,*

$$m \leq H_t(z) \leq M^{(0)}, \quad |\partial^k H_t(z)| \leq M^{(k)}.$$

Observe that the parameters  $m$  and  $M^{(k)}$ , for  $k \in \{0, 1, \dots\}$  only depend on the algorithmic objects (the kernel function  $K$  and the normalization operator  $n$ ) introduced



in Section 3. On the contrary, they do not depend on the assumptions on the model from Section 2.

*Proof.* We only prove the estimates on  $\min H_t$  and  $\max H_t$ , since the proof of the estimates on the derivatives is similar. Note that

$$h_t(z) = h_0(z) + \int_0^t K(z, \xi_{\mathcal{S}}(x_\tau)) F_\tau(\xi(x_\tau)) d\tau,$$

where  $h_0 = \mathcal{K}(\bar{\mu}_0)$ , resp.  $K$ , are positive and continuous on  $\mathbb{M}_m$ , resp.  $\mathbb{M}_m \times \mathbb{M}_m$ . Thus for all  $t \in \mathbb{R}^+$

$$\begin{aligned} \min_{z \in \mathbb{M}_m} h_t(z) &\geq \min(\min h_0, m(K)) \left(1 + \int_0^t F_\tau(\xi_{\mathcal{S}}(x_\tau)) d\tau\right) \\ \max_{z \in \mathbb{M}_m} h_t(z) &\leq \max(\max h_0, M^{(0)}(K)) \left(1 + \int_0^t F_\tau(\xi_{\mathcal{S}}(x_\tau)) d\tau\right). \end{aligned}$$

Then the claim follows since  $H_t = \frac{h_t}{n(h_t)}$ , and using the second condition in Assumption 3.2.  $\square$

Define sets of functions  $\mathcal{F}$  and  $\mathcal{A}$  as follows:

$$\begin{cases} \mathcal{F} = \{F \in \mathcal{C}^\infty(\mathbb{M}_m); \min F \geq m > 0, \max |\partial^k F| \leq M^{(k)}, k \geq 0\}, \\ \mathcal{A} = \{A = -\log(\bar{F}); F \in \mathcal{F}\}. \end{cases} \quad (3.3)$$

Note that Lemma 3.1 may be combined with Property 2.1.

We are now in a position to state the main result of this section.

**THEOREM 3.1** (Well-posedness of (3.2)). *Grant assumptions of Section 2 concerning the model, and assumptions of Section 3 concerning the algorithm.*

- There exists a unique continuous process  $t \in [0, \infty) \mapsto (X_t, \bar{\mu}_t, F_t, A_t)$ , with values in  $\mathcal{S} \times \mathcal{P}(\mathcal{S}) \times \mathcal{C}^0(\mathbb{M}_m, (0, \infty))^2$ , which is solution of the ABP system (3.2).
- For all  $k \geq 1$ ,  $\sup_{t \geq 0} \mathbb{E}|X_t|^k < +\infty$ .
- For all  $t \in \mathbb{R}^+$ ,  $F_t \in \mathcal{F}$  and  $A_t \in \mathcal{A}$ , almost surely, where  $\mathcal{F}$  and  $\mathcal{A}$  are given by (3.3).

We provide a sketch of proof of Theorem 3.1. In the arguments presented below, we emphasize the key role played by Lemma 3.1 combined with Property 2.1.

*Proof.* Let  $T \in (0, \infty)$ , and define the mapping  $\Psi^T$  as follows. For all  $(X, F) \in L^2(\Omega, \mathcal{C}([0, T], \mathcal{S})) \times L^2(\Omega, \mathcal{C}([0, T], \mathcal{C}^1(\mathbb{M}_m, (0, \infty))))$ , set  $\Psi^T(X, F) = (Z, H)$  with

$$\begin{aligned} Z_t &= x + \sqrt{2}W_t + \int_0^t \mathcal{D}(V, A_\tau)(X_\tau) d\tau, \quad A_\tau = -\log(\bar{F}_\tau), \\ \mu_t &= \bar{\mu}_0 + \int_0^t F_\tau(\xi_{\mathcal{S}}(X_\tau)) \delta_{X_\tau} d\tau, \quad H_t = \mathcal{K}(\mu_t), \end{aligned}$$

where the mapping  $\mathcal{K}$  defined by (3.1) is extended to positive measures. Thanks to Lemma 3.1, the process  $H$  takes values in  $\mathcal{F}$ . Thus any fixed point  $(X, F)$  of the mapping  $\Psi^T$  satisfies  $F_t \in \mathcal{F}$  for all  $t \geq 0$ , and in the sequel we may assume that  $F \in L^2(\Omega, \mathcal{C}([0, T], \mathcal{F} \cap \mathcal{C}^1(\mathbb{M}_m, (0, \infty))))$ .

First, assume that  $V$  has a bounded second order derivative: then  $\nabla V$  is globally Lipschitz continuous. More precisely,  $\mathcal{D}(V, A)$  is globally Lipschitz continuous, uniformly with respect to  $A \in \mathcal{A}$ :

$$\sup_{A \in \mathcal{A}} \sup_{x_1 \neq x_2} \frac{|\mathcal{D}(V, A)(x_2) - \mathcal{D}(V, A)(x_1)|}{|x_2 - x_1|} < \infty.$$

We claim that there exists  $C \in (0, \infty)$  such that for all  $T \in (0, \infty)$ , for all  $(X^1, F^1)$  and  $(X^2, F^2)$ , such that  $F_t^1, F_t^2 \in \mathcal{F}$  for all  $t \geq 0$ , then

$$\left( \mathbb{E} \sup_{0 \leq t \leq T} |Z_t^2 - Z_t^1|^2 \right)^{\frac{1}{2}} \leq CT \left( \left( \mathbb{E} \sup_{0 \leq t \leq T} |X_t^2 - X_t^1|^2 \right)^{\frac{1}{2}} + \mathbb{E} \sup_{0 \leq t \leq T} \|\partial A_t^2 - \partial A_t^1\|_{\infty}^2 \right)^{\frac{1}{2}}$$

where  $A_t^2 = -\log(\overline{F}_t^2)$ ,  $A_t^1 = -\log(\overline{F}_t^1)$ . The structure of the mapping  $A \mapsto \mathcal{D}(V, A)$  for each example of diffusion processes is exploited to obtain this estimate.

Since  $F_t^1 \in \mathcal{F}$  and  $F_t^2 \in \mathcal{F}$ , note that there exists  $C' \in (0, \infty)$  such that for all  $t \geq 0$ ,

$$\|\partial A_t^2 - \partial A_t^1\| \leq C' (\|F_t^2 - F_t^1\|_{\infty} + \|\partial F_t^2 - \partial F_t^1\|_{\infty}).$$

Moreover, let  $h_t^1 = \mathcal{K}(\mu_t^1)$  and  $h_t^2 = \mathcal{K}(\mu_t^2)$ . Then

$$\|h_t^2 - h_t^1\|_{\infty} \leq M^{(0)}(K)T \|F_t^2 - F_t^1\|_{\infty} + M^{(0)}M^{(1)}(K)T \sup_{s \in [0, t]} \|X_s^2 - X_s^1\|,$$

more generally, for all  $k \in \{0, 1, \dots\}$ ,

$$\|\partial^k h_t^2 - \partial^k h_t^1\|_{\infty} \leq M^{(k)}(K)T \|F_t^2 - F_t^1\|_{\infty} + M^{(0)}M^{(k+1)}(K)T \sup_{s \in [0, t]} \|X_s^2 - X_s^1\|.$$

From the proof of Lemma 3.1 and thanks to Assumption 3.2,

$$\min(\mathfrak{n}(h_t^1), \mathfrak{n}(h_t^2)) \geq \gamma_{\mathfrak{n}}^{-1} \min(\min h_0, m(K)).$$

Then, writing

$$H_t^2 - H_t^1 = \frac{h_t^2 - h_t^1}{\mathfrak{n}(h_t^2)} + h_t^1 \frac{\mathfrak{n}(h_t^1) - \mathfrak{n}(h_t^2)}{\mathfrak{n}(h_t^1)\mathfrak{n}(h_t^2)},$$

and thanks to Assumption 3.2,

$$\sup_{0 \leq t \leq T} \|H_t^2 - H_t^1\|_{\infty} + \sup_{0 \leq t \leq T} \|\partial H_t^2 - \partial H_t^1\|_{\infty} \leq CT \left( \sup_{0 \leq t \leq T} \|F_t^2 - F_t^1\|_{\infty} + \sup_{0 \leq t \leq T} \|X_t^2 - X_t^1\| \right).$$

Note that the parameter  $C \in (0, \infty)$  does not depend on the time  $T$ . If  $CT < 1$ ,  $\Psi^T$  is a contraction mapping, and thus admits a unique fixed point, which yields a unique local solution for the ABP system (3.2).

In fact, a proof that the solution is in fact global, with no restriction on  $T$ , can be obtained by introducing a family of equivalent metrics  $d_{\alpha}$  on  $L^2(\Omega, \mathcal{C}([0, T], \mathcal{S})) \times L^2(\Omega, \mathcal{C}([0, T], \mathcal{C}^1(\mathbb{M}_m, (0, \infty))))$ :

$$\begin{aligned} d_{\alpha, T}((X^1, F^1), (X^2, F^2)) &= \left\| \sup_{0 \leq t \leq T} e^{-\alpha t} \|X_t^2 - X_t^1\| \right\|_{L^2(\Omega)} \\ &+ \left\| \sup_{0 \leq t \leq T} e^{-\alpha t} \|F_t^2 - F_t^1\| \right\|_{L^2(\Omega)} + \left\| \sup_{0 \leq t \leq T} e^{-\alpha t} \|\partial F_t^2 - \partial F_t^1\| \right\|_{L^2(\Omega)}. \end{aligned}$$

For any fixed  $T$ , for large enough  $\alpha$ , the mapping  $\Psi^T$  is a contraction when the distance  $d_{\alpha,T}$  is used. The computations are left to the reader.

This argument concludes the treatment of the simpler case where  $\nabla V$  is globally Lipschitz continuous (and in particular the case where the state space is compact).

The general case, when the state space is not compact, may be treated by a localization procedure. Precisely, this consists in replacing the drift coefficient  $\mathcal{D}(V,A)$  with  $\mathcal{D}_R(V,A)$ , where  $R \in (0, \infty)$ , such that  $\mathcal{D}_R(V,A)$  is globally Lipschitz continuous and coincides with  $\mathcal{D}(V,A)$  on a ball  $\mathcal{B}(0,R)$  of radius  $R$ . Let  $(X_t^R, F_t^R)_{t \geq 0}$  denote the unique solution of the system 3.2 where  $\mathcal{D}(V,A)$  is replaced with  $\mathcal{D}_R(V,A)$ . This solution is global.

In each of the examples treated in this article (see Section 2.2), the modified coefficients are constructed with applying a truncation operator to  $\nabla V$  only. The result of Lemma 3.1 is not modified by this procedure.

It remains to consider exit times  $\tau_R = \inf \{t; X_t^R \notin \mathcal{B}(0,R)\}$ , and to prove that, for any  $T \in (0, \infty)$ ,  $\lim_{R \rightarrow \infty} \mathbb{P}(\tau_R < T) \rightarrow 0$ . This result is proved thanks to moment estimates of the type

$$\sup_{R \in (0, \infty)} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_t^R|^2 \right] < \infty.$$

Such estimates are consequences of the assumptions on the potential energy function  $V$ , see Assumption 2.1 and 2.2. Details are left to the reader (see also the proof of Lemma 8.2).

Note also that for  $R \leq R'$ , then  $(X_t^R, F_t^R) = (X_t^{R'}, F_t^{R'})$  for  $t \leq \tau_R$ . Thanks to this property and the result above, it is straightforward to check that passing to the limit  $R \rightarrow \infty$  provides the unique solution of (3.2), on arbitrary  $T \in (0, \infty)$ .

This concludes the sketch of proof of Theorem 3.1.  $\square$

#### 4. Convergence results

This section contains the main results of this article, concerning the asymptotic behavior, when  $t \rightarrow \infty$ , of the solution of the ABP system (3.2). We first study consistency, then the efficiency, of the approach. The most important result dealing with consistency is Theorem 4.1: it states almost sure convergence of averages  $\bar{\mu}_t(\varphi)$  to  $\mu_\star(\varphi)$  (where  $\mu_\star = \mu_\star^0$ , see (2.3)).

Section 4.1.3 is devoted to an interpretation of the ABP system (3.2) as an Adaptive Biasing Force method, and to the interpretation of the consistency results presented here in this context.

In the remainder of this section, all the assumptions from Section 2, on the model, and of Section 3, on the algorithm, are considered to be satisfied. In particular, Theorem 3.1 ensures that the ABP system (3.2) is well defined. Moreover, the state space  $\mathcal{S}$  is finite dimensional.

##### 4.1. Consistency of ABP.

**4.1.1. Convergence of weighted empirical averages.** The main result of this article concerns the consistency of the approach, for estimating averages  $\mu_\star(\varphi)$  using weighted empirical averages  $\bar{\mu}_t(\varphi)$  (defined by (3.2)).

**THEOREM 4.1.** *Let  $\varphi \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R})$  be a bounded function, with bounded derivatives of any order. Then, almost surely,*

$$\bar{\mu}_t(\varphi) \xrightarrow[t \rightarrow \infty]{} \mu_\star(\varphi).$$

This result is a generalization in the adaptive case of (2.7). The proof of Theorem 4.1 requires the introduction of auxiliary tools, and is provided in Section 5. Several straightforward consequences of Theorem 4.1 are stated and proved in the next sections.

**4.1.2. Consequences of Theorem 4.1.**

**COROLLARY 4.1.** *We have the almost sure convergence*

$$\bar{\mu}_t \xrightarrow[t \rightarrow \infty]{} \mu_\star.$$

The notation  $\implies$  for convergence of probability distributions is introduced in Section 2.3.

*Proof.* We first state an auxiliary result: for every  $\varphi : \mathcal{S} \rightarrow \mathbb{R}$ , bounded and Lipschitz continuous function, almost surely

$$\bar{\mu}_t(\varphi) \xrightarrow[t \rightarrow \infty]{} \mu_\star(\varphi), \text{ almost surely.}$$

Indeed, apply Theorem 4.1 for an approximating sequence  $\varphi_\epsilon = \rho_\epsilon \star \varphi$ , defined by convolution with smooth functions  $\rho_\epsilon(\cdot) = \frac{1}{\epsilon} \rho_1(\frac{\cdot}{\epsilon})$ , where  $\rho$  is of class  $\mathcal{C}^\infty$ , with compact support, and  $\int_{\mathcal{S}} \rho d\lambda = 1$ .

Let  $\text{BL}(\mathcal{S}, \mathbb{R}) = \{\varphi : \mathcal{S} \rightarrow \mathbb{R} ; \varphi \text{ bounded and Lipschitz continuous}\}$ . Then there exists a sequence of functions  $(\varphi_n)_{n \geq 0}$  defined from  $\mathcal{S}$  to  $\mathbb{R}$ , bounded and Lipschitz continuous, such that

$$\bar{\mu}_t \xrightarrow[t \rightarrow \infty]{} \mu_\star \iff d(\bar{\mu}_t, \mu_\star) \xrightarrow[t \rightarrow \infty]{} 0,$$

where

$$d(\mu^1, \mu^2) = \sum_{n=0}^{\infty} \frac{1}{2^n} \min(1, |\int_{\mathcal{S}} \varphi_n d\mu^1 - \int_{\mathcal{S}} \varphi_n d\mu^2|).$$

Thanks to the convergence result above, almost surely, for every  $n \geq 0$ ,  $\bar{\mu}_t(\varphi_n) \xrightarrow[t \rightarrow \infty]{} \mu_\star(\varphi)$ , and thus  $d(\bar{\mu}_t, \mu_\star) \xrightarrow[t \rightarrow \infty]{} 0$  almost surely.

This concludes the proof of Corollary 4.1. □

The following result deals with the almost sure convergence of the functions  $\bar{F}_t$  and  $A_t$ . Note that contrary to Theorem 4.1 and Corollary 4.1, the limits  $\bar{F}_\infty$  and  $A_\infty$  depend on the parameters of the algorithm, precisely on the kernel function  $K$ . Note that these almost sure limits are not random.

The convergence of  $A_t$  to  $A_\infty$ , which is close to the free energy function  $A_\star$  for well-chosen kernel functions, is one of the nice features of the ABP method, in particular when one is interested in computing free energy differences.

**COROLLARY 4.2.** *Define, for all  $z \in \mathbb{M}_m$ ,*

$$\begin{cases} \bar{F}_\infty(z) = \mu_\star(K(z, \cdot)), \\ A_\infty(z) = -\log(\bar{F}_\infty(z)). \end{cases}$$

*Then, almost surely, for every  $\ell \in \{0, 1, \dots\}$ , uniformly on  $\mathbb{M}_m$ ,*

$$\begin{cases} \partial^\ell \bar{F}_t \xrightarrow[t \rightarrow \infty]{} \partial^\ell \bar{F}_\infty, \\ \partial^\ell A_t \xrightarrow[t \rightarrow \infty]{} \partial^\ell A_\infty. \end{cases}$$

*Proof.* The result is a consequence of the regularity properties of the kernel mapping  $K$ , of Ascoli's theorem, and of Theorem 4.1.

Let  $\mathcal{K}: \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{C}^\infty(\mathbb{M}_m)$  be the mapping defined by (3.1).

