Long-time convergence of an adaptive biasing force method: Variance reduction by Helmholtz projection

Houssam Alrachid, Tony Lelièvre

Long-time convergence of an adaptive biasing force method: Variance reduction by Helmholtz projection

HOUSSAM ALRACHID
École des Ponts ParisTech, Université Paris Est, 6-8 Avenue Blaise Pascal, Cité Descartes Marne-la-Vallée, F-77455 , France
alrachih@cermics.enpc.fr

Université Libanaise, Ecole Doctorale des Sciences et de Technologie, Campus Universitaire de Rafic Hariri Hadath, Lebanon
houssam.alrachid@hotmail.com

TONY LELIEVRE
École des Ponts ParisTech, Université Paris Est, 6-8 Avenue Blaise Pascal, Cité Descartes Marne-la-Vallée, F-77455 , France
lelievre@cermics.enpc.fr

In this paper, we propose an improvement of the adaptive biasing force (ABF) method, by projecting the estimated *mean force* onto a gradient. The associated stochastic process satisfies a non linear stochastic differential equation. Using entropy techniques, we prove exponential convergence to the stationary state of this stochastic process. We finally show on some numerical examples that the variance of the approximated *mean force* is reduced using this technique, which makes the algorithm more efficient than the standard ABF method.

Keywords: Adaptive biasing force; Helmholtz projection; Free energy; Variance reduction.

1. Introduction

1.1. The model

Let us consider the *Boltzmann-Gibbs* measure :

$$\mu(dx) = Z_\mu^{-1} e^{-\beta V(x)} dx, \tag{1.1}$$

where $x \in \mathcal{D}^N$ denotes the position of N particles in $\mathcal{D}$. The space $\mathcal{D}$ is called the configuration space. One should think of $\mathcal{D}$ as a subset of $\mathbb{R}^n$, or the n-dimensional torus $\mathbb{T}^n$ (where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denotes the one dimensional torus). The potential energy function $V : \mathcal{D} \longrightarrow \mathbb{R}$ associates to the positions of the particles $x \in \mathcal{D}$ its energy $V(x)$. In addition, $Z_\mu = \displaystyle\int_{\mathcal{D}} e^{-\beta V(x)} dx$ (assumed to be finite) is the normalization constant and $\beta = 1/(k_B T)$ is proportional to the inverse of the temperature T, k_B being the Boltzmann constant.

The probability measure μ is the equilibrium measure sampled by the particles in the canonical statistical ensemble. A typical dynamics that can be used to sample this measure is the *Overdamped Langevin Dynamics*:

$$dX_t = -\nabla V(X_t)dt + \sqrt{\frac{2}{\beta}}dW_t, \tag{1.2}$$

where $X_t \in \mathcal{D}^N$ and W_t is a Nn-dimensional standard Brownian motion. Under loose assumptions on V, the dynamics $(X_t)_{t\geq 0}$ is ergodic with respect to the measure μ, which means: for any smooth test function φ,

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T \varphi(X_t)dt = \int \varphi d\mu, \tag{1.3}$$

i.e. trajectory averages converge to canonical averages.

1.2. Metastability, reaction coordinate and free energy

In many cases of interest, there exists regions of the configuration space where the dynamics (1.2) remains trapped for a long time, and jumps only occasionally to another region, where it again remains trapped for a long time. This typically occurs when there exist high probability regions separated by very low probability areas. The regions where the process $(X_t)_{t\geq 0}$ remains trapped for very long times, are called metastable.

Because of the metastability, trajectorial averages (1.3) converge very slowly to their ergodic limit. Many methods have been proposed to overcome this difficulty, and we concentrate here on the Adaptive Biasing Force (denoted ABF) method (see [5, 7]). In order to introduce the ABF method, we need another ingredient: a *reaction coordinate* (also known as an order parameter), $\xi = (\xi_1, ..., \xi_m) : \mathcal{D} \longrightarrow \mathbb{R}^m$, $\xi(x) = z$, where $m < nN$. Typically, in (1.2), the time-scale for the dynamics on $\xi(X_t)$ is larger than the time-scale for the dynamics on X_t due to the metastable states, so that ξ can be understood as a function such that $\xi(X_t)$ is in some sense a slow variable compared to X_t. We can say that ξ describes the metastable states of the dynamics associated to the potential V. For a given configuration x, $\xi(x)$ represents some macroscopic information. For example, it could represent angles or bond lengths in a protein, positions of defects in a material, etc ... In any case, it is meant to be a function with values in a small dimensional space (i.e. $m \leq 4$), since otherwise, it is difficult to approximate accurately the associated free energy which is a scalar function defined on the range of ξ (see equation (1.5) below). The choice of a "good" reaction coordinate is a highly debatable subject in the literature. One aim of the mathematical analysis conducted here or in previous papers (see for example [9]) is to quantify the efficiency of numerical algorithms once a reaction coordinate has been chosen.