Let  $(z_n)_{n \in \mathbb{N}}$  denote a dense sequence in  $\mathbb{M}_m$ , and define, for all  $\mu^1, \mu^2 \in \mathcal{P}(\mathcal{S})$ ,

$$d_\infty(\mu^1, \mu^2) = \sum_{\ell, n=0}^{\infty} \frac{1}{2^{\ell+n}} \min(1, \left| \int_{\mathcal{S}} \partial_z^\ell K(z_n, \xi_{\mathcal{S}}(\cdot)) d\mu^1 - \int_{\mathcal{S}} \partial_z^\ell K(z_n, \xi_{\mathcal{S}}(\cdot)) d\mu^2 \right|).$$

Then for any sequence  $(\mu^k)_{k \in \mathbb{N}}$  and any  $\mu$  in  $\mathcal{P}(\mathcal{S})$ ,

- if  $\mu^k \xrightarrow[k \rightarrow \infty]{} \mu$ , then  $d_\infty(\mu^k, \mu) \xrightarrow[k \rightarrow \infty]{} 0$ ;
- if  $d_\infty(\mu^k, \mu) \xrightarrow[k \rightarrow \infty]{} 0$ , then for every  $\ell \in \{0, 1, \dots\}$ ,

$$\partial^\ell \mathcal{K}(\mu^k) \xrightarrow[k \rightarrow \infty]{} \partial^\ell \mathcal{K}(\mu),$$

uniformly on  $\mathbb{M}_m$ , thanks to Ascoli's theorem and the bound  $\|\partial_z^{k+1} K\|_\infty \leq M^{(k+1)}(K)$ .

Thanks to Theorem 4.1, it is straightforward to conclude that almost surely

$$d_\infty(\bar{\mu}_t, \mu_\star) \xrightarrow[t \rightarrow \infty]{} 0.$$

These arguments yield the convergence of  $\bar{F}_t$ . The convergence of  $A_t = -\log(\bar{F}_t)$  is then obtained thanks to the almost sure lower bound from Theorem 3.1,

$$\min_{\mathbb{M}_m} \bar{F}_t \geq m > 0. \quad \square$$

**4.1.3. A remark concerning convergence of the gradient of  $A_t$ .** To keep notation simple, consider the framework of Section 1.1: the diffusion process is the Brownian dynamics on  $\mathbb{T}^d$ , and  $\xi(x_1, \dots, x_d) = x_1 \in \mathbb{T}$ , *i.e.*  $m=1$ . Assume in addition that the kernel  $K$  is symmetric,  $K(z, \zeta) = K(\zeta, z)$ .

The main observation in this section is that, for the ABP method, one may write the derivative  $\partial_{x_1} A_t(x_1)$  of  $A_t$  as a conditional expectation, up to introducing an additional variable.

This observation is motivated by the following statement: the free energy function  $A_\star$  satisfies the identity (expression of the equilibrium mean force)

$$A'_\star(x_1) = \int_{\mathbb{T}^{d-1}} (\partial_{x_1} V(x)) e^{-(V(x) - A_\star(x_1))} dx_2 \dots dx_d = \mathbb{E}_{X \sim \mu_\star} [\partial_{x_1} V(X) | X_1 = x_1],$$

where in the conditional expectation the random variable  $X$  is distributed according to  $\mu_\star$ . This identity is the starting point for constructions of Adaptive Biasing Force (ABF) methods mentioned in Section 1.

Such a formula does not hold for  $A_t$ , when  $t < \infty$ . However, the following generalization may be used. On the one hand, for all  $t \geq 0$ ,

$$A'_t(z) = - \frac{\int_{\mathbb{T}^{d-1}} \partial_z K(z, x_1) \bar{\mu}_t(dx)}{\int_{\mathbb{T}^{d-1}} K(z, x_1) \bar{\mu}_t(dx)} = \mathbb{E}_{(X, Z) \sim \eta_t} \left[ - \frac{\partial_z K(Z, \xi(X))}{K(Z, \xi(X))} \Big| Z = z \right],$$

where  $\eta_t(dx, dz) = K(z, \xi(x)) \bar{\mu}_t(dx) dz$  is a probability distribution on  $\mathbb{T}^d \times \mathbb{T}$ , which depends on the kernel function  $K$ . Observe that if  $(X, Z) \sim \eta_t$ , in general  $Z \neq \xi(X)$ , hence

the need of the new notation instead of conditional expectations. On the other hand, the expression above for the equilibrium mean force can be rewritten in the similar form

$$A'_\star(z) = \mathbb{E}_{(X,Z) \sim \eta_\star} [\partial_{x_1} V(X) | Z = z],$$

where  $\eta_\star(dx, dz) = 1_{z=x_1} \mu_\star(dx) dz$ . Note that if  $(X, Z) \sim \eta_\star$ , then the equality  $Z = \xi(X)$  is now satisfied.

Let us now check that these expressions are consistent with Corollary 4.2. Letting  $t \rightarrow \infty$ , thanks to Corollary 4.1, it is straightforward to check that  $\eta_t$  converges almost surely to  $\eta_\infty(dx, dz) = K(z, x_1) \mu_\star(dx) dz$ . We thus obtain different expressions of  $A'_\infty(z)$ :

$$\begin{aligned} A'_\infty(z) &= \mathbb{E}_{(X,Z) \sim \eta_t} \left[ -\frac{\partial_z K(Z, \xi(X))}{K(Z, \xi(X))} \Big| Z = z \right] \\ &= -\frac{\int_{\mathbb{T}^{d-1}} \partial_z K(z, x_1) \mu_\star(dx)}{\int_{\mathbb{T}^{d-1}} K(z, x_1) \mu_\star(dx)} \\ &= \frac{\int_{\mathbb{T}^{d-1}} \partial_{x_1} V(x) K(z, x_1) \mu_\star(dx)}{\int_{\mathbb{T}^{d-1}} K(z, x_1) \mu_\star(dx)} \\ &= \mathbb{E}_{(X,Z) \sim \eta_\infty} [\partial_{x_1} V(X) | Z = z], \end{aligned}$$

thanks to the use of an integration by parts formula. Due to the presence of the kernel function  $K$ ,  $\eta_\star \neq \eta_\infty$ , and thus  $A'_\star(z) \neq A'_\infty(z)$ .

The observation above may be the starting point for other types of Adaptive Biasing methods, based on a single realization of the stochastic process and a self-interaction mechanism using an empirical distribution.

**4.2. Applications to the diffusion processes of Section 2.2.** The aim of this section is to specify, for each of the examples of diffusion processes from Section 2.2:

- the convergence result of Theorem 4.1, for well chosen test functions  $\varphi$ ;
- the expression of the limit  $\bar{F}_\infty = e^{-A_\infty}$ , in terms of the kernel  $K$  and of the free energy function  $A_\star$ .

We introduce the probability distribution  $\mu_\star^{\text{ref}}(dx) = \frac{e^{-V(x)}}{\int_{E_d} e^{-V(y)} dy} dx$  on  $E_d$ . Observe that in all the examples  $\mu_\star^{\text{ref}}$  is the marginal of the distribution  $\mu_\star$  with respect to its  $E_d$ -valued component (the equality  $\mu_\star^{\text{ref}} = \mu_\star$  holds true only in the Brownian case). As a consequence, the practitioner may choose one of the three dynamics (Brownian, Langevin or extended dynamics) of Section 2.2 to estimate averages  $\mu_\star^{\text{ref}}(\varphi)$ .

We also denote by  $A_\star^{\text{ref}}$  the free energy function associated with the reaction coordinate  $\xi$  and the probability distribution  $\mu_\star^{\text{ref}}$ : by definition,  $e^{-A_\star^{\text{ref}}}$  is the Radon-Nikodym derivative of the image of  $\mu_\star^{\text{ref}}$  by  $\xi$ , with respect to the probability distribution  $\pi$  on  $\mathbb{M}_m$ .

Assume that the kernel  $K = K_\delta$  depends on  $\delta > 0$ , and is such that the probability distribution  $K_\delta(z, \zeta) \pi(dz) \pi(d\zeta)$  converges when  $\delta \rightarrow 0$ , to  $\delta_z(d\zeta) \pi(dz)$ . Then, when  $\delta \rightarrow 0$  (and also  $\epsilon \rightarrow 0$ , in the extended dynamics case), the expressions below prove that  $A_\infty$  is an approximation of  $A_\star^{\text{ref}}$ . We do not provide quantitative estimates.

**4.2.1. Brownian dynamics (Section 2.2.1).**

- Computation of averages: for every  $\varphi \in \mathcal{C}^\infty(E_d, \mathbb{R})$ , bounded and with bounded derivatives, almost surely

$$\int \varphi d\mu_\star^{\text{ref}} = \lim_{t \rightarrow \infty} \frac{1 + \int_0^t F_\tau(\xi(X_\tau)) \varphi(X_\tau) d\tau}{1 + \int_0^t F_\tau(\xi(X_\tau)) d\tau}.$$

- Free energy function:

$$e^{-A_\infty(\cdot)} = \overline{F}_\infty(\cdot) = \int_{E_d} K_\delta(\cdot, \xi(x)) \mu_\star^{\text{ref}}(dx) = \int_{\mathbb{M}_m} K_\delta(\cdot, \zeta) e^{-A_\star^{\text{ref}}(\zeta)} \pi(d\zeta).$$

In particular, Theorem 1.1, stated in Section 1.1 and taken from [3], is a consequence of Corollaries 4.1 and 4.2, in the simplified context.

**4.2.2. Langevin dynamics (Section 2.2.2).** We use the notation  $X_t = (q_t, p_t)$ .

- Computation of averages: for every  $\varphi \in \mathcal{C}^\infty(E_d, \mathbb{R})$ , bounded and with bounded derivatives, almost surely

$$\int \varphi d\mu_\star^{\text{ref}} = \lim_{t \rightarrow \infty} \frac{1 + \int_0^t F_\tau(\xi(q_\tau)) \varphi(q_\tau) d\tau}{1 + \int_0^t F_\tau(\xi(q_\tau)) d\tau}.$$

- Free energy function:

$$e^{-A_\infty(\cdot)} = \overline{F}_\infty(\cdot) = \int_{E_d} K_\delta(\cdot, \xi(q)) \mu_\star^{\text{ref}}(dq) = \int_{\mathbb{M}_m} K_\delta(\cdot, \zeta) e^{-A_\star^{\text{ref}}(\zeta)} \pi(d\zeta).$$

Observe that the free energy function  $A_\infty$  is the same for the Brownian and the Langevin dynamics. This identity is in fact obtained since  $\xi_S(q, p) = \xi(q)$  only depends on  $q \in E_d$ .

**4.2.3. Extended dynamics (Section 2.2.3).** We use the notation  $(X_t, Z_t)$ . Recall that  $\xi_S(x, z) = z$  in this case.

- Computation of averages: for every  $\varphi \in \mathcal{C}^\infty(E_d, \mathbb{R})$ , bounded and with bounded derivatives, almost surely

$$\int \varphi d\mu_\star^{\text{ref}} = \lim_{t \rightarrow \infty} \frac{1 + \int_0^t F_\tau(Z_\tau) \varphi(X_\tau) d\tau}{1 + \int_0^t F_\tau(Z_\tau) d\tau}.$$

- Free energy function:

$$\begin{aligned} e^{-A_\infty(\cdot)} = \overline{F}_\infty(\cdot) &= \int_{E_d \times \mathbb{M}_m} K(\cdot, z) \mu_\star(dx dz) \\ &= \int_{E_d \times \mathbb{M}_m} K(\cdot, z) K_\epsilon^{\text{ext}}(z, \xi(x)) \mu_\star^{\text{ref}}(dx) \pi(dz) \\ &= \int_{\mathbb{M}_m} \left( \int_{\mathbb{M}_m} K(\cdot, z) K_\epsilon^{\text{ext}}(z, \zeta) \pi(dz) \right) e^{-A_\star^{\text{ref}}(\zeta)} \pi(d\zeta), \end{aligned}$$

where we have introduced the auxiliary kernel  $K_\epsilon^{\text{ext}}: \mathbb{M}_m \times \mathbb{M}_m \rightarrow (0, \infty)$ , such that  $\mu_\star(dx dz) = K_\epsilon^{\text{ext}}(z, \xi(x)) \mu_\star^{\text{ref}}(dx) \pi(dz)$ : up to a multiplicative constant,  $K_\epsilon^{\text{ext}}(z, \zeta) = \exp(-\frac{1}{2\epsilon} |z - \zeta|^2)$ . Note that the expression of  $A_\infty$  is not the same as in the previous examples, due to the additional term in the definition of the extended potential energy function on  $E_d \times \mathbb{M}_m$ . However, when  $\epsilon \rightarrow 0$ ,  $A_\infty$  converges to  $A_\star^{\text{ref}}$ : this observation justifies the use of the extended dynamics in the context of free energy computations.

**4.3. Efficiency.** We now state and prove a series of results concerning the efficiency of the approach, first in a qualitative way, second with a more quantitative approach. Corollary 4.3 deals with the convergence of the non-weighted empirical distribution  $\bar{\rho}_t$ , defined by (4.1); it is a straightforward consequence of Corollary 4.1.

Proposition 4.1 deals with the mean-square error, and identifies an asymptotic variance. Since the proof of Proposition 4.1 requires tools introduced in Section 5, we postpone its proof to Section 6.

In terms of the behavior of the occupation measure and of the asymptotic variance, the results stated below may be interpreted as follows: in the asymptotic regime  $t \rightarrow \infty$ , the Adaptive Biasing Potential method (3.2) performs in the same way as the non-adaptive Biasing Potential method (2.2), with the bias  $A = A_\infty$ .

Note that these results are asymptotic, when  $t \rightarrow \infty$ ; it would also be interesting to study more quantitatively the convergence, for each of the results. This question is left for future works.

**4.3.1. Convergence of non-weighted empirical distributions.** In this section, we focus on the convergence of non-weighted empirical averages  $\bar{\rho}_t(\varphi)$ , where  $\bar{\rho}_t$  is the probability distribution on  $\mathcal{S}$  defined by

$$\bar{\rho}_t = \frac{\bar{\mu}_0 + \int_0^t \delta_{X_\tau} d\tau}{1+t}. \tag{4.1}$$

We refer to  $\bar{\rho}_t$  as the non-weighted empirical distribution, or as the occupation measure, associated with the diffusion process  $(X_t)_{t \geq 0}$  defined by (3.2). We have the following result.

**COROLLARY 4.3.** *Let  $\varphi \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R})$  be a bounded function, with bounded derivatives of any order. Then*

$$\bar{\rho}_t(\varphi) \xrightarrow[t \rightarrow +\infty]{} \mu_\star^{A_\infty}(\varphi), \text{ almost surely,} \tag{4.2}$$

where  $A_\infty = \lim_{t \rightarrow \infty} A_t$  (see Corollary 4.2), and  $\mu_\star^{A_\infty}$  is given by (2.3).

Moreover, almost surely,  $\bar{\rho}_t \xrightarrow[t \rightarrow \infty]{} \mu_\star^{A_\infty}$ .

The arguments below justify that Corollary 4.3 can be interpreted, qualitatively, as an efficiency property of the ABP method.

First, observe that considering the biased dynamics  $(X_t^A)_{t \geq 0}$  given by (2.2), and setting

$$\bar{\rho}_t^A = \frac{\bar{\mu}_0 + \int_0^t \delta_{X_\tau^A} d\tau}{1+t},$$

then almost surely  $\bar{\rho}_t^A(\varphi) \xrightarrow[t \rightarrow \infty]{} \mu_\star^A(\varphi)$ . The limit in (4.2), when the adaptive dynamics is used, is the same as when using the non-adaptive dynamics (2.2), with  $A = A_\infty$ .

Second, observe that the image by the mapping  $\xi_{\mathcal{S}}: \mathcal{S} \rightarrow \mathbb{M}_m$  of the probability distribution  $\mu_\star^A$  has density with respect to  $\pi$  proportional to

$$\exp(-A_\star + A).$$

This density is constant, equal to 1, when  $A = A_\star$ : this means that in the asymptotic limit  $t \rightarrow \infty$ , the values of  $\xi_{\mathcal{S}}(X_t^{A_\star})$  are distributed according to the reference probability distribution  $\pi$ . On the contrary, when  $A = 0$ , the values of  $\xi_{\mathcal{S}}(X_t^0)$  are distributed according to  $\pi_\star^0 = e^{-A_\star} d\pi$ .

Assume that  $\pi$  is the uniform distribution on  $\mathbb{M}_m = \mathbb{T}^m$ ; assume also that all the metastability of the system is encoded by the reaction coordinate  $\xi$ . If  $A_\star$  has several



local minima, then  $\pi_\star^0$  is a multimodal distribution, and the diffusion process  $(X_t^0)_{t \geq 0}$  is metastable, and does not efficiently sample all the state space. Thus the convergence of  $\bar{\rho}_t^0$  to  $\mu_\star^0$  is expected to be slower than the convergence of  $\bar{\rho}_t^{A_\star}$  to  $\mu_\star^{A_\star}$ . Indeed, the exploration of the metastable states tends to be uniform, when  $t \rightarrow \infty$ , when observed through the reaction coordinate mapping.

Since  $A_\infty$  is an approximation of the free energy function  $A_\star$ , for well-chosen kernel functions  $K$ , efficiency of the ABP method is justified by the observations above.

We now provide the proof of Corollary 4.3, with elementary arguments. The proof of the almost sure convergence of the probability distributions is obtained as in the proof of Corollary 4.1, therefore we only focus on the convergence of averages  $\bar{\rho}_t(\varphi)$ .