The image of the measure μ by ξ is defined by:

$$\xi * \mu := \exp(-\beta A(z))dz, \tag{1.4}$$

where A is the so-called free energy associated with the reaction coordinate ξ. By the co-formula (see [9], Appendix A), the following formula for the free energy can then be obtained: up to an additive constant,

$$A(z) = -\beta^{-1} \ln(Z_{\Sigma_z}), \tag{1.5}$$

where $Z_{\Sigma_z} = \displaystyle\int_{\Sigma_z} e^{-\beta V(x)} \delta_{\xi(x)-z}(dx)$, the submanifold Σ_z is defined by

$$\Sigma_z = \{x = (x_1, ..., x_n) \in \mathcal{D} \,|\, \xi(x) = z\},$$

and $\delta_{\xi(x)-z}(dx)$ represents a measure with support Σ_z, such that $\delta_{\xi(x)-z}(dx)dz = dx$ (for further details on delta measures, we refer to [10], Section 3.2.1). We assume henceforth that ξ and V are such that $Z_{\Sigma_z} < \infty$, for all $z \in \mathbb{R}^m$.

The idea of free energy biasing methods, such as the adaptive biasing force method (see [5, 7]) or the Wang Landau algorithm (see [14]), is that, if ξ is well chosen, the dynamics associated to $V - A \circ \xi$ is less metastable than the dynamics associated to V. Indeed, from the definition of the free energy (1.4), for any compact subspace $\mathcal{M} \subset \mathbb{R}^m$, the image of $\tilde{Z}^{-1} \mathrm{e}^{(-\beta(V - A\circ\xi)(x))} 1_{\xi(x)\in\mathcal{M}}$ by ξ is the uniform law $\dfrac{1_{\mathcal{M}}}{|\mathcal{M}|}$, where $\tilde{Z} = \displaystyle\int_{Z_{\Sigma_z}} \mathrm{e}^{(-\beta(V - A\circ\xi)(x))} 1_{\xi(x)\in\mathcal{M}}$ and $|\mathcal{M}|$ denotes the Lebesgue measure on $\mathcal{M}$. The uniform law is typically easier to sample than the original measure $\xi * \mu$. If the function ξ is well chosen (i.e. if the dynamics in the direction orthogonal to ξ is not too metastable), the free energy can be used as a biasing potential to accelerate the sampling of the dynamics (see [9]). The difficulty is of course that the free energy A is unknown and difficult to approximate using the original dynamics (1.2) because of metastability. Actually, in many practical cases, it is the quantity of interest that one would like to approximate by molecular dynamics simulations (see [4, 10]). The principle of adaptive biasing methods is thus to approximate A (or its gradient) on the fly in order to bias the dynamics and to reduce the metastable features of the original dynamics (1.2).

1.3. *Adaptive biasing force method (ABF)*

In order to introduce the ABF method, we need a formula for the derivatives of A. The so called *mean force* $\nabla A(z)$, can be obtained from (1.5) as (see [10], Section 3.2.2):

$$\nabla A(z) = \int_{\Sigma_z} f(x) d\mu_{\Sigma_z}, \tag{1.6}$$

where $d\mu_{\Sigma_z}$ is the probability measure μ conditioned to a fixed value z of the reaction coordinate:

$$d\mu_{\Sigma_z} = Z_{\Sigma_z}^{-1} \mathrm{e}^{-\beta V(x)} \delta_{\xi(x)-z}(dx), \tag{1.7}$$

and f is the so-called *local mean force* defined by

$$f_i = \sum_{j=1}^{m} G^{-1}_{i,j} \nabla \xi_j . \nabla V - \beta^{-1} \mathrm{div} \left(\sum_{j=1}^{m} G^{-1}_{i,j} \nabla \xi_j \right), \tag{1.8}$$

where $G = (G_{i,j})_{i,j=1,...,m}$, has components $G_{i,j} = \nabla \xi_i \cdot \nabla \xi_j$. This can be rewritten in terms of conditional expectation as: for a random variable X with law μ (defined by (1.1)),

$$\nabla A(z) = \mathbb{E}(f(X)|\xi(x) = z). \tag{1.9}$$

We are now in position to introduce the standard adaptive biasing force (ABF) technique, applied to the overdamped Langevin dynamics (1.2):

$$\begin{cases} dX_t = -\left(\nabla V - \sum_{i=1}^{m} F_t^i \circ \xi \nabla \xi_i + \nabla (W \circ \xi) \right)(X_t) dt + \sqrt{2\beta^{-1}} dW_t, \\ F_t^i(z) = \mathbb{E}[f_i(X_t)|\xi(X_t) = z], \; i = 1, ..., m, \end{cases} \tag{1.10}$$

where f is defined in (1.8). Compared with the original dynamics (1.2), two modifications have been made to obtain the ABF dynamics (1.10):