*Proof. (Proof of Corollary 4.3.)* Introduce the auxiliary measure

$$\rho_t = \frac{\bar{\mu}_0 + \int_0^t \delta_{X_\tau} d\tau}{1 + \int_0^t F_\tau \circ \xi_S(X_\tau) d\tau} = \frac{(1+t)\bar{\rho}_t}{1 + \int_0^t F_\tau \circ \xi_S(X_\tau) d\tau}. \quad (4.3)$$

Since the measures  $\rho_t$  and  $\bar{\rho}_t$  only differ by a multiplicative (normalization) constant, one has the identity  $\bar{\rho}_t = \frac{\rho_t}{\rho_t(1)}$ . Then, note that

$$\begin{aligned} \rho_t(\varphi) &= \frac{\bar{\mu}_0(\varphi) + \int_0^t F_\tau \circ \xi_S(X_\tau) \frac{\varphi(X_\tau)}{F_\tau \circ \xi_S(X_\tau)} d\tau}{1 + \int_0^t F_\tau \circ \xi_S(X_\tau) d\tau} \\ &= \frac{\bar{\mu}_0(\varphi) + \int_0^t F_\tau \circ \xi_S(X_\tau) \frac{\varphi(X_\tau)}{F_\infty \circ \xi_S(X_\tau)} d\tau}{1 + \int_0^t F_\tau \circ \xi_S(X_\tau) d\tau} \\ &\quad + \frac{1}{1 + \int_0^t F_\tau \circ \xi_S(X_\tau) d\tau} \int_0^t F_\tau \circ \xi_S(X_\tau) \varphi(X_\tau) \left( \frac{1}{F_\tau \circ \xi_S(X_\tau)} - \frac{1}{F_\infty \circ \xi_S(X_\tau)} \right) d\tau \\ &= \bar{\mu}_t \left( \frac{\varphi}{F_\infty \circ \xi_S} \right) + o(1), \end{aligned}$$

using the following version of Cesaro's Lemma: if  $a: [0, \infty) \rightarrow \mathbb{R}$  is a continuous function such that  $a(t) \xrightarrow{t \rightarrow \infty} 0$ , then  $\frac{1}{t} \int_0^t a(\tau) d\tau \xrightarrow{t \rightarrow \infty} 0$ . This result may be applied, thanks to the almost sure lower bound  $1 + \int_0^t F_\tau \circ \xi_S(X_\tau) d\tau \geq 1 + mt$ ; moreover thanks to Corollary 4.2,  $F_t = \mathcal{N}(\bar{F}_t) \xrightarrow{t \rightarrow +\infty} F_\infty = \mathcal{N}(\bar{F}_\infty)$ , uniformly on  $\mathbb{M}_m$ , almost surely.

Moreover, the function  $\frac{\varphi}{F_\infty \circ \xi_S}$  is bounded and of class  $\mathcal{C}^\infty$ , with bounded derivatives (using  $\min F_\infty \geq m > 0$  thanks to Theorem 3.1). Applying Theorem 4.1, almost surely

$$\begin{aligned} \bar{\rho}_t(\varphi) &= \frac{\rho_t(\varphi)}{\rho_t(1)} \xrightarrow{t \rightarrow +\infty} \frac{\mu_\star \left( \varphi / F_\infty(\xi_S(\cdot)) \right)}{\mu_\star \left( 1 / F_\infty(\xi_S(\cdot)) \right)} = \frac{\mu_\star \left( \varphi / \bar{F}_\infty(\xi_S(\cdot)) \right)}{\mu_\star \left( 1 / \bar{F}_\infty(\xi_S(\cdot)) \right)} \\ &= \frac{\int_{\mathcal{S}} \varphi(x) \exp(-(\mathcal{E}(V)(x) - A_\infty(\xi_S(x)))) \lambda(dx)}{\int_{\mathcal{S}} \exp(-(\mathcal{E}(V)(x) - A_\infty(\xi_S(x)))) \lambda(dx)} \\ &= \frac{\int_{\mathcal{S}} \varphi(x) \exp(-(\mathcal{E}(V, A_\infty)(x))) \lambda(dx)}{\int_{\mathcal{S}} \exp(-(\mathcal{E}(V, A_\infty)(x))) \lambda(dx)} \\ &= \mu_\star^{A_\infty}(\varphi), \end{aligned}$$

thanks to the identity (2.4), and to (2.3). This concludes the proof.  $\square$

**4.3.2. Asymptotic mean-square error.** This section is devoted to a more quantitative approach, concerning the behavior when  $t \rightarrow \infty$  of the mean-square error

$$\mathbb{E}|\bar{\mu}_t(\varphi) - \mu_*(\varphi)|^2,$$

for functions  $\varphi \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R})$ , bounded and with bounded derivatives.

In order to compare the performance of the adaptive and non-adaptive versions of the biasing potential approach, introduce the following quantity

$$\mathbf{V}_\infty(\varphi, A) = \limsup_{t \rightarrow \infty} t \mathbb{E}|\bar{\mu}_t^A(\varphi) - \mu_*(\varphi)|^2 \in [0, \infty],$$

where  $A: \mathbb{M}_m \rightarrow \mathbb{R}$  is fixed,  $\bar{\mu}_t^A(\varphi)$  is the estimator of  $\mu_*(\varphi)$  defined by the left-hand side of (2.7), for every  $t \geq 0$ , using the biased dynamics (2.2).

In Section 6, it will be proved that in fact

$$\mathbf{V}_\infty(\varphi, A) = \lim_{t \rightarrow \infty} t \mathbb{E}|\bar{\mu}_t^A(\varphi) - \mu_*(\varphi)|^2 \in (0, \infty)$$

is a non-degenerate limit.

The following result, concerning the asymptotic mean-square error of the estimator  $\bar{\mu}_t(\varphi)$  of  $\mu_*(\varphi)$ , constructed using the adaptively biased dynamics (3.2).

**PROPOSITION 4.1.** *Let  $\varphi \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R})$  be a bounded function, with bounded derivatives of any order. Then*

$$t \mathbb{E}|\bar{\mu}_t(\varphi) - \mu_*(\varphi)|^2 \xrightarrow{t \rightarrow \infty} \mathbf{V}_\infty(\varphi, A_\infty),$$

where  $A_\infty = \lim_{t \rightarrow \infty} A_t$  almost surely, see Corollary 4.2.

As already explained, the asymptotic mean-square error for the adaptive version is the same as that for the non-adaptive version, where the bias is chosen as  $A = A_\infty$ . Note that the dependence of  $\mathbf{V}_\infty(\varphi, A)$  with respect to  $A$  depends a lot on the choice of the function  $\varphi$ ; therefore no optimality result is stated.

The proof of Proposition 4.1 is postponed to Section 6; explicit expressions for  $\mathbf{V}_\infty(\varphi, A)$ , in terms of the solutions of Poisson equations, are given there.

### 5. Proof of Theorem 4.1

The aim of this section is to provide a detailed proof of Theorem 4.1.

First, in Sections 5.1.1 and 5.1.2, we present the strategy, and in particular we establish a connexion with the analysis of self-interacting diffusions from [4], and more generally of stochastic algorithms, see [2, 8, 21, 31]. More precisely, Section 5.1.1 presents a (random) change of time variable,  $s = \theta(t)$ , which transforms the weighted empirical distributions  $\bar{\mu}_t$  associated with the process  $X_t$ , into non-weighted empirical distributions  $\bar{\nu}_s$  associated with a process  $Y_s$ , with modified dynamics. In Section 5.1.2, we explain how the so-called ODE method can be exploited: the asymptotic behavior of  $\bar{\nu}_s$ , when  $s \rightarrow \infty$ , is related to the behavior of a differential equation of the type  $\dot{\nu} = -\nu + \Pi(\nu)$ . A crucial result, Proposition 5.1, states that  $\Pi(\nu) = \mu_*$  is a constant mapping, and the dynamics of the differential equation above is extremely simple.

The analysis is thus expected to be much simpler than in [4]. Indeed, in Section 5.2, we directly prove the almost sure convergence of  $\bar{\mu}_t(\varphi) - \mu_*(\varphi)$  to 0 when  $t \rightarrow \infty$ . Results concerning Poisson equations are stated, their proofs being postponed to Section 8.1.

Even if it is not explicitly used in the technical part of the proof of Theorem 4.1, the description of the change of time variable strategy is included for pedagogical purpose.

Moreover, in our opinion, it is an elegant way to justify the consistency of the approach. Moreover, it may be a useful strategy in other similar situations. Readers only interested in the proof of Theorem 4.1 may skip Sections 5.1.1 and 5.1.2 – except for Proposition 5.1 which is used in the sequel.

**5.1. Approach from a stochastic approximation perspective.**

**5.1.1. Change of time variable.** In this section, we introduce a random change of time variable, and describe some of its nice properties. This is only a mathematical tool, and does not need to be performed in practice when implementing the method. In addition, as explained above, this change of variable has only a pedagogical role, and will not be used in the technical details of the proof.

Consider the solution of the ABP system (3.2). Then the mapping  $t \mapsto \bar{\mu}_t \in \mathcal{P}(\mathcal{S})$  is the unique solution of the following ordinary differential equation (ODE)

$$\frac{d\bar{\mu}_t}{dt} = \frac{\theta'(t)}{1+\theta(t)} (\delta_{X_t} - \bar{\mu}_t), \quad \theta(t) = \int_0^t F_\tau(\xi_{\mathcal{S}}(X_\tau)) d\tau. \tag{5.1}$$

The ODE (5.1) is interpreted in the following weak sense: for every bounded continuous test function  $\varphi: \mathcal{S} \rightarrow \mathbb{R}$ , the real-valued function  $t \mapsto \bar{\mu}_t(\varphi) = \int_{\mathcal{S}} \varphi d\bar{\mu}_t \in \mathbb{R}$  is the unique solution of the differential equation

$$\frac{d\bar{\mu}_t(\varphi)}{dt} = \frac{\theta'(t)}{1+\theta(t)} (\varphi(X_t) - \bar{\mu}_t(\varphi)),$$

with the initial condition  $\bar{\mu}_0(\varphi)$ .

Define the measure  $\mu_t = \bar{\mu}_0 + \int_0^t F_\tau(\xi_{\mathcal{S}}(X_\tau)) \delta_{X_\tau} d\tau$ . Then observe that  $\theta(t) = \mu_t(1)$  can be interpreted as a normalizing constant.

The presence of the random variable  $\theta(t)$  in the ODE (5.1) suggests that the analysis will be not trivial. However, we can remove this quantity thanks to a change of time variable. Simultaneously, this procedure removes the weights in the definition of the measure  $\bar{\mu}_t$ , and the dynamics of the stochastic process  $X_t$  is modified.

Thanks to Theorem 3.1, there exist two non-random real numbers  $0 < m \leq M$  such that almost surely  $\theta'(t) = F_t(\xi_{\mathcal{S}}(X_t)) \in [m, M]$  for all  $t \geq 0$ . Moreover,  $\theta(0) = 0$ , and  $\theta(t) \geq mt \xrightarrow{t \rightarrow \infty} \infty$ . As a consequence, almost surely,  $\theta: [0, \infty) \rightarrow [0, \infty)$  is a  $\mathcal{C}^1$ -diffeomorphism, with inverse denoted by  $\theta^{-1}$ . Define, for every  $s \geq 0$ ,  $\tilde{W}(s) = \int_0^{\theta^{-1}(s)} \sqrt{\theta(t)} dW(t)$ . Note that for every  $s \geq 0$ ,  $\theta^{-1}(s) = \inf \{t \geq 0; \theta(t) \geq s\}$  is a bounded stopping time, associated with the filtration generated by the Wiener process  $W$ . Then, it is straightforward to check that  $(\tilde{W}(s))_{s \geq 0}$  is a standard Wiener process on  $\mathcal{S}$ .

We introduce the following system:

$$\begin{cases} dY_s = \mathcal{D}(V, B_s)(Y_s) \frac{1}{G_s(\xi_{\mathcal{S}}(Y_s))} ds + \sqrt{\frac{2}{G_s(\xi_{\mathcal{S}}(Y_s))}} \Sigma d\tilde{W}_s, \\ \bar{\nu}_s = \frac{1}{1+s} (\bar{\mu}_0 + \int_0^s \delta_{Y_\sigma} d\sigma), \\ G_s = \mathcal{N}(\mathcal{K}(\bar{\nu}_s)), \\ B_s = -\log(\bar{G}_s). \end{cases} \tag{5.2}$$

Then the following identities are satisfied almost surely:

$$\begin{cases} X_t = Y_{\theta(t)}, \quad \bar{\mu}_t = \bar{\nu}_{\theta(t)}, \quad F_t = G_{\theta(t)}, \quad A_t = B_{\theta(t)}, \quad \forall t \geq 0 \\ Y_s = X_{\theta^{-1}(s)}, \quad \bar{\nu}_s = \bar{\mu}_{\theta^{-1}(s)}, \quad G_s = F_{\theta^{-1}(s)}, \quad B_s = A_{\theta^{-1}(s)}, \quad \forall s \geq 0. \end{cases} \tag{5.3}$$

The system (5.2) may thus be considered as the time-changed version of the original ABP system (3.2), with the new time variable  $s = \theta(t)$ , and the new unknowns  $Y_s, \bar{\nu}_s, G_s$  and  $B_s$ , replacing  $X_t, \bar{\mu}_t, F_t$  and  $A_t$ .

Observe that, in (5.2), the weight  $F_t(\xi_S(X_t)) = G_s(\xi_S(Y_s))$  does not appear anymore in the definition of the measure  $\bar{\nu}_s$ . Instead, the weight appears in the dynamics of the diffusion process  $(Y_s)_{s \geq 0}$ . In terms of new variables, the ODE (5.1) has a simpler formulation:

$$\frac{d\bar{\nu}_s}{ds} = \frac{1}{1+s} (\delta_{Y_s} - \bar{\nu}_s). \tag{5.4}$$

We are interested in the convergence of  $\bar{\mu}_t$  (or  $\bar{\mu}_t(\varphi)$ ) when  $t \rightarrow \infty$ . Since  $\bar{\mu}_t = \bar{\nu}_{\theta(t)}$ , and  $\theta(t) \xrightarrow{t \rightarrow \infty} \infty$  almost surely, the asymptotic behavior ( $s \rightarrow \infty$ ) of  $\bar{\nu}_s$  needs to be analyzed. In the remainder of this section, we work only with the system (5.2), and consider  $s$  as the natural (but fictive in practice) time variable. Observe that proving Theorem 4.1 is equivalent to proving that

$$\bar{\nu}_s(\varphi) \xrightarrow{s \rightarrow \infty} \mu_\star(\varphi), \text{ almost surely,}$$

which is done in Section 5.1.2 using the ODE method.

**5.1.2. Consistency via the ODE method.** Thanks to the change of time variable  $s = \theta(t)$  introduced above, the structure of the system (5.2) is closer to the formulation of self-interacting diffusions (see [4] for instance), depending on the normalized occupation measure, than for the initial system (3.2). However, in the specific situation considered in the present article, arguments need to be modified, in particular the coupling of the evolutions of the diffusion process and of the empirical distributions does not have the same structure (here it depends on the kernel  $K$ ).

Thanks to the ODE (5.4), observe that there is an asymptotic time scale separation (in the limit  $s \rightarrow \infty$ ) between slow variables  $\bar{\nu}_s, G_s$  and  $B_s$ , and fast variables  $Y_s$ . It is reasonable to focus on the asymptotic behavior of the diffusion process when the other variables are frozen; when its unique invariant distribution (in general depending on the frozen variables) is introduced in place of the Dirac mass in (5.4), a limit ODE is obtained: the rationale behind the ODE method is that its asymptotic behavior provides information on the asymptotic behavior of the solution of (5.4).

The ODE method allows us to make rigorous the discussion above, and to identify the appropriate limit ODE. In this article, one of the main specific properties is that the invariant distribution of the fast equation with frozen variables is equal to  $\mu_\star$ , the target probability distribution, and thus does not depend on the frozen variables.

**REMARK 5.1.** The asymptotic time scale separation (when  $t \rightarrow \infty$ ) between slow variables  $\bar{\mu}_t, F_t$  and  $A_t$ , and the fast variable  $X_t$ , already appears in the original system (1.2). The change of time variable  $s = \theta(t)$  allows us to remove the random quantity  $\theta(t)$ , and to identify the correct limit equation for the application of the ODE method.

Precisely, for every  $G \in \mathcal{C}^\infty(\mathbb{M}_m, \mathbb{R}) \cap \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$ , let  $(Y_s^G)_{s \geq 0}$  denote the diffusion process which is the unique solution of

$$dY_s^G = \frac{\mathcal{D}(V, B)(Y_s^G)}{G(\xi_S(Y_s^G))} ds + \sqrt{\frac{2}{G(\xi_S(Y_s^G))}} \Sigma d\tilde{W}_s, \tag{5.5}$$

where  $B = -\log(\bar{G})$ .

PROPOSITION 5.1. *For every  $G \in \mathcal{C}^\infty(\mathbb{M}_m, \mathbb{R}) \cap \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$ , the unique invariant probability distribution for (5.5) is equal to  $\mu_\star$ .*

*Proof.* First note that  $G = \frac{\bar{G}}{n(\bar{G})} = \frac{\exp(-B)}{n(\bar{G})}$  is equal to  $\exp(-B)$  up to a multiplicative constant, and thus a probability distribution  $\mu$  is invariant for (5.5) if and only if it is invariant for

$$d\mathcal{Y}_s^B = \frac{\mathcal{D}(V, B)(\mathcal{Y}_s^B)}{e^{-B(\xi_s(\mathcal{Y}_s^B))}} ds + \sqrt{\frac{2}{e^{-B(\xi_s(\mathcal{Y}_s^B))}}} \Sigma d\tilde{W}_s. \quad (5.6)$$

Let  $\mathcal{L}_\mathcal{Y}^B$  denote the associated infinitesimal generator: then for every function  $\varphi \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R})$ ,

$$\mathcal{L}_\mathcal{Y}^B \varphi(y) = \frac{1}{e^{-B(\xi_s(y))}} \mathcal{L}_X^B \varphi(y), \quad (5.7)$$

where  $\mathcal{L}_X^B$  is the infinitesimal generator of the biased diffusion process  $X^B$  defined by (2.2), with  $A=B$ .