(1) First and more importantly, the force $\sum_{i=1}^{m} F_t^i \circ \xi \nabla \xi_i$ has been added to the original force $-\nabla V$. At time t, F_t approximates ∇A defined in (1.6).
(2) Second, a potential $W \circ \xi$ has been added. This is actually needed in the case when ξ lives in an unbounded domain. In this case, a so-called confining potential W is introduced so that the law of $\xi(X_t)$ admits a long-time limit $Z_W^{-1} \mathrm{e}^{-\beta W(z)} dz$ (see Remark 2 at the end of Section 2.2), where $Z_W = \int_{\mathrm{Ran}(\xi)} \mathrm{e}^{-\beta W}$ is assumed to be finite. When ξ is living in a compact subspace of $\mathbb{R}^m$, there is no need to introduce such a potential and the law of $\xi(X_t)$ converges exponentially fast to the uniform law on the compact subspace (as explained in Section 1.2 and Section 2.2). Typically, W is zero in a chosen compact subspace $\mathcal{M}$ of $\mathbb{R}^m$ and is harmonic outside $\mathcal{M}$. For example, in dimension two, suppose that $\xi = (\xi_1, \xi_2)$ and $\mathcal{M} = [\xi_{min}, \xi_{max}] \times [\xi_{min}, \xi_{max}]$, then W can be defined as:

$$W(z_1, z_2) = \sum_{i=1}^{2} 1_{z_i \geq \xi_{max}} (z_i - \xi_{max})^2 + \sum_{i=1}^{2} 1_{z_i \leq \xi_{min}} (z_i - \xi_{min})^2. \tag{1.11}$$

It is proven in [9] that, under appropriate assumptions, F_t converges exponentially fast to ∇A. In addition, for well chosen ξ, the convergence to equilibrium for (1.10) is much quicker than for (1.2). This can be quantified using entropy estimates and Logarithmic Sobolev Inequalities, see [9].

Notice that even though F_t converges to a gradient (∇A), there is no reason why F_t would be a gradient at time t. In this paper, we propose an alternative method,

where we approximate ∇A, at any time t, by a gradient denoted ∇A_t. The gradient ∇A_t is defined as the Helmholtz projection of F_t. One could expect improvements compared to the original ABF method since the variance of ∇A_t is then smaller than the variance of F_t (since A_t is a scalar function). Reducing the variance is important since the conditional expectation in (1.10) is approximated by empirical averages in practice.

1.4. *Projected adaptive biasing force method (PABF)*

A natural algorithm to reconstruct A_t from F_t, consists in solving the following Poisson problem:

$$\Delta A_t = \mathrm{div} F_t \quad \text{on } \mathcal{M}, \tag{1.12}$$

with appropriate boundary conditions depending on the choice of ξ and $\mathcal{M}$. More precisely, if ξ is periodic and $\mathcal{M}$ is the torus $\mathbb{T}^m$, then we are working with periodic boundary conditions. If ξ is with values in $\mathbb{R}^m$ and $\mathcal{M}$ is a bounded subset of $\mathbb{R}^m$, then Neumann boundary conditions are needed (see Remark 8 at the end of Section 2.3.2). To solve this Poisson problem, standard methods such as finite difference methods, finite element methods, spectral methods or Fourier transforms can be used. Note that (1.12) is the Euler equation associated with the minimization problem:

$$A_t = \underset{g \in H^1(\mathcal{M})/\mathbb{R}}{\mathrm{argmin}} \int_{\mathcal{M}} |\nabla g - F_t|^2, \tag{1.13}$$

where $H^1(\mathcal{M})/\mathbb{R} = \left\{ g \in H^1(\mathcal{M}) \,|\, \int_{\mathcal{M}} g = 0 \right\}$ denotes the subspace of $H^1(\mathcal{M})$ of zero average functions. In view of (1.13), A_t can be interpreted as the function such that its gradient is the closest to F_t. Solving (1.12) amounts to computing the so-called *Helmholtz-Hodge* decomposition of the vector field F_t as (see [6], Section 3):

$$F_t = \nabla A_t + R_t, \quad \text{on } \mathcal{M}, \tag{1.14}$$

where R_t is a divergence free vector field.

Finally, the *projected ABF dynamics* we propose to study is the following non linear stochastic differential equation:

$$\begin{cases} dX_t = -\nabla(V - A_t \circ \xi + W \circ \xi)(X_t)dt + \sqrt{2\beta^{-1}} dW_t, \\ \Delta A_t = \mathrm{div} F_t \quad \text{on } \mathcal{M}, \text{ with appropriate boundary conditions}, \\ F_t^i(z) = \mathbb{E}[f_i(X_t) | \xi(X_t) = z], \; i = 1, ..., m. \end{cases} \tag{1.15}$$

Compared with the standard ABF dynamics (1.10), the only modification is that the mean force F_t is replaced by ∇A_t, which is meant to be an approximation of ∇A at time t.

The main theoretical result of this paper is that A_t converges exponentially fast to the free energy A in $\mathcal{M}$ (at least in a specific setting and for a slightly

modified version of (1.15), see Section 2 for more details). Moreover, we illustrate numerically this result on a typical example. From a numerical point of view, the interest of the method is that the variance of the projected estimated mean force (i.e. ∇A_t) is smaller than the variance of the estimated mean force (i.e. F_t). We observe numerically that this variance reduction enables a faster convergence to equilibrium for PABF compared with the original ABF.