Since the unique invariant probability distribution of (2.2) with  $A=B$  is  $\mu_\star^B$ , the unique invariant probability distribution of (5.5) is proportional to

$$\begin{aligned} e^{-B(\xi_s(y))} \mu_\star^B(dy) &= e^{-B(\xi_s(y))} \frac{\exp(-\mathcal{E}(V, B)(y))}{Z^B} \lambda(dy) \\ &= \frac{\exp(-\mathcal{E}(V)(y))}{Z^B} \lambda(dy) = \frac{Z^0}{Z^B} \mu_\star(dy), \end{aligned}$$

using (2.3) (expression of  $\mu_\star^B$ ) and (2.4) (expression of  $\mathcal{E}(V, B)$ ). Identifying the normalizing constants then concludes the proof of Proposition 5.1.  $\square$

Following the ODE method leads to study the following equation:

$$\frac{d\gamma_s}{ds} = \frac{1}{1+s} (\Pi(\gamma_s) - \gamma_s) = \frac{1}{1+s} (\mu_\star - \gamma_s). \quad (5.8)$$

Indeed, thanks to Proposition 5.1,  $\Pi(\gamma) = \mu_\star$  is the unique invariant distribution of (5.5), where  $G = \mathcal{K}(\gamma)$ . This property justifies the consistency of the approach, *i.e.* the almost sure convergence of  $\bar{\nu}_s$  to  $\mu_\star$ . Indeed, it is straightforward to check that, for any initial condition  $\gamma_0 \in \mathcal{P}(\mathcal{S})$ , one has

$$\gamma_s = \frac{1}{1+s} (\bar{\gamma}_0 + s\mu_\star) \xrightarrow{s \rightarrow \infty} \mu_\star.$$

Moreover, a rigorous connexion between the asymptotic behaviors of  $\bar{\nu}_s$  and of  $\gamma_s$  may be stated for instance using the notion of asymptotic pseudo-trajectories (see [2, 4]); or by proving direct estimates on the  $L^p$  norm of the random variable  $\bar{\nu}_s(\varphi) - \mu_\star(\varphi)$ .

In Section 5.2 below, instead, we prove directly estimates on the  $L^p$  norm of the random variable  $\bar{\mu}_t(\varphi) - \mu_\star(\varphi)$ ; indeed, thanks to Proposition 5.1, the situation is rather simple and the error is analyzed using straightforward computations, combined with a powerful auxiliary tool: the use of the solutions of associated Poisson equation.

**5.2. Analysis of the error and convergence.**

**5.2.1. The error in terms of the solutions of Poisson equations.** In order to prove that

$$\bar{\mu}_t(\varphi) - \mu_\star(\varphi) = \frac{\int_0^t F_\tau(\xi_S(X_\tau)) [\varphi(X_\tau) - \mu_\star(\varphi)] d\tau}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau}$$

converges to 0 when  $t \rightarrow \infty$ , it is convenient to introduce a family of Poisson equations depending on the integrand on the numerator. Let  $\Phi : (s, y) \in [1, \infty) \times \mathcal{S} \mapsto \Phi(s, y) \in \mathbb{R}$  be a  $\mathcal{C}^{1,2}$  function, *i.e.* of class  $\mathcal{C}^1$  with respect to the variable  $s$  and of class  $\mathcal{C}^2$  with respect to the variable  $y$ , with bounded associated derivatives. The application of Itô's formula yields the equality

$$\begin{aligned} \Phi(t, X_t) - \Phi(0, X_0) &= \int_0^t \mathcal{L}_X^{A_\tau}(\tau, X_\tau) d\tau + \int_0^t \frac{\partial \Phi}{\partial \tau}(\tau, X_\tau) d\tau \\ &\quad + \int_0^t \sqrt{2} \langle \nabla \Phi(\tau, X_\tau), \Sigma dW(\tau) \rangle, \end{aligned}$$

where  $\mathcal{L}_X^A$  is the infinitesimal generator of the biased diffusion process  $X^A$ , see (2.2).

Assume that the function  $\Phi$  satisfies, for all  $t \geq 0, x \in \mathcal{S}$

$$\mathcal{L}_X^{A_t} \Phi(t, X_t) = F_t(\xi_S(x)) [\varphi(x) - \mu_\star(\varphi)]; \tag{5.9}$$

then one obtains

$$\begin{aligned} \bar{\mu}_t(\varphi) - \mu_\star(\varphi) &= \frac{\Phi(t, X_t) - \Phi(0, X_0)}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau} - \frac{\int_0^t \sqrt{2} \langle \nabla \Phi(\tau, X_\tau), \Sigma dW(\tau) \rangle}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau} \\ &\quad - \frac{\int_0^t \frac{\partial \Phi}{\partial \tau}(\tau, X_\tau) d\tau}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau}. \end{aligned}$$

Recall that,  $1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau \geq mt$ , for all  $t \geq 0$ , almost surely, thanks to Theorem 3.1, with  $m > 0$ . Convergence of  $\bar{\mu}_t(\varphi) - \mu_\star(\varphi)$  to 0, in a  $L^p$  sense, then follows from appropriate estimates on the function  $\Phi$  and its derivatives, which are stated below.

**5.2.2. Properties of solutions of Poisson equations.** This section is devoted to the statement of the properties concerning solutions of Poisson equations which are used in the analysis. We emphasize that the estimates are uniform with respect to  $A \in \mathcal{A}$ , which is defined by (3.3).

The Equation (5.9) can be written as

$$\Phi(t, x) = \frac{1}{n(F_t)} \Psi(A_t, x) \tag{5.10}$$

where, for any  $A \in \mathcal{A}$  (see (3.3)),  $\Psi(A, \cdot)$  is solution of the Poisson equation

$$\begin{cases} \mathcal{L}_X^A \Psi(A, \cdot) = e^{-A(\xi_S(x))} [\varphi - \mu_\star(\varphi)], \\ \int \Psi(A, \cdot) d\mu_\star^A = 0. \end{cases} \tag{5.11}$$

Introduce the set

$$\mathcal{C} = \mathcal{C}_{\text{pol}}^\infty(\mathcal{S}, \mathbb{R}) = \left\{ \varphi \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R}) ; \forall k \in \mathbb{N}, \exists p_k \in \mathbb{N}, \sup_{x \in \mathcal{S}} \frac{|D^k \varphi(x)|}{1 + |x|^{p_k}} < \infty \right\} \quad (5.12)$$

of functions  $\varphi : \mathcal{S} \rightarrow \mathbb{R}$ , of class  $\mathcal{C}^\infty$ , with at most polynomial growth, and all derivatives with at most polynomial growth. Note that the average  $\mu_\star(\varphi)$  is well-defined, since the probability distribution  $\mu_\star$  admits finite moments of any order.

We first state the following well-posedness result.

**PROPOSITION 5.2.** *For every  $A \in \mathcal{A}$  and every  $\varphi \in \mathcal{C}$ , there exists a unique solution  $\Psi(A, \cdot) \in \mathcal{C}$  of the Poisson Equation (5.11).*

We only provide a sketch of the proof. Define the auxiliary function  $\bar{\varphi}^A = e^{-A \circ \xi \mathcal{S}}(\varphi - \mu_\star(\varphi))$ , and note that  $\bar{\varphi}^A \in \mathcal{C}$ . Thanks to (2.6), the centering condition

$$\int \bar{\varphi}^A d\mu_\star^A = \int [\varphi - \mu_\star(\varphi)] d\mu_\star^0 = 0$$

is satisfied. It is then well-known that the unique solution of the Poisson Equation (5.11) is given by

$$\Psi(A, x) = - \int_0^\infty \mathbb{E}_x[\bar{\varphi}^A(X_t^A)] dt, \quad (5.13)$$

where  $X^A$  is the biased process given by (2.2).

In Section 5.2.3 below, bounds on  $\Psi(A_t, \cdot)$ , and its derivatives are required. The analysis is performed using the two following claims. On the one hand, thanks to Proposition 2.1, almost surely  $A_t \in \mathcal{A}$ , where  $\mathcal{A}$  is defined by (3.3). On the other hand, Proposition 5.3 states estimates which are uniform for  $A \in \mathcal{A}$ . The proof is postponed to Appendix 8.1.

**PROPOSITION 5.3.** *Let  $\varphi \in \mathcal{C}$ , of class  $\mathcal{C}^\infty$ , with at most polynomial growth.*

*There exist  $C \in (0, \infty)$  and  $p \in \mathbb{N}^\star$ , such that the following results hold true.*

(i) *For every  $A \in \mathcal{A}$  and every  $x \in \mathcal{S}$*

$$|\Psi(A, x)| \leq C(1 + |x|^p). \quad (5.14)$$

(ii) *For every  $A \in \mathcal{A}$  and every  $x \in \mathcal{S}$*

$$|\nabla_x \Psi(A, x)| \leq C(1 + |x|^p). \quad (5.15)$$

(iii) *The function  $(t, x) \in [0, \infty) \times \mathcal{S} \mapsto \Psi(A_t, x)$  is of class  $\mathcal{C}^{1,2}$ , and for every  $x \in \mathcal{S}$  and every  $t \geq 0$ , almost surely*

$$\left| \frac{\partial \Psi(A_t, x)}{\partial t} \right| \leq \frac{C(1 + |x|^p)}{1 + t}, \quad (5.16)$$

where  $(A_t)_{t \geq 0}$  is the  $\mathcal{A}$ -valued process defined in (3.2).

**5.2.3. Proof of convergence.** An approximation procedure is required to deal with the low regularity properties of the normalization operator  $\mathfrak{n}$ , see Assumption 3.2. For  $k \in \mathbb{N}$ , define

$$\Phi^{(k)}(t, x) = \frac{1}{\mathfrak{n}^{(k)}(\bar{F}_t)} \Psi(A_t, x).$$

Then observe that that

$$\begin{aligned}
 \bar{\mu}_t(\varphi) - \mu_*(\varphi) &= \frac{\int_0^t F_\tau(\xi_S(X_\tau)) [\varphi(X_\tau) - \mu_*(\varphi)] d\tau}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau} \\
 &= \frac{\int_0^t \frac{1}{n(\bar{F}_\tau)} \mathcal{L}_X^{A_\tau} \Psi(A_\tau, \cdot)(X_\tau) d\tau}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau} \\
 &= \lim_{k \rightarrow \infty} \frac{\int_0^t \frac{1}{n^{(k)}(\bar{F}_\tau)} \mathcal{L}_X^{A_\tau} \Psi^{(k)}(A_\tau, X_\tau) d\tau}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau} \\
 &= \lim_{k \rightarrow \infty} \frac{\int_0^t \mathcal{L}_X^{A_\tau} \Phi^{(k)}(\tau, X_\tau) d\tau}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau} =: \lim_{k \rightarrow \infty} \epsilon_t^{(k)}(\varphi),
 \end{aligned}$$

where the limit  $k \rightarrow \infty$  is understood in an almost sure sense, thanks to Assumption 3.2, and the fact that  $\bar{F}_t \in \mathcal{F}$  for all  $t \geq 0$ , almost surely, thanks to Theorem 3.1.

Itô's formula can be used, since  $(t, x) \mapsto \Phi^{(k)}(t, x)$  is of class  $\mathcal{C}^{1,2}$  thanks to Proposition 5.3. Then

$$\begin{aligned}
 \epsilon_t^{(k)}(\varphi) &= \frac{\Phi^{(k)}(t, X_t) - \Phi^{(k)}(0, X_0)}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau} - \frac{\int_0^t \sqrt{2} \langle \nabla \Phi^{(k)}(\tau, X_\tau), \Sigma dW(\tau) \rangle}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau} - \frac{\int_0^t \frac{\partial \Phi^{(k)}}{\partial \tau}(\tau, X_\tau) d\tau}{1 + \int_0^t F_\tau(\xi_S(X_\tau)) d\tau} \\
 &=: \epsilon_t^{(k),1}(\varphi) + \epsilon_t^{(k),2}(\varphi) + \epsilon_t^{(k),3}(\varphi).
 \end{aligned}$$

We now prove the following result.

LEMMA 5.1. *Let  $\varphi \in \mathcal{C}$ . There exists  $C(\varphi) \in (0, \infty)$  such that for every  $t \geq 0$  and  $k \in \mathbb{N}$*

$$\mathbb{E} |\epsilon_t^{(k)}(\varphi)|^2 \leq \frac{C(\varphi)}{t}.$$

Observe that Lemma 5.1 is valid for test functions  $\varphi$  in the set  $\mathcal{C}$  defined by (5.12). To prove Theorem 4.1, *i.e.* an almost sure convergence result, we will only use it with test functions which are bounded and have bounded derivatives. However, to prove Proposition 4.1, we will need this Lemma for test functions with polynomial growth.

*Proof. (Proof of Lemma 5.1.)* The proof of that result consists in using the estimates of Proposition 5.3.

- Thanks to item (i) from Proposition 5.3, for every  $t \geq 0$ ,  $\Phi^{(k)}(t, \cdot)$  has at most polynomial growth, and moments of the process  $X$  are bounded, see Theorem 3.1. More precisely, the parameters  $C$  and  $p$  in the right-hand side of the inequality (5.14) do not depend on  $A = A_\tau$ . Moreover, thanks to Assumption 3.2, for every  $k \in \mathbb{N}$  and  $\tau \geq 0$ , one has  $n^{(k)}(\bar{F}_\tau) \geq m > 0$  almost surely. It is then straightforward to conclude that

$$\mathbb{E} |\epsilon_t^{(k),1}(\varphi)|^2 \leq \frac{C(\varphi)}{t^2}.$$

- To have an estimate of the stochastic integral, we use Itô's formula, and we obtain

$$\mathbb{E} |\epsilon_t^{(k),2}(\varphi)|^2 \leq C \frac{1 + \int_0^t |\Sigma^* \nabla_x \Phi^{(k)}(\tau, X_\tau)|^2 d\tau}{1 + t^2}$$



$$\leq \frac{C(\varphi)}{t},$$

thanks to (5.15), and arguments similar to the term above.

- Finally, using (5.16), and similar arguments, one obtains

$$\mathbb{E}|\epsilon_t^{(k),3}(\varphi)|^2 \leq \frac{C(\varphi) \left( \int_0^t \frac{1}{1+\tau} d\tau \right)^2}{t^2} \leq \frac{C(\varphi)(1+\log(t))^2}{t^2}.$$

Gathering estimates then concludes the proof of Lemma 5.1. □

We are now in a position to deduce Theorem 4.1 from Lemma 5.1. First, note that it is straightforward to obtain

$$\mathbb{E}|\bar{\mu}_t(\varphi) - \mu_\star(\varphi)|^2 \leq \frac{C(\varphi)}{t}.$$

Indeed, the right-hand side in the estimate of Lemma 5.1 does not depend on  $k$ , and taking the limit  $k \rightarrow \infty$  in the right-hand side gives the result, thanks to Assumption 3.2 which ensures the required uniform convergence properties for the application of the bounded convergence theorem.

This estimate ensures the convergence in mean-square sense, and in probability, of  $\bar{\mu}_t(\varphi)$  to  $\mu_\star(\varphi)$ . To go further, and obtain the almost sure convergence, we use the following arguments. First, note that it is sufficient to prove that  $\bar{\mu}_{\exp(t)}$  converges almost surely to  $\mu_\star(\varphi)$  when  $t \rightarrow \infty$ . Using the estimate

$$\mathbb{E}|\bar{\mu}_{\exp(t)}(\varphi) - \mu_\star(\varphi)|^2 \leq C(\varphi)e^{-t},$$

and Borel-Cantelli's Lemma, then almost surely, for every  $\delta \in \mathbb{Q} \cap (0, \infty)$ ,

$$\bar{\mu}_{\exp(n\delta)}(\varphi) \xrightarrow[n \rightarrow \infty]{} \mu_\star(\varphi).$$

Finally, thanks to the differential Equation (5.1) and boundedness of the function  $\varphi$ , the mapping  $t \mapsto \bar{\mu}_{\exp(t)}(\varphi)$  is Lipschitz continuous, with constant smaller than  $C(\varphi)$ , almost surely, for some  $C(\varphi) \in (0, \infty)$  depending only on  $\varphi$ , and on the parameters appearing in the definition of the set  $\mathcal{F}$ , see (3.3). It is then straightforward to obtain the almost sure convergence

$$\bar{\mu}_{\exp(t)}(\varphi) \xrightarrow[t \rightarrow \infty]{} \mu_\star(\varphi).$$

This concludes the proof of Theorem 4.1.

### 6. Analysis of the mean-square error. Proof of Proposition 4.1

In this section, we give a proof of Proposition 4.1, concerning the asymptotic behavior of the mean-square error, which is decomposed as

$$\mathbb{E}|\bar{\mu}_t(\varphi) - \mu_\star(\varphi)|^2 = (\mathbb{E}\bar{\mu}_t(\varphi) - \mu_\star(\varphi))^2 + \text{Var}(\bar{\mu}_t(\varphi)), \tag{6.1}$$

when  $t \rightarrow \infty$ , for functions  $\varphi \in \mathcal{C}$ , of class  $\mathcal{C}^\infty$ , with at most polynomial growth.

In Section 6.1, we prove that the bias satisfies

$$\mathbb{E}\bar{\mu}_t(\varphi) - \mu_\star(\varphi) = \mathcal{O}\left(\frac{1 + \log(t)}{t}\right). \tag{6.2}$$

In Section 6.2, we then prove that

$$t\mathbb{E}|\bar{\mu}_t(\varphi) - \mu_\star(\varphi)|^2 \xrightarrow{t \rightarrow \infty} \mathbf{V}_\infty(\varphi) \in [0, \infty). \tag{6.3}$$

In particular, thanks to (6.2), we may interpret the limit as the asymptotic variance, since

$$\mathbf{V}_\infty(\varphi) = \lim_{t \rightarrow \infty} t\text{Var}(\bar{\mu}_t(\varphi)).$$

The asymptotic variance is expressed in terms of the solution of a Poisson Equation (5.11), with  $A = A_\infty = \lim_{t \rightarrow \infty} A_t$  (defined in Corollary 4.2).

In Section 6.3, we check that  $\mathbf{V}_\infty(\varphi) = \mathbf{V}_\infty(\varphi, A_\infty)$ , where  $\mathbf{V}_\infty(\varphi, A) \in [0, \infty)$  is the asymptotic variance associated with the non-adaptively biasing method, using (2.2) and (2.7), with  $A = A_\infty$ .

**6.1. Asymptotic behavior of the bias.** Let us prove (6.2). Using the same arguments as in Section 5.2.3, note that

$$\begin{aligned} \mathbb{E}\bar{\mu}_t(\varphi) - \mu_\star(\varphi) &= \mathbb{E}\left[\lim_{k \rightarrow \infty} \epsilon_t^{(k)}(\varphi)\right] \\ &= \lim_{k \rightarrow \infty} \mathbb{E}\left[\epsilon_t^{(k)}(\varphi)\right] \\ &= \lim_{k \rightarrow \infty} \mathbb{E}\left[\epsilon_t^{(k),1}(\varphi) + \epsilon_t^{(k),3}(\varphi)\right]. \end{aligned}$$

Indeed, using Assumption 3.2 and the property that  $\bar{F}_t \in \mathcal{F}$ , for all  $t \geq 0$ , almost surely, allows us to use the bounded convergence theorem. Moreover, note that  $\mathbb{E}[\epsilon_t^{(k),2}(\varphi)] = 0$ , for all  $k \in \mathbb{N}$  and  $t \geq 0$ . Then (6.2) is an immediate corollary of Lemma 5.1.

**6.2. Asymptotic behavior of the mean-square error.** Let us now prove (6.3). Like in Sections 5.2.3 and 6.1 above, we use the decomposition of  $\bar{\mu}_t(\varphi) - \mu_\star(\varphi)$  in terms of the auxiliary function  $\Phi^{(k)}$ ; we prove error bounds which are uniform with respect to  $k \in \mathbb{N}$ , and pass to the limit  $k \rightarrow \infty$ , thanks to Assumption 3.2 and Theorem 3.1.

It is straightforward to check that, uniformly in  $k \in \mathbb{N}$ ,

$$t\mathbb{E}|\bar{\mu}_t(\varphi) - \mu_\star(\varphi)|^2 - t\mathbb{E}\left|\frac{\int_0^t \sqrt{2}\langle \nabla_x \Phi^{(k)}(\tau, X_\tau), \Sigma dW(\tau) \rangle}{1 + \int_0^t F_\tau(\xi_S(X_\tau^A)) d\tau}\right|^2 = O\left(\frac{1}{t}\right),$$

thanks to Lemma 5.1, *i.e.* only the stochastic integral contributes to the asymptotic variance. We can directly pass to the limit  $k \rightarrow \infty$  at this stage.

Let  $R(t) = \frac{1 + \int_0^t F_\tau(\xi_S(X_\tau^A)) d\tau}{t}$ . Then almost surely, thanks to Corollary 4.2 (uniform convergence of  $F_\tau$  to  $F_\infty$  when  $\tau \rightarrow \infty$ ), Corollary 4.3, and the version of Cesaro’s Lemma stated in the proof of Corollary 4.3,

$$R(t) \xrightarrow{t \rightarrow \infty} \int_S F_\infty \circ \xi_S d\mu_\star^{A_\infty} = \frac{\mu_\star^{A_\infty}(e^{-\star A_\infty \circ \xi_S})}{\mathfrak{n}(\bar{F}_\infty)} = \frac{1}{\mu_\star^0(e^{A_\infty \circ \xi_S}) \mathfrak{n}(\bar{F}_\infty)},$$

using (2.6). We then obtain

$$t\mathbb{E}\left|\frac{\int_0^t \frac{\sqrt{2}}{\mathfrak{n}(\bar{F}_\tau)} \langle \nabla_x \Psi(A_\tau, X_\tau), \Sigma dW(\tau) \rangle}{1 + \int_0^t F_\tau(\xi_S(X_\tau^A)) d\tau}\right|^2 = \mathbb{E}\frac{\left|\int_0^t \frac{\sqrt{2}}{\mathfrak{n}(\bar{F}_\tau)} \langle \nabla_x \Psi(A_\tau, X_\tau), \Sigma dW(\tau) \rangle\right|^2}{t|R(t)|^2}$$

$$\begin{aligned}
&= \mathbb{E} \frac{\left| \int_0^t \frac{\sqrt{2}}{n(\bar{F}_\tau)} \langle \nabla_x \Psi(A_\tau, X_\tau), \Sigma dW(\tau) \rangle \right|^2}{t} + o(1) \\
&= \frac{2 \int_0^t \mathbb{E} \frac{|\Sigma^* \nabla_x \Psi(A_\tau, X_\tau)|^2}{n(\bar{F}_\tau)^2} d\tau}{t} + o(1) \\
&= \frac{\frac{2}{n(\bar{F}_\infty)^2} \int_0^t \mathbb{E} |\Sigma^* \nabla_x \Psi(A_\infty, X_\tau)|^2 d\tau}{t} + o(1) \\
&= 2\mathbb{E} \bar{\mu}_t (|\Sigma^* \nabla_x \Psi(A_\infty, \cdot)|^2) + o(1) \\
&\xrightarrow{t \rightarrow \infty} 2\mu_\star (|\Sigma^* \nabla_x \Psi(A_\infty, \cdot)|^2),
\end{aligned}$$

thanks to Lemma 5.1, applied to the function  $|\Sigma^* \nabla_x \Psi(A_\infty, \cdot)|^2 \in \mathcal{C}$ , thanks to Proposition 5.2. Observe that the limit does not depend on the normalization operator  $n$ . We thus obtain (6.3), more precisely,

$$t\mathbb{E} |\bar{\mu}_t(\varphi) - \mu_\star(\varphi)|^2 \xrightarrow{t \rightarrow \infty} \mathbf{V}_\infty(\varphi) = 2\mu_\star (|\Sigma^* \nabla_x \Psi(A_\infty, \cdot)|^2).$$

**6.3. Comparison with the non-adaptive biasing method.** We now check that the expression obtained above for the asymptotic variance in the adaptive method, coincides with the expression of the asymptotic variance in the non-adaptive method, when choosing  $A = A_\infty$ .

Let  $A \in \mathcal{A}$  (see (3.3)), and  $\varphi \in \mathcal{C}$ . Using (2.7), and the solution of the Poisson Equation (5.11),

$$\begin{aligned}
\bar{\mu}_t^A(\varphi) - \mu_\star(\varphi) &= \frac{1 + \int_0^t \exp(-A \circ \xi_S(X_\tau^A)) [\varphi(X_\tau^A) - \mu_\star(\varphi)] d\tau}{1 + \int_0^t \exp(-A \circ \xi_S(X_\tau)) d\tau} \\
&= \frac{1 + \int_0^t \mathcal{L}_X^A \Psi(A, X_\tau^A) d\tau}{1 + \int_0^t \exp(-A \circ \xi_S(X_\tau)) d\tau} \\
&= \frac{\Psi(A, X_t^A) - \Psi(A, X_0^A) - \int_0^t \sqrt{2} \langle \nabla_x \Psi(A, X_\tau^A), \Sigma dW(\tau) \rangle}{1 + \int_0^t \exp(-A \circ \xi_S(X_\tau)) d\tau}
\end{aligned}$$

Since  $A \in \mathcal{A}$ ,  $\exp(-A \circ \xi_S) \geq m > 0$ , for some  $m > 0$ . Then

$$t\mathbb{E} |\bar{\mu}_t^A(\varphi) - \mu_\star(\varphi)|^2 - t\mathbb{E} \left| \frac{\int_0^t \sqrt{2} \langle \nabla_x \Psi(A, X_\tau^A), \Sigma dW(\tau) \rangle}{1 + \int_0^t \exp(-A \circ \xi_S(X_\tau^A)) d\tau} \right|^2 = O\left(\frac{1}{t}\right).$$

Let  $R^A(t) = \frac{1 + \int_0^t \exp(-A \circ \xi_S(X_\tau^A)) d\tau}{t}$ . Then almost surely  $R^A(t) \xrightarrow{t \rightarrow \infty} \mu_\star^A(e^{-A \circ \xi_S}) = \frac{1}{\mu_\star(e^{A \circ \xi_S})}$ .

With the same arguments as in Section 6.2 above,

$$\begin{aligned}
t\mathbb{E} \left| \frac{\int_0^t \sqrt{2} \langle \nabla_x \Psi(A, X_\tau^A), \Sigma dW(\tau) \rangle}{1 + \int_0^t \exp(-A \circ \xi_S(X_\tau^A)) d\tau} \right|^2 &= \mathbb{E} \frac{\left| \int_0^t \sqrt{2} \langle \nabla_x \Psi(A, X_\tau^A), \Sigma dW(\tau) \rangle \right|^2}{t |R^A(t)|^2} \\
&= \mathbb{E} \frac{\left| \int_0^t \sqrt{2} \langle \nabla_x \Psi(A, X_\tau^A), \Sigma dW(\tau) \rangle \right|^2}{t} + o(1) \\
&= \frac{2 \int_0^t \mathbb{E} |\Sigma^* \nabla_x \Psi(A, X_\tau^A)|^2 d\tau}{t} + o(1)
\end{aligned}$$

$$\begin{aligned}
 &= 2\mathbb{E}\bar{\mu}_t^A(|\Sigma^*\nabla_x\Psi(A,\cdot)|^2) + o(1) \\
 &\xrightarrow{t\rightarrow\infty} 2\mu_\star(|\Sigma^*\nabla_x\Psi(A,\cdot)|^2).
 \end{aligned}$$

We thus conclude that

$$t\mathbb{E}|\bar{\mu}_t^A(\varphi) - \mu_\star(\varphi)|^2 \xrightarrow{t\rightarrow\infty} \mathbf{V}_\infty(\varphi, A) = 2\mu_\star(|\Sigma^*\nabla_x\Psi(A,\cdot)|^2).$$

The asymptotic variance in the ABP method is thus equal to the asymptotic variance in the non-adaptive method with  $A = A_\infty = \lim_{t\rightarrow\infty} A_t$ , as expected:

$$\mathbf{V}_\infty(\varphi) = 2\mu_\star(|\Sigma^*\nabla_x\Psi(A_\infty,\cdot)|^2) = \mathbf{V}_\infty(\varphi, A_\infty).$$

**7. The SPDE case**

The aim of this section is to generalize the approach developed in other sections of this article, to deal with metastable stochastic processes in infinite dimension. More precisely, we describe an ABP method designed to compute averages  $\mu_\star(\varphi)$ , where  $\mu_\star$  is a probability distribution defined on an infinite dimensional (Hilbert) space; the corresponding diffusion processes are given by some parabolic semilinear stochastic partial differential equations (SPDEs).

In Section 7.1, we describe the model, and we explain how it fits in the framework of Section 2. In particular, this description justifies the introduction of the abstract objects in Section 2.

Some arguments and some statements need to be substantially modified, compared with the finite dimensional situation, see Sections 7.2 and 7.3, and details in Section 8.2.

**7.1. The model.** In this section, we consider infinite dimensional diffusion processes, which are solutions of parabolic, semilinear, SPDEs, driven by space-time white noise, in space dimension 1, which may be written in the following form:

$$du^0(t, x) = \frac{\partial^2 u^0(t, x)}{\partial x^2} dt - \nabla\mathcal{V}(u^0(t, x))dt + \sqrt{2}dW(t, x), \tag{7.1}$$

for  $x \in (0, 1)$ , with (for instance) homogeneous Dirichlet boundary conditions.

The function  $\mathcal{V} : \mathbb{R} \rightarrow \mathbb{R}$  is a smooth mapping. With the choice  $\mathcal{V}(x) = \frac{x^4}{4} - \frac{x^2}{2}$ , one obtains the Allen-Cahn equation, which is the paradigmatic example of metastable SPDE considered in the literature: see for instance [9–11, 22]. In particular, (non-adaptive) importance sampling techniques are considered for this problem in [42].

In this article,  $\mathcal{V}$  is assumed to have bounded derivatives, in order to simplify the presentation and the functional setting. Metastable states are solutions of the stationary PDE

$$\frac{\partial^2 u(x)}{\partial x^2} - \nabla\mathcal{V}(u(x)) = 0.$$

Assume that the potential energy function  $\mathcal{V}$  is even; then  $x \mapsto u^0(x) = 0$  is one solution. Moreover, if there exists another solution  $x \mapsto u^+(x)$ ,  $x \mapsto u^-(x) = -u^+(x)$  is also a solution. These solutions are critical points of the energy functional

$$u \mapsto \int_0^1 \left[ \frac{1}{2} \left| \frac{\partial u(x)}{\partial x} \right|^2 + \mathcal{V}(u(x)) \right] dx,$$

and may be local minima, saddle points, etc...

It is convenient and standard to write (7.1) as a stochastic evolution equation in the Hilbert space  $H = L^2(0,1)$ , see for instance the monograph [17]:

$$du_t^0 = Lu_t^0 dt - DV(u_t^0)dt + \sqrt{2}dW(t), \quad (7.2)$$

where  $D$  denotes the Fréchet derivative, and

- $(e_n)_{n \in \mathbb{N}^*}$  is the complete orthonormal system of  $H$  given by  $e_n(x) = \sqrt{2} \sin(n\pi x)$ ;
- the unbounded linear operator  $L: H \rightarrow H$  satisfies  $Lu = -\sum_{n \in \mathbb{N}^*} \pi^2 n^2 \langle u, e_n \rangle e_n$ ;
- $V(u) = \int_0^1 \langle \nabla \mathcal{V}(\theta u), u \rangle d\theta$  for all  $u \in H$ ;
- $(W(t))_{t \geq 0}$  is a cylindrical Wiener process on  $H$ , i.e.  $W(t) = \sum_{n \in \mathbb{N}^*} \beta_n(t) e_n$  for a family  $(\beta_n)_{n \in \mathbb{N}^*}$  of independent, one-dimensional, standard Wiener processes.

Equation (7.2) admits a unique mild solution (see [17]) with values in  $H$ , defined for  $t \geq 0$ , i.e.  $u^0$  is the unique process solution satisfying the equation

$$u_t^0 = e^{tL} u_0 - \int_0^t e^{(t-s)L} DV(u_s^0) ds + \sqrt{2} \int_0^t e^{(t-s)L} dW(s),$$

where  $(e^{tL})_{t \in [0, +\infty)}$  is the semi-group on  $H$  generated by

$$L: e^{tL} u = \sum_{n \in \mathbb{N}^*} e^{-\pi^2 n^2 t} \langle u, e_n \rangle e_n.$$

In the context of this section, Assumption 2.1 is satisfied when the following condition is satisfied:

$$\sup_{x \in \mathbb{R}} |\mathcal{V}''(x)| < \pi^2. \quad (7.3)$$

In other words, the Lipschitz constant of the non-linear coefficient  $u \in H \mapsto DV(u) \in H$  is bounded from above by all the eigenvalues of  $-L$ . Ergodicity of the SPDE (7.2) is then obtained by the following arguments, see for instance [16, Section 6.3] for additional details. Let  $u_0, v_0 \in H$  denote two initial conditions, and define  $(u_t^0)_{t \geq 0}$  and  $(v_t^0)_{t \geq 0}$  the solutions of (7.2) driven by the same Wiener process  $(W(t))_{t \geq 0}$ . Then  $r_t = u_t^0 - v_t^0$  satisfies

$$\frac{dr_t}{dt} = Lr_t + DV(v_t^0) - DV(u_t^0),$$

and thus

$$\begin{aligned} \frac{1}{2} \frac{\|dr_t\|_H^2}{dt} &= -\langle (-L)r_t, r_t \rangle + \langle DV(v_t^0) - DV(u_t^0), r_t \rangle \\ &\leq -\pi^2 \|r_t\|_H^2 + \sup_{x \in \mathbb{R}} |V''(x)| \|r_t\|_H^2 \leq -\gamma \|r_t\|_H^2, \end{aligned}$$

with  $\gamma > 0$ , thanks to the condition (7.3). By Grönwall's Lemma,  $\mathbb{E} \|v_t^0 - u_t^0\|_H^2 \leq e^{-\gamma t} \|v_0 - u_0\|_H^2$ , which yields uniqueness of an invariant distribution for the SPDE (7.2), as well as exponential convergence to equilibrium. There are several general ways to prove the existence of an invariant distribution, see [16, Chapter 6]. Alternatively, in the situation treated in the present work, the SPDE has a gradient structure and an explicit formula for the invariant distribution is available, see [16, Theorem 8.6.3]:

$$\mu_\star(du) = \frac{\exp(-V(u))}{Z} \lambda(du)$$

for some  $Z \in (0, \infty)$ , where the reference measure  $\lambda$  is a Gaussian measure on  $H$ , defined below in Section 7.1.2.

We are now in a position to explain how the SPDE dynamics fits into the general framework presented in this article, in Section 2.

**7.1.1. Setting.** In the SPDE example, one has the following elements, see Section 2.1.

- **State space:**  $\mathcal{S} = H$  (infinite dimensional, separable, Hilbert space).
- **Reaction coordinate:** assume that  $\mathbb{M}_m = \mathbb{T}$  (with  $m = 1$ ),  $E_1 = \mathbb{R}$  (with  $d = 1$ ). Then for instance  $\xi(u) = \xi_{\mathcal{S}}(u) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{1}{2} \int_0^1 u(x) dx\right)$ .
- **Drift coefficient:**  $\mathcal{D}(V, A) = Lu - D(V - A \circ \xi_{\mathcal{S}})$ . **Diffusion operator:**  $\Sigma$  is the identity on  $\mathcal{S}$ .