The paper is organized as follows. In Section 2, the longtime convergence of the projected ABF method is proven. Section 3 is devoted to a numerical illustration of the interest of the projected ABF compared to the standard ABF approach. Finally, the proofs of the results presented in Section 2 are provided in Section 4.

2. Longtime convergence of the projected ABF method

For the sake of simplicity, we assume in this Section that $\mathcal{D} = \mathbb{T}^n$ and that $\xi(x) = (x_1, x_2)$. Then ξ lives in the compact space $\mathcal{M} = \mathbb{T}^2$ and we therefore take $W = 0$. The free energy can be written as:

$$A(x_1, x_2) = -\beta^{-1} \ln(Z_{\Sigma_{(x_1,x_2)}}), \tag{2.1}$$

where $Z_{\Sigma_{(x_1,x_2)}} = \displaystyle\int_{\Sigma_{(x_1,x_2)}} e^{-\beta V(x)} dx_3 ... dx_n$ and $\Sigma_{(x_1,x_2)} = \{x_1, x_2\} \times \mathbb{T}^{n-2}$. The mean force becomes:

$$\nabla A(x_1, x_2) = \int_{\Sigma_{(x_1,x_2)}} f(x) d\mu_{\Sigma_{(x_1,x_2)}}, \tag{2.2}$$

where $f = (f_1, f_2) = (\partial_1 V, \partial_2 V)$ and the conditional probability measure $d\mu_{\Sigma_{(x_1,x_2)}}$ is:

$$d\mu_{\Sigma_{(x_1,x_2)}} = Z^{-1}_{\Sigma_{(x_1,x_2)}} \mathrm{e}^{-\beta V} dx_3 ... dx_n.$$

Finally, the vector field $F_t(x_1, x_2)$ writes $\displaystyle\int_{\Sigma_{(x_1,x_2)}} f d\mu_{\Sigma_{(x_1,x_2)}}(t, .)$, or equivalently:

$$F^i_t(x_1, x_2) = \mathbb{E}\big(\partial_i V(X_t) | \xi(X_t) = (x_1, x_2)\big),\ i = 1, 2.$$

2.1. *Helmholtz projection*

In section 2.1.1, weighted Helmholtz-Hodge decomposition of F_t is presented. In section 2.1.2, the associated minimization problem and projection operator are introduced.

Let us first fix some notations. For $x \in \mathbb{R}^n$ and $1 \leq i < j \leq n$, x^j_i denotes the vector $(x_i, x_{i+1}, ..., x_j)$ and dx^j_i denotes $dx_i\, dx_{i+1} \ldots dx_j$. Moreover, $\nabla_{x^2_1}$, $\mathrm{div}_{x^2_1}$ and $\Delta_{x^2_1}$ represent respectively the gradient, the divergence and the laplacian in dimension two for the first two variables (x_1, x_2). Likewise, $\nabla_{x^n_3} = (\partial_3, ..., \partial_n)^T$ represents the gradient vector starting from the third variable of $\mathbb{R}^n$.

2.1.1. *Helmholtz decomposition*

The space $\mathbb{T}^2$ is a bounded and connected space. For any smooth positive probability density function $\varphi : \mathbb{T}^2 \to \mathbb{R}$, let us define the weighted Hilbert space: $L^2_\varphi(\mathbb{T}^2) = \{f : \mathbb{T}^2 \to \mathbb{R},\ \int_{\mathbb{T}^2} |f|^2\varphi < \infty\}$. Let us also introduce the Hilbert space $H_\varphi(\mathrm{div}; \mathbb{T}^2) = \{g \in L^2_\varphi(\mathbb{T}^2) \times L^2_\varphi(\mathbb{T}^2),\ \mathrm{div}_{x_1^2}(g) \in L^2(\mathbb{T}^2)\}$. It is well-known that any vector field $F_t : \mathbb{T}^2 \to \mathbb{R}^2 \in H_1(\mathrm{div}, \mathbb{T}^2)$ can be written (see [6], Section 3 for example) as (Helmholtz decomposition): $F_t = \nabla_{x_1^2} A_t + R_t$, where R_t is a divergence free vector field. We will need a generalization of the standard Helmholtz decomposition to the weighted Hilbert spaces $L^2_\varphi(\mathbb{T}^2)$ and $H_\varphi(\mathrm{div}; \mathbb{T}^2))$:

$$F_t\varphi = \nabla_{x_1^2}(A_t)\varphi + R_t, \tag{2.3}$$

s.t. $\mathrm{div}_{x_1^2}(R_t) = 0$. This weighted Helmholtz decomposition is required to simplify calculations when studying the longtime convergence (see Remark 10 in Section 4.1 for more details). Recall the space: $H^1(\mathbb{T}^2)/\mathbb{R} = \left\{g \in H^1(\mathbb{T}^2) \,|\, \int_{\mathbb{T}^2} g = 0\right\}$. The function A_t is then the solution to the following problem:

$$\int_{\mathbb{T}^2} \nabla_{x_1^2} A_t \cdot \nabla_{x_1^2} g\, \varphi = \int_{\mathbb{T}^2} F_t \cdot \nabla_{x_1^2} g\, \varphi, \quad \forall g \in H^1(\mathbb{T}^2)/\mathbb{R}, \tag{2.4}$$

which is the weak formulation of the Poisson problem:

$$\mathrm{div}_{x_1^2}(\nabla_{x_1^2} A_t \varphi(t,.)) = \mathrm{div}_{x_1^2}(F_t\varphi(t,.)), \tag{2.5}$$

with periodic boundary conditions. Using standard arguments (Lax-Milgram theorem), it is straight forward to check that (2.4) admits a unique solution $A_t \in H^1(\mathbb{T}^2)/\mathbb{R}$.

2.1.2. *Minimization problem and projection on a gradient*

Proposition 1. *Suppose that $F_t \in H_\varphi(\mathrm{div}; \mathbb{T}^2)$. Then for any smooth positive probability density function φ, the equation* (2.4) *is the Euler Lagrange equation associated with the following minimization problem:*

$$A_t = \min_{h \in H^1(\mathbb{T}^2)/\mathbb{R}} \int_{\mathbb{T}^2} |\nabla_{x_1^2} h - F_t|^2 \varphi = \min_{h \in H^1(\mathbb{T}^2)/\mathbb{R}} ||\nabla_{x_1^2} h - F_t||^2_{L^2_\varphi(\mathbb{T}^2)}. \tag{2.6}$$

Furthermore, A_t belongs to $H^2(\mathbb{T}^2)$.

Proof. Let us introduce the application $I : H^1(\mathbb{T}^2)/\mathbb{R} \to \mathbb{R}^+$, defined by $I(g) = ||\nabla_{x_1^2} h - F_t||^2_{L^2_\varphi(\mathbb{T}^2)}$. It is easy to prove that I is α-convex and coercive, i.e. $\lim_{\|g\|_{H^1} \to +\infty} I(g) = +\infty$. Thus I admits a unique global minimum $A_t \in H^1(\mathbb{T}^2)/\mathbb{R}$.

Furthermore, $\forall \varepsilon > 0, \forall g \in H^1(\mathbb{T}^2)$,

$$\begin{aligned} I(A_t + \varepsilon g) &= \int_{\mathbb{T}^2} |\nabla_{x_1^2}(A_t + \varepsilon g) - F_t|^2 \varphi \\ &= \int_{\mathbb{T}^2} |\nabla_{x_1^2} A_t - F_t|^2 \varphi - 2\varepsilon \int_{\mathbb{T}^2} (\nabla_{x_1^2} A_t - F_t) \cdot \nabla_{x_1^2} g \, \varphi + \varepsilon^2 \int_{\mathbb{T}^2} |\nabla_{x_1^2} g|^2 \varphi \\ &= I(A_t) - 2\varepsilon \int_{\mathbb{T}^2} (\nabla_{x_1^2} A_t - F_t) \cdot \nabla_{x_1^2} g \, \varphi + \varepsilon^2 \int_{\mathbb{T}^2} |\nabla_{x_1^2} g|^2 \varphi. \end{aligned} \tag{2.7}$$

Since A_t is the minimum of I, then $I(A_t + \varepsilon g) \geq I(A_t)$, $\forall \varepsilon > 0, \forall g \in H^1(\mathbb{T}^2)$. By considering the asymptotic regime $\varepsilon \to 0$ in the last equation, one thus obtains the equation (2.4):

$$\int_{\mathbb{T}^2} \nabla_{x_1^2} A_t \cdot \nabla_{x_1^2} g \, \varphi = \int_{\mathbb{T}^2} F_t \cdot \nabla_{x_1^2} g \, \varphi, \, \forall g \in H^1(\mathbb{T}^2)/\mathbb{R}.$$

This is the weak formulation of the following problem (2.5) in $H^1(\mathbb{T}^2)/\mathbb{R}$:

$$\mathrm{div}_{x_1^2}(\nabla_{x_1^2} A_t \varphi(t,.)) = \mathrm{div}_{x_1^2}(F_t \varphi(t,.)).$$

Since φ is a smooth positive function, then $\exists \, \delta > 0$, s.t. $\varphi > \delta$. Furthermore, since $\mathrm{div}_{x_1^2}(F_t \varphi(t,.)) \in L^2(\mathbb{T}^2)$, thus $\Delta_{x_1^2} A_t \in L^2(\mathbb{T}^2)$. Therefore, using standard elliptic regularity results, $A_t \in H^2(\mathbb{T}^2)$. □

For any positive probability density function φ, the estimation vector field $\nabla_{x_1^2} A_t$ is the projection of F_t onto a gradient. In the following, we will use the notation:

$$\mathcal{P}_\varphi(F_t) = \nabla_{x_1^2} A_t, \tag{2.8}$$

where the projection operator $\mathcal{P}_\varphi$ is a linear projection defined from $H_\varphi(\mathrm{div}; \mathbb{T}^2)$ to $H^1(\mathbb{T}^2) \times H^1(\mathbb{T}^2)$. Notice in particular that $\mathcal{P}_\varphi \circ \mathcal{P}_\varphi = \mathcal{P}_\varphi$.