Since  $L$  is an unbounded linear operator on  $H$ , note that the drift is only defined on a domain  $D(L) \subset H$ . This is one of the technical issues which are specific to the infinite dimensional framework.

REMARK 7.1. Note that, in general, there does not exist a function  $\mathcal{V}_A : \mathbb{R} \rightarrow \mathbb{R}$  such that the function  $V - A \circ \xi : L^2(0, 1) \rightarrow \mathbb{R}$  satisfies  $D(V - A \circ \xi)(u)(x) = \nabla \mathcal{V}_A(u(x))$ : the bias is a nonlocal function of  $u$ , since it depends on the spatial average  $\int_0^1 u(y) dy$ , instead of  $u(x)$  only.

The biased version (2.2) of the SPDE (7.1) is written as

$$du^A(t) = Lu^A(t)dt - D(V - A \circ \xi_{\mathcal{S}})(u^A(t))dt + \sqrt{2}dW(t), \tag{7.4}$$

with mild formulation

$$u^A(t) = e^{tL}u_0 + \sqrt{2} \int_0^t e^{(t-s)L}dW(s) - \int_0^t e^{(t-s)L}DV(u^A(s))ds + \int_0^t e^{(t-s)L}D(A \circ \xi_{\mathcal{S}})(u^A(s))ds.$$

**7.1.2. Invariant probability distribution.** We now construct the total energy function, and the reference measure  $\lambda$  on  $H$ .

First, the definition of the mapping  $V \mapsto \mathcal{E}(V)$  is straightforward:  $\mathcal{E}(V) = V$ . The reference measure  $\lambda$  on  $\mathcal{S}$  is defined as follows: it is the centered Gaussian probability distribution on  $H$  with covariance operator  $L^{-1}$ . This measure can be constructed as follows: let  $(g_n)_{n \in \mathbb{N}^*}$  be a sequence of independent standard real-valued Gaussian random variables (centered and with variance 1); then  $\lambda$  is the probability distribution of the  $H$ -valued random variable  $\sum_{n \in \mathbb{N}^*} \frac{1}{n\pi} g_n e_n$ .

REMARK 7.2. One may check that  $\lambda$  defined as above is the distribution of the Brownian Bridge on  $(0, 1)$ . This interpretation is specific to the choice of  $L$  and plays no role in this article. On the contrary, the construction above, based on eigenvalues and eigenvectors of  $L$ , is general.

It is straightforward to check that  $\lambda$  is the unique invariant distribution of (7.2) when  $V = 0$ . More generally, for any function  $A : \mathbb{T} \rightarrow \mathbb{R}$  of class  $\mathcal{C}^\infty$ , the probability distribution  $\mu_\star^A$  on  $H$ , defined by

$$\mu_\star^A(du) = \frac{\exp(-(V(u) - A(\xi(u))))}{Z^A} \lambda(du)$$

where  $Z^A \in (0, \infty)$  thanks to (7.3), is the unique invariant distribution of the biased SPDE (7.4), see for instance [16].

**7.1.3. Free Energy function.** It remains to discuss how the free energy function  $A_\star$  is defined: this is done using Definition 2.1, like in the finite dimensional case. It is natural to choose  $\pi$  to be the Lebesgue measure on  $\mathbb{T}$ . Indeed, the image measure of the Gaussian distribution  $\lambda$  by the linear mapping  $u \mapsto \int_0^1 u(y) dy$  is a non-degenerate Gaussian distribution on  $\mathbb{R}$ ; thus the image of  $\lambda$  by  $\xi_S$  is equivalent to the Lebesgue measure on  $\mathbb{T}$ . Then  $\pi_\star^0$  the image of  $\mu_\star^0$  by  $\xi$  is equivalent to  $\pi$ .

**7.2. ABP dynamics and convergence results.** Let us first describe the dynamics of the ABP method, which generalizes (3.2) in the case where the diffusion process is governed by a SPDE:

$$\begin{cases} du(t) = Lu(t)dt - D(V - A_t \circ \xi_S)(u(t))dt + \sqrt{2}dW(t) \\ \bar{\mu}_t = \frac{\bar{\mu}_0 + \int_0^t \exp(-A_r \circ \xi_S(u(r))) \delta_{u(r)} dr}{1 + \int_0^t \exp(-A_r \circ \xi_S(u(r))) dr} \\ \exp(-A_t(z)) = \int_{\mathbb{T}^d} K(z, \xi_S(u)) \bar{\mu}_t(du), \quad \forall z \in \mathbb{T}^m, \end{cases} \tag{7.5}$$

For simplicity, we have chosen the normalization operator  $\mathcal{N}$ , with  $n(F) = \int_{\mathbb{T}} F(z) dz$ . The kernel function  $K : \mathbb{T} \times \mathbb{T} \rightarrow (0, \infty)$  satisfies Assumption 3.1.

As explained in Section 7.1 above, it is convenient to consider the mild formulation for the SPDE dynamics: the first equation in (7.5) is understood as

$$u(t) = e^{tL} u_0 - \int_0^t e^{(t-s)L} D(V - A_s \circ \xi)(u(s)) ds + \sqrt{2} \int_0^t e^{(t-s)L} dW(s).$$

Using Lemma 3.1 and standard techniques, the following generalization of Theorem 3.1 is obtained.

**THEOREM 7.1.** *Consider the framework of Section 7.1 (in particular assume that (7.3) is satisfied), and assume that the kernel function  $K$  satisfies Assumption 3.1.*

- *There exists a unique continuous process  $t \mapsto (u(t), \bar{\mu}_t, A_t)$ , taking values in  $H \times \mathcal{P}(H) \times \mathcal{C}^0(\mathbb{M}_m, (0, \infty))$ , which is solution of the ABP system (7.5).*
- *For all  $k \geq 1$ ,  $\sup_{t \geq 0} \mathbb{E} \|u(t)\|_H^k < +\infty$ .*
- *There exist  $m \in (0, \infty)$  and  $(M^{(r)})_{r \in \{0, 1, \dots\}} \in (0, \infty)$  such that, almost surely,  $A_t \in \mathcal{A}$ , for all  $t \in \mathbb{R}^+$ , where*

$$\begin{cases} \mathcal{F} = \{F \in \mathcal{C}^\infty(\mathbb{M}_m); \min F \geq m, \max |\partial^k F| \leq M^{(k)}, k \geq 0\}, \\ \mathcal{A} = \{A = -\log(\bar{F}); F \in \mathcal{F}\}. \end{cases} \tag{7.6}$$

We are able to prove generalizations of Theorem 4.1 and of Corollary 4.2.

**THEOREM 7.2.**

- *Let  $\varphi \in \mathcal{C}^\infty(H, \mathbb{R})$  be a bounded function, with bounded derivatives of any order. Then, almost surely,*

$$\bar{\mu}_t(\varphi) \xrightarrow[t \rightarrow \infty]{} \mu_\star(\varphi).$$

- *Define, for all  $z \in \mathbb{M}_m$ ,*

$$A_\infty(z) = -\log(\mu_\star(K(z, \cdot))).$$

Then, almost surely, for every  $\ell \in \{0, 1, \dots\}$ , uniformly on  $\mathbb{M}_m$ ,

$$\partial^\ell A_t \xrightarrow{t \rightarrow \infty} \partial^\ell A_\infty.$$

The efficiency results from Section 4.3 may also be generalized: more precisely, the convergence result (4.2), and Proposition 4.1 remain valid.

**7.3. Some modifications for SPDEs.** Compared with the finite dimensional situation, observe that we only state the almost sure convergence of averages  $\bar{\mu}_t(\varphi)$  in Theorem 7.2. The arguments used in the proof of Corollary 4.1 do not easily generalize to the infinite dimensional setting, to prove almost sure convergence of  $\bar{\mu}_t$ .

There are also modifications when dealing with the solutions of the Poisson equations. To simplify the discussion, assume first that  $A=0$ . Then the Poisson Equation (5.11) is written in the infinite dimensional setting, as

$$\langle Lu - DV(u), D\Psi(u) \rangle + \frac{1}{2} \sum_{n=1}^{\infty} D^2\Psi(u) \cdot (e_n, e_n) = \varphi(u) - \mu_\star(\varphi), \quad \forall u \in H, \quad (7.7)$$

where the unknown is the function  $\Psi : H \rightarrow \mathbb{R}$ . In the Poisson equation above, the first order derivative  $D\Psi(u) \in H$  is interpreted as an element of  $H$  thanks to Riesz's theorem.

For an arbitrary function  $\Psi$  of class  $\mathcal{C}^2$  on  $H$ , it is not true in general that the left-hand side is well-defined, for all  $u \in H$ , or even when  $u = u(t)$  is the diffusion process evaluated at a time  $t \geq 0$ . Indeed,  $L$  is an unbounded operator, so  $Lu$  is not an element of  $H$  in general. Moreover, the series may not be convergent.

In fact, the Poisson equation may be solved and all the terms make sense thanks to regularity properties, which may be written in the form (7.8), where auxiliary norms are introduced: for any  $\alpha \in (0, 1)$ , and any  $h \in H$ , let

$$\|h\|_\alpha^2 = \sum_{n=1}^{\infty} \lambda_n^{2\alpha} |\langle h, e_n \rangle|^2 \in [0, \infty], \quad \|h\|_{-\alpha}^2 = \sum_{n=1}^{\infty} \lambda_n^{-2\alpha} |\langle h, e_n \rangle|^2 < \infty.$$

We refer to [13, Chapters 4,5], for general results concerning the smoothing properties of the transition semi-group in infinite dimension, and to [12, Proposition 6.1], and [30, Chapter 4, Section 8], for their application to the analysis of Poisson equations. Rigorous properties are often stated for spatial Galerkin approximations, with bounds not depending on the dimension. We do not provide such details here, and directly write the results in the Hilbert space  $H$ .

Using arguments from the references mentioned above, and taking care of the dependence with respect to the function  $A$  to obtain uniform bounds on the set  $\mathcal{A}$ , generalizations of Propositions 5.2 and 5.3 are obtained.

**PROPOSITION 7.1.** *Let  $A \in \mathcal{A}$ , and  $\varphi : H \rightarrow \mathbb{R}$ , of class  $\mathcal{C}^\infty$ , bounded and with bounded derivatives of any order.*

*There exists a unique solution  $\Psi(A, \cdot)$  of the Poisson Equation (5.11),*

$$\langle Lu - D(V - A \circ \xi)(u), D\Psi(u) \rangle + \frac{1}{2} \sum_{n=1}^{\infty} D^2\Psi(u) \cdot (e_n, e_n) = e^{-A(\xi(u))} [\varphi(u) - \mu_\star(\varphi)],$$

for all  $u \in H$ , with the condition  $\int_H \Psi d\mu_\star = 0$ .

*This solution is given by*

$$\Psi(A, u) = - \int_0^\infty \mathbb{E}_u [\bar{\varphi}^A(u(t))] dt,$$



for all  $u \in H$ , where  $\bar{\varphi}^A(u) = e^{-A(\xi(u))}[\varphi(u) - \mu_\star(\varphi)]$ .

Moreover, the following properties are satisfied.

- There exists  $C(\varphi) \in (0, \infty)$  such that, for all  $A \in \mathcal{A}$  and  $u \in H$ ,

$$|\Psi(A, u)| \leq C(\varphi)(1 + \|u\|_H^2).$$

- For every  $\alpha \in (0, \frac{1}{2})$ , there exists  $C(\alpha, \varphi) \in (0, \infty)$ , such that, for all  $A \in \mathcal{A}$  and  $u \in H$

$$\begin{cases} |\langle D_u \Psi(A, u), h \rangle| \leq C(\alpha, \varphi)(1 + \|u\|_H^2) \|h\|_{-2\alpha}, \quad \forall h \in H, \\ |D_u^2 \Psi(A, u) \cdot (h, k)| \leq C(\alpha, \varphi)(1 + \|u\|_H^2) \|h\|_{-\alpha} \|k\|_{-\alpha}, \quad \forall h, k \in H. \end{cases} \quad (7.8)$$

- For every  $\alpha \in (0, \frac{1}{4})$  and every  $n \in \mathbb{N}$ , there exists  $C(\alpha, n, \varphi) \in (0, \infty)$ , such that  $\mathbb{E} \|u(t)\|_\alpha^n \leq C(\alpha, n, \varphi)(1 + \frac{\|u(0)\|_H}{t^\alpha})^n$ .
- The function  $(t, u) \in [0, \infty) \times H \mapsto \Psi(A_t, u)$  is of class  $\mathcal{C}^{1,2}$ , and for every  $u \in H$  and every  $t \geq 0$ , almost surely

$$\left| \frac{\partial \Psi(A_t, u)}{\partial t} \right| \leq \frac{C(1 + \|u\|^2)}{1+t},$$

where  $(A_t)_{t \geq 0}$  is the  $\mathcal{A}$ -valued process defined in (7.5).

A sketch of proof is postponed to Appendix 8.2. More precisely, we focus there on the estimates (7.8), with  $A=0$  (to simplify the presentation), for  $\alpha > 0$ , since they are the main novelty in the infinite dimensional framework.

The estimates (7.8) justify that all the terms in the left-hand side of (7.7) make sense. First, if one assumes that  $\|u\|_\epsilon < \infty$  for some  $\epsilon > 0$ , choosing  $2\alpha = 1 - \epsilon < 1$ , gives  $|\langle Lu, D\Psi(u) \rangle| < \infty$ . Second,  $\sum_{n=1}^\infty \|e_n\|_{-\alpha}^2 = \sum_{n=1}^\infty \lambda_n^{-2\alpha} < \infty$  for  $\alpha > \frac{1}{4}$ .

Adapting the strategy of proof of Theorem 4.1, developed in Section 5.2, and using Proposition 7.1 to control the terms, it is then straightforward to prove that

$$\mathbb{E} |\bar{\mu}_t(\varphi) - \mu_\star(\varphi)|^2 \leq \frac{C(\varphi)}{t} \xrightarrow{t \rightarrow \infty} 0.$$

The proof of the almost sure convergence result is obtained using the boundedness of  $\varphi$ , and the same argument as in the finite dimensional case. This concludes the proof of the first part of Theorem 7.2. The second part of Theorem 7.2, concerning the almost sure convergence of  $A_t$ , is proved exactly as Corollary 4.2.

Thanks to the general framework developed in Section 2, the ABP method is also applicable in the infinite dimensional setting, for metastable stochastic PDEs.

**Acknowledgments.** The authors would like to thank the referees for their valuable comments which helped us improve the presentation of this manuscript. They would also like to thank Tony Lelièvre and Gabriel Stoltz for stimulating discussions on this article and comments on a preliminary version of this manuscript. The work is partially supported by SNF grants 200020/149871 and 200021/163072.

## Appendix. Results concerning Poisson equations.

**8.1. The finite dimensional case.** The aim of this section is to give a proof of Proposition 5.3, stated in Section 5.2. More precisely, the key point is to prove that the estimates are uniform with respect to  $A \in \mathcal{A}$ , where  $\mathcal{A}$  is defined by (3.3).

To simplify the presentation, the analysis is restricted to functions  $\varphi$  which are bounded and have bounded derivatives of any order. The general case  $\varphi \in \mathcal{C}$ , having polynomial growth (see (5.12)), may be treated using weight functions, using Assumptions 2.1 and 2.2. Thanks to Property 2.1, the weight functions may be chosen independently of  $A \in \mathcal{A}$ , hence estimates are uniform for  $A \in \mathcal{A}$ .

Let  $\varphi$  be fixed, and recall the notation  $\bar{\varphi}^A = e^{-A \circ \xi_S}(\varphi - \mu_\star(\varphi))$ . Moreover,  $\Psi(A, \cdot)$  is given by (5.13).

**8.1.1. Auxiliary result: exponential convergence to equilibrium.** Let  $W: \mathcal{S} \rightarrow \mathbb{R}_+$  be defined as follows. If the dynamics is given by the Brownian dynamics (Section 2.2.1), set

$$W(x) = \|x\|.$$

If the dynamics is given by the Langevin dynamics (Section 2.2.2), set

$$W(q, p) = \sqrt{V(q) + Q(q, p)}$$

with  $Q(q, p) = \frac{\gamma^2}{4} \|q\|^2 + \frac{\gamma}{2} \langle q, p \rangle + \frac{1}{2} \|p\|^2$ . Here we assume without loss of generality that  $V \geq 0$ .

The case of the extended dynamics (Section 2.2.3) is treated like the Brownian dynamics case.

Let  $(P_t^A)_{t \geq 0}$  denote the semi-group associated with the SDE (2.2), with  $A \in \mathcal{A}$ . Recall that  $\mu_\star^A$  defined by (2.3) is the unique invariant distribution for this process.

The aim of this section is to prove that convergence to equilibrium is exponentially fast, uniformly with respect to  $A \in \mathcal{A}$ .

**PROPOSITION 8.1.** *There exists  $\vartheta > 0$  and  $C \in (0, \infty)$  such that for every measurable function  $\varphi: \mathcal{S} \rightarrow \mathbb{R}$  with  $\|\varphi\|_W := \sup \frac{|\varphi(x)|}{1+W(x)} < \infty, A \in \mathcal{A}$  and  $t \geq 0$ ,*

$$|P_t^A \varphi(x) - \mu_\star^A(\varphi)| \leq C e^{-\vartheta t} (1 + W(x)) \|\varphi\|_W, \tag{8.1}$$

*In addition, for functions  $\varphi: \mathcal{S} \rightarrow \mathbb{R}$  such that  $\|\varphi\|_W < \infty$ , almost surely*

$$\frac{1}{t} \int_0^t \varphi(X_r^A) dr \xrightarrow[t \rightarrow \infty]{} \int \varphi d\mu_\star^A. \tag{8.2}$$

We first state a version of Harris Theorem. Let  $\mathcal{E}$  be a measurable set, and  $W: \mathcal{E} \rightarrow \mathbb{R}_+$  be a measurable map. Following [26], define for every  $0 \leq \beta \leq 1$  and  $f: \mathcal{E} \rightarrow \mathbb{R}$  measurable (possibly unbounded),

$$\|f\|_{\beta, W} = \sup_{x, y} \frac{|f(x) - f(y)|}{2 + \beta(W(x) + W(y))}.$$

**LEMMA 8.1.** *Let  $P$  and  $Q$  be two Markov kernels over  $\mathcal{E}$ . Assume there exist  $0 \leq \rho < 1, \kappa \geq 0, R \geq \frac{2\kappa}{1-\rho}, \epsilon > 0, \delta \geq 0$  and  $\psi$  a probability distribution over  $\mathcal{E}$  such that*

- (1)  $PW \leq \rho W + \kappa, QW \leq \rho W + \kappa;$
- (2) *For all  $x \in W_R := \{y \in \mathcal{E} : W(y) \leq R\}$   $P(x, dy) \geq \epsilon \psi(dy)$  and  $|\delta_x P - \delta_x Q|_{1, W} \leq \delta.$*

*Then, there exist  $0 \leq \beta \leq 1$  and  $0 \leq \theta < 1$  such that*

$$\|P\|_{\beta, W} \leq \theta, \text{ and } \|Q\|_{\beta, W} \leq \theta + \delta.$$

Here  $\|P\|_{\beta,W}$  stands for  $\sup\{\|Pf\|_{\beta,W} : \|f\|_{\beta,W} \leq 1\}$ .

The first statement (concerning  $P$ ) rephrases Theorem 3.1 in [26]. The proof of the second one (concerning  $Q$ ) easily follows from the proof of the first one.

LEMMA 8.2. *For each  $A \in \mathcal{A}$  there exists  $T_A > 0, 0 \leq \beta_A \leq 1, 0 \leq \theta_A < 1$  and  $\delta_A > 0$  such that for all  $B \in \mathcal{A}$ ,*

$$\|A - B\|_{C^1} \leq \delta \Rightarrow \|P_T^B\|_{\beta,W} \leq \theta,$$

where  $\|A - B\|_{C^1} = \max|A - B| + \max|\partial A - \partial B|$ .

*Proof.* Let  $V_A(q) = V(q) - A(\xi(q))$ . Replacing  $V$  by  $V + c$  for some  $c > 0$  we can assume without loss of generality that  $V_A \geq 0$  for all  $A \in \mathcal{A}$ . Thanks to Property 2.1, there exist positive constants  $\alpha = \alpha_{\mathcal{A}}, \kappa = \kappa_{\mathcal{A}}$  such that for all  $A \in \mathcal{A}$

$$\mathcal{L}^A W_A^2 \leq -2\alpha W_A^2 + 2\alpha\kappa^2,$$

where  $\mathcal{L}^A$  is the infinitesimal generator of the SDE (2.2) and  $W_A$  is defined like  $W$  with  $V_A$  in place of  $V$ . Then, by standard Itô calculus,

$$P_t^A W_A^2 \leq e^{-2\alpha t} W_A^2 + \kappa^2. \tag{8.3}$$

Replacing  $\kappa^2$  by  $\kappa^2 + 2(|m| \wedge |M|)$  and using the fact that  $\|W^2 - W_A^2\| \leq \|A\|_{\infty} \leq |m| \wedge |M|$ , we then obtain

$$P_t^A W^2 \leq e^{-2\alpha t} W^2 + \kappa^2. \tag{8.4}$$

Also, by Hölder's inequality,

$$P_t^A W \leq \sqrt{e^{-2\alpha t} W^2 + \kappa^2} \leq e^{-\alpha t} W + \kappa. \tag{8.5}$$

By classical ellipticity (Brownian) or hypoellipticity (Langevin) results (see e.g [28]), for any given  $A \in \mathcal{A}$ :

- (a) For all  $t > 0$  there exists a smooth function  $(x, y) \mapsto p_t^A(x, y)$  such that

$$P_t^A(x, dy) = p_t^A(x, y) dy$$

- (b)  $(P_t^A)_{t \geq 0}$  is a strong Feller semi-group.

Given  $A \in \mathcal{A}, x_0 \in \mathcal{S}$  and  $t_0 > 0$  one can then find  $y_0 \in \mathcal{S}$  such that

$$p_{t_0}^A(x_0, y_0) > 0 \tag{8.6}$$

The strong Feller property combined with the existence of an invariant probability having full support (here  $\mu_{\star}^A$ ) makes  $(P_t^A)$  positively recurrent (see e.g [28], Section 5). In particular, the almost sure convergence property (8.2) is satisfied, and for all  $x \in \mathcal{S}$  and every neighborhood  $U$  of  $x$ , there exists  $\tau > 0$  such that

$$P_{\tau}^A(x, U) > 0. \tag{8.7}$$

Using (8.6) and (8.7), it is then proved that, for every compact set  $K \subset \mathcal{S}$ , there exist  $T > 0$ , a bounded open set  $V$  and  $\epsilon > 0$ , such that

$$P_T(x, dy) \geq \epsilon 1_V(y) dy, \quad \forall x \in K.$$

Applying Lemma 8.1, with  $P = P_T^A$  and  $K = W_R$ ,

$$\|P_T^A\|_{\beta, W} < \theta$$

for some  $0 \leq \beta \leq 1$  and  $0 \leq \theta < 1$ .

We now claim that  $|\delta_x P_T^A - \delta_x P_T^B|_{1, W} \rightarrow 0$  uniformly in  $x \in W_R$  when  $\|A - B\|_{C^1} \rightarrow 0$ .

Let  $f$  be such that  $\|f\|_{1, W} \leq 1$ . Replacing  $f$  by  $f - f(x_0)$ , without loss of generality it is assumed that  $|f(x)| \leq C + W(x)$ , with  $C = 2 + W(x_0)$ . By Girsanov's theorem,

$$P_T^A f(x) - P_T^B f(x) = \mathbb{E}_x(f(X_T^A) - f(X_T^B)) = \mathbb{E}_x(f(X_T^A)(1 - M_T))$$

where  $(M_t)$  is the martingale defined as

$$M_t = \exp\left(-\int_0^t \langle u_s, d\tilde{W}_s \rangle - \frac{1}{2} \int_0^t \|u_s\|^2 ds\right)$$

and  $u_s = \nabla(A \circ \xi - B \circ \xi)(X_s^B)$  (Brownian case),  $u_s = (2\gamma)^{-1/2}(\nabla(A \circ \xi_s - B \circ \xi_s))(X_s^B)$  (Langevin case). Thus, for all  $x \in W_R$

$$\begin{aligned} |P_T^A f(x) - P_T^B f(x)| &\leq \mathbb{E}_x((C + W(X_T^A))|1 - M_T|) \leq (C + \sqrt{P_T^A W^2(x)}) \sqrt{\mathbb{E}_x(M_T^2 - 1)} \\ &\leq (C + R + \kappa) \sqrt{\mathbb{E}_x(M_T^2 - 1)}, \end{aligned}$$

thanks to Hölder's inequality and to (8.4). Observe that  $M_t^2 = \tilde{M}_t e^{\int_0^t \|u_s\|^2 ds}$  where  $(\tilde{M}_t)$  is a nonnegative martingale. Therefore  $1 \leq \mathbb{E}(M_T^2) \leq e^{c_T \|A - B\|_{C^1}^2}$  with  $c_T = T \max(1, (2\gamma)^{-1}) \|D\xi\|^2$ .

This concludes the proof of the claim. The result then follows from applying Lemma 8.1.  $\square$

We are now in a position to conclude.

*Proof. (Proof of Proposition 8.1.)* By Ascoli's theorem,  $\mathcal{A}$  is relatively compact for the  $C^1$  topology. Thus, thanks to Lemma 8.2, there exists a finite covering of  $\mathcal{A}$  by open sets (for the  $C^1$  topology)  $O_1, \dots, O_N$  (i.e.  $\mathcal{A} \subset \cup_{i=1}^N O_i$ ), and parameters  $0 \leq \beta_i \leq 1, t_i > 0$ , and  $0 \leq \theta_i < 1$ , such that for all  $A \in O_i$

$$\|P_{t_i}^A\|_{\beta_i, W} \leq \theta_i.$$

Let  $\theta = \max_{i=1, \dots, N} \theta_i < 1$ . Note that for all  $\beta > 0$

$$\|\varphi\|_{1, W} \leq \|\varphi\|_{\beta, W} \leq \frac{1}{\beta} \|\varphi\|_{1, W} \leq \frac{1}{\beta} \|\varphi\|_W,$$

while, by (8.5), for all  $r \geq 0$

$$\|P_r^A \varphi\|_W \leq \|\varphi\|_W (1 + \kappa).$$

Thus, for  $A \in O_i, t = kt_i + r, k \in \mathbb{N}$  and  $0 \leq r < t_i$ ,

$$\|P_t^A \varphi\|_{1, W} \leq \|P_{t_i}^A \varphi\|_{\beta_i, W} \leq \theta^k \|P_r^A \varphi\|_{\beta_i, W} \leq \frac{\theta^k}{\beta_i} \|\varphi\|_W (1 + \kappa).$$

That is

$$\|P_t^A \varphi\|_{1,W} \leq e^{-\vartheta t} C \|\varphi\|_W$$

with  $\vartheta = \min_i \frac{-\log(\theta)}{t_i}$  and  $C = \max_i \frac{1+\kappa}{\beta_i \theta}$ . Equivalently, for all  $x, y$ ,

$$|P_t^A \varphi(x) - P_t^A \varphi(y)| \leq e^{-\vartheta t} C \|\varphi\|_W (2 + W(x) + W(y)).$$

Hence, integrating in  $y$ ,

$$|P_t^A \varphi(x) - \mu_A^*(\varphi)| \leq e^{-\vartheta t} C \|\varphi\|_W (2 + W(x) + \mu_A^* W) \leq e^{-\vartheta t} C \|\varphi\|_W (2 + W(x) + \kappa).$$

This concludes the proof.  $\square$

**8.1.2. Proof of Proposition 5.3.** In the proof below, the values of  $C \in (0, \infty)$  and  $p \in \mathbb{N}^*$  may change from line to line. Note that if  $\varphi$  is bounded, then  $\|\varphi\|_W \leq \|\varphi\|_\infty = \sup \varphi(x)$ .

The properties of  $V$  given by Assumptions 2.1 and 2.2 play a key role in the estimate. Recall that Property 2.1 then allows to get estimates which are uniform with respect to  $A \in \mathcal{A}$ . As already explained, the technical computations are not reported here.

- (i) Thanks to Property 2.1, and to Proposition 8.1, applied with  $\varphi = \bar{\varphi}^A$ , there exist  $\vartheta \in (0, \infty)$ ,  $C(\varphi) \in (0, \infty)$  and  $p \in \mathbb{N}^*$ , such that for every  $A \in \mathcal{A}$ , then for all  $x \in \mathcal{S}$  and  $t \geq 0$ , one has

$$|\mathbb{E}_x [\bar{\varphi}^A(X_t^A)]| \leq C \|\varphi\|_\infty e^{-\vartheta t} (1 + W(x)). \tag{8.8}$$

Integrating from  $t=0$  to  $t=\infty$ , using (5.13) and the polynomial growth assumption on  $V$ , gives (5.14).

- (ii) The inequality (8.8) may be rewritten as follows: for all  $A \in \mathcal{A}$ ,  $x \in \mathcal{S}$  and  $t \geq 0$ ,

$$\|P_t^A \bar{\varphi}^A\|_W \leq C \|\varphi\|_\infty e^{-\vartheta t}, \tag{8.9}$$

where we recall that  $(P_t^A)_{t \geq 0}$  is the transition semi-group associated with  $X^A$ .

The elliptic and hypoelliptic cases need to be treated separately.

Consider first the elliptic case (Brownian dynamics). The idea is to adapt the arguments in [30, Chapter 2, Section 6], and to check that all estimates are uniform with respect to  $A \in \mathcal{A}$ . First, by direct estimates of the derivatives (using in particular the semi-convexity property of  $V$ ), when  $t \in [0, 1]$ ,

$$|\nabla_x P_t^A \bar{\varphi}^A(x)| \leq C(1 + W(x)) \|\nabla_x \bar{\varphi}^A\|_\infty \leq C(1 + W(x)) (\|\varphi\|_\infty + \|\nabla_x \varphi\|_\infty).$$

Second, for  $t \geq 1$ , let  $\phi_t^A = P_{t-1}^A \bar{\varphi}^A$ . Using the semi-group property, and the Bismut-Elworthy-Li formula, with constants which do not depend on  $A \in \mathcal{A}$ ,

$$|\nabla_x P_t^A \bar{\varphi}^A(x)| = |\nabla_x P_1 \phi_t^A(x)| \leq C \|\phi_t^A\|_W (1 + W(x)).$$

Using (8.9) to have an estimate of  $\|\phi_t^A\|_W$ , then integrating separately from  $t=0$  to  $t=1$  and from  $t=1$  to  $t=\infty$  gives (5.15).

Consider now the hypoelliptic case (Langevin dynamics). The idea is to adapt the arguments in [30, Chapter 3, Section 6], and to check that all estimates are uniform with respect to  $A \in \mathcal{A}$ . Again (8.9) is a fundamental ingredient. Estimates in Sobolev norms of  $P_t^A \bar{\varphi}^A$  and of its derivatives are obtained. Then pointwise estimates are obtained using a Sobolev imbedding theorem. All the estimates are uniform with respect to  $A \in \mathcal{A}$ . The long and technical calculations are omitted.

(iii) Since  $F_t \in \mathcal{F}$  for all  $t \geq 0$ , almost surely, thanks to Theorem 3.1, then  $\min_{z \in \mathbb{M}_m} \bar{F}_t(z) \geq m$  for all  $t \geq 0$ , almost surely, for some  $m \in (0, \infty)$ .

Moreover,  $\bar{F}_t(z) = \bar{\mu}_t(K(z, \cdot))$ ; thanks to Assumption 3.1 and to the ODE (5.1), for every  $k \in \{0, 1, \dots\}$ , there exists  $C^{(k)} \in (0, \infty)$  such that

$$\sup_{z \in \mathbb{M}_m} \left| \frac{d(\partial^k A_t(z))}{dt} \right| \leq \frac{C^{(k)}}{1+t}. \tag{8.10}$$

For every  $t > 0$ , every  $\epsilon \in (-t, 1)$ , note that

$$\mathcal{L}_y^{A_{t+\epsilon}} \Psi(A_{t+\epsilon}, \cdot) - \mathcal{L}_y^{A_t} \Psi(A_t, \cdot) = 0,$$

thanks to (5.7). Passing to the limit  $\epsilon \rightarrow 0$  yields

$$\begin{aligned} \mathcal{L}_y^{A_t} \frac{\partial \Psi(A_t, \cdot)}{\partial t} &= - \left( \frac{\partial \mathcal{L}_y^{A_t}}{\partial t} \right) \Psi(A_t, \cdot) \\ &= \left( \frac{dA_t \circ \xi_S}{dt} \mathcal{L}_Y^{A_t} + \frac{d}{dt} \left( e^{A_t \circ \xi_S} \langle \mathcal{D}(V, A_t), \nabla \cdot \rangle \right) \right) (\Psi(A_t, \cdot)). \end{aligned}$$

Considering each example for the definition of the drift function  $\mathcal{D}(V, B_s)$ , it is straightforward to check that  $-\left(\frac{d\mathcal{L}_y^{A_t}}{ds}\right)\Psi(A_t, \cdot) \in \mathcal{C}$  is a function of class  $\mathcal{C}^\infty$  with polynomial growth; and, moreover, that for every  $k \in \{0, 1, \dots\}$ , there exist  $p_k \geq 0$  and  $C^{(k)} \in (0, \infty)$  such that

$$\sup_{y \in \mathcal{S}} \left| D^k \left( \frac{\partial \mathcal{L}_y^{A_t}}{\partial t} \right) \Psi(A_t, \cdot)(x) \right| \leq \frac{C^{(k)}(1+|x|^{p_k})}{1+t},$$

thanks to the inequality (8.10), and the estimate (5.15) on the gradient  $\nabla_x \Psi(A_t, x)$ .

Thanks to Proposition 5.2, one then concludes the proof of (5.16).

**8.2. The infinite dimensional case.** The aim of this section is to provide a proof of the estimates (7.8) which are specific to the infinite dimensional framework. As explained above, we only focus on the case  $A=0$ , and to simplify notation,  $\bar{\varphi} = \bar{\varphi}^0$  and  $\Psi = \Psi(0, \cdot)$ .

Introduce the semi-group  $(P_t)_{t \geq 0}$ , such that for all  $t \geq 0$

$$P_t \bar{\varphi}(u) = \mathbb{E}_u[\bar{\varphi}^0(u(t))].$$

**8.2.1. First-order derivative.** We claim that

$$\begin{aligned} |\langle DP_t \bar{\varphi}(u), h \rangle| &\leq e^{-\gamma t} \sup_{v \in H} \|D\bar{\varphi}(v)\| \|h\|, \quad \forall t \geq 0 \\ |\langle DP_t \bar{\varphi}(u), h \rangle| &\leq \frac{C_\alpha \sup_{v \in H} \|D\bar{\varphi}(v)\|}{t^{2\alpha}} \|h\|_{-2\alpha}, \quad \forall t \in (0, 1]. \end{aligned} \tag{8.11}$$

Then using the semi-group property  $P_t = P_1 P_{t-1}$ , for all  $t \geq 1$ ,

$$\begin{aligned} |\langle DP_t \bar{\varphi}(u), h \rangle| &\leq C_\alpha \sup_{v \in H} \|D(P_{t-1} \bar{\varphi})(v)\| \|h\|_{-2\alpha} \\ &\leq C_\alpha e^{-\gamma(t-1)} \sup_{v \in H} \|D\bar{\varphi}(v)\| \|h\|_{-2\alpha}. \end{aligned}$$

By integrating, this yields the first estimate in (7.8).