2.2. *The projected ABF (PABF) method*

We will study the longtime convergence of the following PABF dynamics:

$$\begin{cases} dX_t = -\nabla\big(V - A_t \circ \xi\big)(X_t)dt + \sqrt{2\beta^{-1}} dW_t, \\ \nabla_{x_1^2} A_t = \mathcal{P}_{\psi^\xi}(F_t), \\ F_t^i(x_1, x_2) = \mathbb{E}[\partial_i V(X_t) | \xi(X_t) = (x_1, x_2)], \; i = 1, 2, \end{cases} \tag{2.9}$$

where $\mathcal{P}_{\psi^\xi}$ is the linear projection defined by (2.8) and W_t is a standard nN-dimensional Brownian motion. Thanks to the diffusion term $\sqrt{2\beta^{-1}} dW_t$, X_t admits a smooth density ψ with respect to the the Lebesgue measure on $\mathbb{T}^n$ and ψ^ξ then denotes the marginal distribution of ψ along ξ:

$$\psi^\xi(t, x_1, x_2) = \int_{\Sigma_{(x_1, x_2)}} \psi(t, x) dx_3^n. \tag{2.10}$$

The dynamics (2.9) is the PABF dynamics (1.15) with $\xi(x) = (x_1, x_2)$, $W = 0$ and a weighted Helmholtz projection. The weight ψ^ξ is introduced to simplify the convergence proof (see Remark 10 in Section 4).

Remark 1. If the law of X_t is $\psi(t,x)dx$ then the law of $\xi(X_t)$ is $\psi^\xi(t,x_1,x_2)dx_1dx_2$ and the conditional distribution of X_t given $\xi(X_t) = (x_1,x_2)$ is (see (1.7) for a similar formula when $\psi = Z^{-1}_{\Sigma_{(x_1,x_2)}} \mathrm{e}^{-\beta V}$):

$$d\mu_{t,x_1,x_2} = \frac{\psi(t,x)dx_3^n}{\psi^\xi(t,x_1,x_2)}. \tag{2.11}$$

Indeed, for any smooth functions f and g,

$$\begin{aligned}\mathbb{E}(f(\xi(X_t))g(X_t)) &= \int_{\mathbb{T}^n} f(\xi(x))g(x)\psi(t,x)dx \\ &= \int_{\mathbb{T}^2}\int_{\Sigma_{(x_1,x_2)}} f\circ\xi g\psi dx_3^n dx_1dx_2 \\ &= \int_{\mathbb{T}^2} f(x_1,x_2)\frac{\int_{\Sigma_{(x_1,x_2)}} g\psi dx_3^n}{\psi^\xi(x_1,x_2)}\psi^\xi(x_1,x_2)dx_1dx_2.\end{aligned}$$

$\Diamond$

Let us now introduce the non linear partial differential equation (the so-called *Fokker-Planck equation*) which rules the evolution of the density $\psi(t,x)$ of X_t solution of (2.9):

$$\begin{cases} \partial_t\psi = \mathrm{div}\left(\left(\nabla V - \sum_{i=1}^2 \partial_i A_t\circ\xi\nabla\xi_i\right)\psi + \beta^{-1}\nabla\psi\right), \quad \text{for } (t,x)\in[0,\infty[\times\mathbb{T}^n, \\ \forall t\geq 0,\ \mathrm{div}(\nabla A_t\psi^\xi(t,.)) = \mathrm{div}(F_t\psi^\xi(t,.)),\ \text{in } \mathbb{T}^2 \text{ with periodic boundary conditions}, \\ \forall t\geq 0,\ \forall (x_1,x_2)\in\mathbb{T}^2,\ F_t^i(x_1,x_2) = \dfrac{\int_{\Sigma_{(x_1,x_2)}} \partial_i V\psi dx_3^n}{\psi^\xi(x_1,x_2)},\ i=1,2. \end{cases} \tag{2.12}$$

The first equation of (2.12) rewrites:

$$\partial_t\psi = \mathrm{div}[\nabla V\psi + \beta^{-1}\nabla\psi] - \partial_1((\partial_1 A_t)\psi) - \partial_2((\partial_2 A_t)\psi). \tag{2.13}$$

Suppose that (ψ, F_t, A_t) is a solution of (2.12) and let us introduce the expected long-time limits of ψ, ψ^ξ (defined by (2.10)) and μ_{t,x_1,x_2} (defined by (2.11)) respectively:

(1) $\psi_\infty = e^{-\beta(V - A\circ\xi)}$;
(2) $\psi^\xi_\infty = 1$ (uniform law);
(3)

$$\mu_{\infty,x_1,x_2} = Z^{-1}_{\Sigma_{(x_1,x_2)}} e^{-\beta V} dx_3^n. \tag{2.14}$$

Notice that the probability measure $\psi_\infty^\xi(x_1,x_2)dx_1dx_2$ is the image of the probability measure $\psi_\infty(x)dx$ by ξ and that $\mu_{\infty,x_1,x_2} = \mu_{\Sigma_{(x_1,x_2)}}$ defined in (1.7). Furthermore, we have that

$$\int_{\mathbb{T}^n} \psi_\infty = 1, \int_{\mathbb{T}^2} \psi_\infty^\xi = 1 \text{ and } \forall (x_1,x_2) \in \mathbb{T}^2, \quad \int_{\Sigma_{(x_1,x_2)}} d\mu_{\infty,x_1x_2} = 1.$$

Remark 2. In the case when $W \neq 0$, the first equation of the Fokker-Plank problem (2.12) becomes:

$$\partial_t \psi = \operatorname{div}\left(\nabla\left(V - A_t \circ \xi - W \circ \xi\right)\psi + \beta^{-1}\nabla\psi\right).$$

The expected long-time limits of ψ, ψ^ξ and μ_{t,x_1,x_2} are respectively:

(1) $\psi_\infty = Z_W^{-1} e^{-\beta(V - A\circ\xi - W\circ\xi)}$;
(2) $\psi_\infty^\xi = Z_W^{-1} e^{-\beta W}$;
(3) $\mu_{\infty,x_1,x_2} = Z_{\Sigma_{(x_1,x_2)}}^{-1} e^{-\beta V} dx_3^n$,

where $Z_W = \displaystyle\int_{\mathbb{T}^2} e^{-\beta W}$.

2.3. *Precise statements of the longtime convergence results*

In section 2.3.1, some well-known results on entropy techniques are presented. For a general introduction to logarithmic Sobolev inequalities, their properties and their relation to long-time behaviours of solutions to partial differential equations, we refer to [1, 2, 13]. Section 2.3.2 presents the main theorem of convergence.

2.3.1. *Entropy and Fisher information*

Define the relative entropy $H(.|.)$ as follows: for any probability measures μ and ν such that μ is absolutely continuous with respect to ν (denoted $\mu \ll \nu$),

$$H(\mu|\nu) = \int \ln\left(\frac{d\mu}{d\nu}\right) d\mu.$$

Abusing the notation, we will denote $H(\varphi|\psi)$ for $H(\varphi(x)dx|\psi(x)dx)$ in case of probability measures with densities. Let us recall the Csiszar-Kullback inequality (see [2]):

$$\|\mu - \nu\|_{TV} \leq \sqrt{2H(\mu|\nu)}, \tag{2.15}$$

where $\|\mu-\nu\|_{TV} = \displaystyle\sup_{\|f\|_{L^\infty}\leq 1}\left\{\int f d(\mu-\nu)\right\}$ is the total variation norm of the signed measure $\mu - \nu$. When both μ and ν have densities with respect to the Lebesque measure, $\|\mu-\nu\|_{TV}$ is simply the L^1 norm of the difference between the two densities. The entropy $H(\mu|\nu)$ can be understood as a measure of how close μ and ν are.

Now, let us define the Fisher information of μ with respect to ν:

$$I(\mu|\nu) = \int \left|\nabla \ln\left(\frac{d\mu}{d\nu}\right)\right|^2 d\mu. \tag{2.16}$$

The Wasserstein distance is another way to compare two probability measures μ and ν defined on a space Σ,

$$W(\mu,\nu) = \sqrt{\inf_{\pi\in\prod(\mu,\nu)} \int_{\Sigma\times\Sigma} d_\Sigma(x,y)^2 d\pi(x,y)},$$

where the geodesic distance d_Σ on Σ is defined as: $\forall\, x, y \in \Sigma$,

$$d_\Sigma(x,y) = \inf\left\{ \sqrt{\int_0^1 |\dot{w}(t)|^2 dt} \;\mid\; w \in C^1([0,1],\Sigma),\, w(0)=x, w(1)=y \right\},$$

and $\prod(\mu,\nu)$ denotes the set of coupling probability measures, namely probability measures on $\Sigma\times\Sigma$ such that their marginals are μ and ν: $\forall\pi \in \prod(\mu,\nu)$, $\int_{\Sigma\times\Sigma}\phi(x)d\pi(x,y) = \int_\Sigma \phi d\mu$ and $\int_{\Sigma\times\Sigma}\psi(y)d\pi(x,y) = \int_\Sigma \psi d\nu$.