It remains to prove the claim (8.11). Note that

$$\langle DP_t \bar{\varphi}(u), h \rangle = \mathbb{E}[\langle D\bar{\varphi}^0(u(t)), \eta^h(t, u) \rangle],$$

where  $\eta^h(0, u) = h$  and

$$d\eta^h(t, u) = L\eta^h(t, u)dt - \mathcal{V}''(u(t))\eta^h(t, u)dt.$$

The proof of the first inequality of (8.11) is straightforward:

$$\begin{aligned} \frac{1}{2} \frac{d\|\eta^h(t, u)\|^2}{dt} &= \langle L\eta^h(t, u), \eta^h(t, u) \rangle - \langle \mathcal{V}''(u(t))\eta^h(t, u), \eta^h(t, u) \rangle \\ &\leq -(\lambda_1 - \sup_{x \in \mathbb{R}} |\mathcal{V}''(x)|) \|\eta^h(t, u)\|^2 \\ &\leq e^{-\gamma t} \|h\|^2, \end{aligned}$$

with  $\gamma = \lambda_1 - \sup_{x \in \mathbb{R}} |\mathcal{V}''(x)| > 0$  thanks to (7.3).

The second inequality of (8.11) is obtained using the mild formulation of the equation for  $\eta^h(t, u)$  and regularization properties of the semi-group  $(e^{tL})_{t \geq 0}$ :

$$\begin{aligned} \|\eta^h(t, u)\|_H &\leq \|e^{tL}h\|_H + \int_0^t \|e^{(t-s)L}(\mathcal{V}''(u(s))\eta^h(s, u))\|_H ds \\ &\leq C(\alpha)t^{-2\alpha} \|(-L)^{-2\alpha}h\|_H + C \int_0^t \|\eta^h(s, u)\|_H ds, \end{aligned}$$

and  $\|(-L)^{-2\alpha}h\|_H = \|h\|_{-2\alpha}$ . Thanks to Grönwall's Lemma, there exists  $C(\alpha) \in (0, \infty)$  such that for all  $t \in (0, 1]$ ,

$$\|\eta^h(t, u)\|_H \leq C(\alpha)t^{-2\alpha} \|h\|_{-2\alpha},$$

which yields the required estimate.

**8.2.2. Second-order derivative.** We claim that, for some  $\tilde{\gamma} \in (0, \gamma)$ ,

$$\begin{aligned} |D^2 P_t \bar{\varphi}(u).(h, k)| &\leq C \left( \sup_{v \in H} \|D\bar{\varphi}(v)\| + \sup_{v \in H} \|D^2 \bar{\varphi}(v)\| \right) e^{-\tilde{\gamma}t} \|h\| \|k\|, \quad \forall t \geq 0 \\ |D^2 P_t \bar{\varphi}(u).(h, k)| &\leq \frac{C_\alpha (\sup_{v \in H} \|D\bar{\varphi}(v)\| + \sup_{v \in H} \|D^2 \bar{\varphi}(v)\|)}{t^{2\alpha}} \|h\|_{-\alpha} \|k\|_{-\alpha}, \quad \forall t \in (0, 1]. \end{aligned} \tag{8.12}$$

The proof uses the following identity:

$$D^2 P_t \bar{\varphi}(u).(h, k) = \mathbb{E}[\langle D\bar{\varphi}(u(t)), \zeta^{h,k}(t, u) \rangle] + \mathbb{E}[D^2 \bar{\varphi}(u(t)).(\eta^h(t, u), \eta^k(t, u))],$$

where  $\zeta^{h,k}(0, u) = 0$  and

$$d\zeta^{h,k}(t, u) = L\zeta^{h,k}(t, u)dt - \mathcal{V}''(u(t))\zeta^{h,k}(t, u)dt - \mathcal{V}^{(3)}(u(t))\eta^h(t, u)\eta^k(t, u)dt.$$

The two following inequalities are used in the proof:

- the Gagliardo-Nirenberg inequality, for every  $v \in H_0^1(0, 1)$

$$\|v\|_{L^\infty(0,1)} \leq C \|v\|_{L^2(0,1)}^{\frac{1}{2}} \|v\|_{H^1(0,1)}^{\frac{1}{2}},$$

$$\text{combined with } \|v\|_{H^1(0,1)} \leq C \|(-L)^{\frac{1}{2}}v\|_{L^2(0,1)} = C \|v\|_{\frac{1}{2}},$$

- the Sobolev inequality  $\|\cdot\|_{L^4(0,1)} \leq C\|\cdot\|_{\frac{1}{8}}$ , which implies

$$\|e^{tL}\|_{\mathcal{L}(L^2(0,1),L^4(0,1))} \leq Ct^{-\frac{1}{8}}.$$

To prove the first inequality in (8.12), an energy estimate and the use of the Gagliardo-Nirenberg and Young's inequalities, give

$$\begin{aligned} & \frac{1}{2} \frac{d\|\zeta^{h,k}(t,u)\|_{L^2(0,1)}^2}{dt} + \|(-L)^{\frac{1}{2}}\zeta^{h,k}(t,u)\|_{L^2(0,1)}^2 \\ & \leq \|\mathcal{V}''\|_{\infty} \|\zeta^{h,k}(t,u)\|_{L^2(0,1)}^2 \\ & \quad + \|\mathcal{V}^{(3)}\|_{\infty} \|\zeta^{h,k}(t,u)\|_{L^\infty(0,1)} \|\eta^h(t,u)\|_{L^2(0,1)} \|\eta^k(t,u)\|_{L^2(0,1)} \\ & \leq (\lambda_1 - \gamma + \epsilon) \|\zeta^{h,k}(t,u)\|_{L^2(0,1)}^2 + \epsilon \|(-L)^{\frac{1}{2}}\zeta^{h,k}(t,u)\|_{L^2(0,1)}^2 \\ & \quad + \frac{C}{\epsilon} \|\eta^h(t,u)\|_{L^2(0,1)}^2 \|\eta^k(t,u)\|_{L^2(0,1)}^2, \end{aligned}$$

for  $\epsilon > 0$  sufficiently small.

Using the Poincaré inequality  $\|(-A)^{\frac{1}{2}} \cdot\|_{L^2(0,1)}^2 \geq \lambda_1 \|\cdot\|_{L^2(0,1)}^2$ , then

$$\frac{1}{2} \frac{d\|\zeta^{h,k}(t,u)\|_{L^2(0,1)}^2}{dt} \leq -(\gamma - (1 + \lambda_1)\epsilon) \|\zeta^{h,k}(t,u)\|_{L^2(0,1)}^2 + C_\epsilon e^{-4\gamma t} \|h\|_{L^2(0,1)}^2 \|k\|_{L^2(0,1)}^2.$$

By Grönwall's Lemma, since  $\zeta^{h,k}(0,u) = 0$ ,

$$\begin{aligned} \|\zeta^{h,k}(t,u)\|_{L^2(0,1)}^2 & \leq \int_0^t e^{-2(\gamma - (1 + \lambda_1)\epsilon)(t-s)} e^{-4\gamma s} ds \|h\|_{L^2(0,1)}^2 \|k\|_{L^2(0,1)}^2 \\ & \leq C_\epsilon e^{-2(\gamma - (1 + \lambda_1)\epsilon)t} \|h\|_{L^2(0,1)}^2 \|k\|_{L^2(0,1)}^2. \end{aligned}$$

This concludes the proof of the first estimate in (8.12).

To obtain the second inequality in (8.12), it is sufficient to estimate (using the mild formulation)

$$\begin{aligned} \|\zeta^{h,k}(t,u)\| & \leq \int_0^t \|e^{(t-s)L} (\mathcal{V}''(u(s))\zeta^{h,k}(s,u))\| ds \\ & \quad + \int_0^t \|e^{(t-s)L} (\mathcal{V}^{(3)}(u(s))\eta^h(s,u)\eta^k(s,u))\| ds \\ & \leq C \int_0^t \|\zeta^{h,k}(s,u)\| ds + \int_0^t \|e^{(t-s)L}\|_{\mathcal{L}(L^1,L^2)} \|\eta^h(s,u)\| \|\eta^k(s,u)\| ds \\ & \leq C \int_0^t \|\zeta^{h,k}(s,u)\| ds + \int_0^t \frac{C_{\alpha,\epsilon}}{(t-s)^{\frac{1}{4} + \epsilon} s^{2\alpha}} ds \|(-L)^{-\alpha} h\| \|(-L)^{-\alpha} k\| \\ & \leq C \int_0^t \|\zeta^{h,k}(s,u)\| ds + \frac{C_\alpha}{t^{2\alpha}} \|h\|_{-\alpha} \|k\|_{-\alpha}, \end{aligned}$$

and the conclusion follows from Grönwall's Lemma.

## REFERENCES

- [1] A. Barducci, G. Bussi, and M. Parrinello, *Well-tempered metadynamics: a smoothly converging and tunable free-energy method*, Phys. Rev. Lett., **100(2):020603**, 2008. [1](#)
- [2] M. Benaïm, *Dynamics of stochastic approximation algorithms*, in Séminaire de Probabilités, XXXIII, Lecture Notes in Math., Springer, Berlin, **1709:1–68**, 1999. [1.1](#), [1.3](#), [5](#), [5.1.2](#)



- [3] M. Benaïm and C.-E. Bréhier, *Convergence of adaptive biasing potential methods for diffusions*, C.R. Math. Acad. Sci. Paris, **354(8):842–846**, 2016. 1, 1.1, 1.3, 4.2.1
- [4] M. Benaïm, M. Ledoux, and O. Raimond, *Self-interacting diffusions*, Probab. Theory Related Fields, **122(1):1–41**, 2002. 1.1, 1.3, 5, 5.1.2, 5.1.2
- [5] M. Benaïm and O. Raimond, *Self-interacting diffusions. II. Convergence in law*, Ann. Inst. H. Poincaré Probab. Statist., **39(6):1043–1055**, 2003. 1.1
- [6] M. Benaïm and O. Raimond, *Self-interacting diffusions. III. Symmetric interactions*, Ann. Probab., **33(5):1716–1759**, 2005. 1.1
- [7] M. Benaïm and O. Raimond, *Self-interacting diffusions IV: Rate of convergence*, Electron. J. Probab., **16(66):1815–1843**, 2011. 1.1
- [8] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Appl. Math. (New York), Springer-Verlag, Berlin, **22**, 1990. 1.1, 1.3, 5
- [9] N. Berglund and B. Gentz, *Sharp estimates for metastable lifetimes in parabolic SPDEs: Kramers’ law and beyond*, Electron. J. Probab., **18(24):58**, 2013. 7.1
- [10] A. Bovier and F. den Hollander, *Metastability*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer, Cham, **351**, 2015. 7.1
- [11] C.-E. Bréhier, M. Gazeau, L. Goudenège, and M. Rousset, *Analysis and simulation of rare events for SPDEs*, CEMRACS 2013 - Modelling and Simulation of Complex Systems: Stochastic and Deterministic, ESAIM Proc. Surveys, EDP Sci., Les Ulis, **48:364–384**, 2015. 7.1
- [12] C.-E. Bréhier and M. Kopec, *Approximation of the invariant law of PDE’s: error analysis using a Poisson equation for a full-discretization scheme*, IMA J. Numer. Anal., **37(3):1375–1410**, 2017. 7.3
- [13] S. Cerrai, *Second order PDE’s in finite and infinite dimension: A probabilistic approach*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, **1762**, 2001. 7.3
- [14] C. Chipot and T. Lelièvre, *Enhanced sampling of multidimensional free-energy landscapes using adaptive biasing forces*, SIAM J. Appl. Math., **71(5):1673–1695**, 2011. 1
- [15] J. Comer, J. C. Gumbart, J. Hénin, T. Lelièvre, A. Pohorille, and C. Chipot, *The adaptive biasing force method: Everything you always wanted to know but were afraid to ask*, J. Phys. Chem. B, **119(3):1129–1151**, 2014. 1
- [16] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite Dimensional Systems*, London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, **229**, 1996. 7.1, 7.1.2
- [17] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, Second Edition, **152**, 2014. 7.1, 7.1
- [18] E. Darve and A. Pohorille, *Calculating free energies using average force*, J. Chem. Phys., **115(20):9169–9183**, 2001. 1
- [19] B. Dickson, F. Legoll, T. Lelièvre, G. Stoltz, and P. Fleurat-Lessard, *Free energy calculations: An efficient adaptive biasing potential method*, J. Phys. Chem. B, **114:5823–5830**, 2010. 1
- [20] B. M. Dickson, *Survey of adaptive biasing potentials: comparisons and outlook*, Curr. Opin. Struct. Biol., **43:63–67**, 2017. 1
- [21] M. Duflo, *Random Iterative Models*, Appl. Math. (New York), Springer-Verlag, Berlin, **34**, 1997. 1.1, 1.3, 5
- [22] W. G. Faris and G. Jona-Lasinio, *Large fluctuations for a nonlinear heat equation with noise*, J. Phys. A: Math. General, **15(10):3025**, 1982. 7.1
- [23] G. Fort, B. Jourdain, E. Kuhn, T. Lelièvre, and G. Stoltz, *Efficiency of the Wang-Landau algorithm: a simple test case*, Appl. Math. Res. Express. AMRX, **(2):275–311**, 2014. 1
- [24] G. Fort, B. Jourdain, E. Kuhn, T. Lelièvre, and G. Stoltz, *Convergence of the Wang-Landau algorithm*, Math. Comp., **84(295):2297–2327**, 2015. 1
- [25] G. Fort, B. Jourdain, T. Lelièvre, and G. Stoltz, *Self-healing umbrella sampling: convergence and efficiency*, Stat. Comput., **27(1):147–168**, 2017. 1
- [26] M. Hairer and J. C. Mattingly, *Yet another look at Harris’ ergodic theorem for Markov chains*, in Seminar on Stochastic Analysis, Random Fields and Applications VI, Progr. Probab., Birkhäuser/Springer Basel AG, Basel, **63:109–117**, 2011. 8.1.1, 8.1.1
- [27] J. Hénin and C. Chipot, *Overcoming free energy barriers using unconstrained molecular dynamics simulations*, J. Chem. Phys., **121(7):2904–2914**, 2004. 1
- [28] K. Ichihara and H. Kunita, *A classification of the second order degenerate elliptic operators and its probabilistic characterization*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, **30:235–254**, 1974. 8.1.1, 8.1.1
- [29] B. Jourdain, T. Lelièvre, and R. Roux, *Existence, uniqueness and convergence of a particle approximation for the adaptive biasing force process*, M2AN Math. Model. Numer. Anal., **44(5):831–865**, 2010. 1
- [30] M. Kopec, *Quelques contributions à l’analyse numérique d’équations stochastiques*, PhD Thesis,

- ENS Rennes., 2014. [7.3](#), [8.1.2](#)
- [31] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Second Edition, Appl. Math., Stochastic Modelling and Applied Probability, Springer-Verlag, New York, [35](#), [2003](#). [1.1](#), [1.3](#), [5](#)
- [32] A. Laio and M. Parrinello, *Escaping free-energy minima*, Proc. Natl. Acad. Sci. USA, [99\(20\):12562–12566](#), 2002. [1](#)
- [33] T. Lelièvre and K. Minoukadeh, *Long-time convergence of an adaptive biasing force method: the bi-channel case*, Arch. Ration. Mech. Anal., [202\(1\):1–34](#), 2011. [1](#)
- [34] T. Lelièvre, M. Rousset, and G. Stoltz, *Computation of free energy profiles with parallel adaptive dynamics*, J. Chem. Phys., [126\(13\):134111](#), 2007. [1](#)
- [35] T. Lelièvre, M. Rousset, and G. Stoltz, *Long-time convergence of an adaptive biasing force method*, Nonlinearity, [21\(6\):1155–1181](#), 2008. [1](#)
- [36] T. Lelièvre, M. Rousset, and G. Stoltz, *Free Energy Computations: A Mathematical Perspective*, Imperial College Press, London, 2010. [1](#)
- [37] T. Lelièvre and G. Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*, Acta Numer., [25:681–880](#), 2016. [1](#)
- [38] A. Lesage, T. Lelièvre, G. Stoltz, and J. Hénin, *Smoothed biasing forces yield unbiased free energies with the extended-system adaptive biasing force method*, J. Phys. Chem. B, [121\(15\):3676–3685](#), 2017. [1.2](#)
- [39] S. Marsili, A. Barducci, R. Chelli, P. Procacci, and V. Schettino, *Self-healing umbrella sampling: a non-equilibrium approach for quantitative free energy calculations*, J. Phys. Chem. B, [110\(29\):14011–14013](#), 2006. [1](#)
- [40] F. Wang and D. Landau, *Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram*, Phys. Rev. E, [64\(5\):056101](#), 2001. [1](#)
- [41] F. Wang and D. Landau, *Efficient, multiple-range random walk algorithm to calculate the density of states*, Phys. Rev. Lett., [86\(10\):2050](#), 2001. [1](#)
- [42] E. Vanden-Eijnden and J. Weare, *Rare event simulation of small noise diffusions*, Comm. Pure Appl. Math., [65\(12\):1770–1803](#), 2012. [7.1](#)