Definition 1. We say that a probability measure ν satisfies a logarithmic Sobolev inequality with constant $\rho > 0$ (denoted LSI(ρ)) if for all probability measure μ such that $\mu \ll \nu$,

$$H(\mu|\nu) \le \frac{1}{2\rho} I(\mu|\nu).$$

Definition 2. We say that a probability measure ν satisfies a Talagrand inequality with constant $\rho > 0$ (denoted T(ρ)) if for all probability measure μ such that $\mu \ll \nu$,

$$W(\mu,\nu) \le \sqrt{\frac{2}{\rho} H(\mu|\nu)}.$$

Remark 3. We implicitly assume in the latter definition, that the probability measures have finite moments of order 2. This is the case for the probability measures used in this paper.

The following lemma is proved in [11], Theorem 1:

Lemma 1. *If ν satisfies LSI(ρ), then ν satisfies T(ρ).*

Recall that X_t solution to (2.9) has a density $\psi(t,.)$. In the following, we denote the *Total Entropy* by

$$E(t) = H(\psi(t,.)|\psi_\infty) = \int_{\mathbb{T}^n} \ln(\psi/\psi_\infty)\psi, \tag{2.17}$$

the *Macroscopic Entropy* by

$$E_M(t) = H(\psi^\xi(t,.)|\psi^\xi_\infty) = \int_{\mathbb{T}^2} \ln(\psi^\xi/\psi^\xi_\infty)\psi^\xi, \tag{2.18}$$

and the *Microscopic Entropy* by

$$E_m(t) = \int_{\mathcal{M}} e_m(t,x_1,x_2)\psi^\xi(t,x_1,x_2)dx_1dx_2, \tag{2.19}$$

where $e_m(t,x_1,x_2) = H(\mu_{t,x_1,x_2}|\mu_{\infty,x_1,x_2})$. The following result is straightforward to check:

Lemma 2. *It holds, $\forall t \geq 0$,*

$$E(t) = E_M(t) + E_m(t).$$

Note that the Fisher information of μ_{t,x_1,x_2} with respect to μ_{∞,x_1,x_2} can be written as (see (2.16)):

$$I(\mu_{t,x_1,x_2}|\mu_{\infty,x_1,x_2}) = \int_{\Sigma_{(x_1,x_2)}} |\nabla_{x_3^n} \ln(\psi(t,.)/\psi_\infty)|^2 d\mu_{t,x_1,x_2}.$$

2.3.2. *Convergence of the PABF dynamics* (2.12)

The following proposition shows that the density function ψ^ξ satisfies a simple diffusion equation.

Proposition 2. *Suppose that (ψ, F_t, A_t) is a smooth solution of* (2.12)*. Then ψ^ξ satisfies:*

$$\begin{cases} \partial_t \psi^\xi = \beta^{-1} \Delta_{x_1^2} \psi^\xi, \ \textit{in } [0,\infty[\times\mathbb{T}^2, \\ \psi^\xi(0,.) = \psi_0^\xi, \ \textit{on } \mathbb{T}^2. \end{cases} \tag{2.20}$$

Remark 4. If $\psi_0^\xi = 0$ at some points or is not smooth, then F at time 0 may not be well defined or $I(\psi^\xi(0,.)/\psi_\infty^\xi)$ may be infinite. Since, by Proposition 2, ψ^ξ satisfies a simple diffusion equation these difficulties disappear as soon as $t > 0$. Therefore, up to considering the problem for $t > t_0 > 0$, we can suppose that ψ_0^ξ is a smooth positive function. We also have that for all $t > 0$, $\psi^\xi(t,.) > 0$, $\int_{\mathbb{T}^2} \psi^\xi = 1$ and $\psi^\xi(t,.) \in C^\infty(\mathbb{T}^2)$.

Remark 5. In the case where $W \neq 0$, the probability density function ψ^ξ satisfies the modified diffusion equation:

$$\partial_t \psi^\xi = \nabla_{x_1^2} \cdot \left(\beta^{-1} \nabla_{x_1^2} \psi^\xi + \psi^\xi \nabla_{x_1^2} W \right).$$

Here are two simple corollaries of Proposition 2.

Corollary 1. *There exists $t_0 > 0$ and $I_0 > 0$ (depending on ψ_0^ξ), such that*

$$\forall t > t_0, \quad I(\psi^\xi(t,.)|\psi_\infty^\xi) < I_0 e^{-\beta^{-1} 8\pi^2 t}.$$

Corollary 2. *The macroscopic entropy $E_M(t)$, defined by* (2.18)*, converges exponentially fast to zero:*

$$\forall t > t_0, \quad E_M(t) \leq \frac{I_0}{8\pi^2} e^{-\beta^{-1} 8\pi^2 t},$$

where I_0 is the constant introduced in Corollary 1.

The assumptions we need to prove the longtime convergence of the biasing force ∇A_t to the mean force ∇A are the following:

[**H1**] $V \in C^2(\mathbb{T}^n)$ and satisfies:

$$\exists \gamma > 0,\, \forall\, 3 \leq j \leq n, \forall x \in \mathbb{T}^n, \quad \max(|\partial_1 \partial_j V(x)|, |\partial_2 \partial_j V(x)|) \leq \gamma.$$

[**H2**] V is such that $\exists \rho > 0$, $\forall (x_1, x_2) \in \mathbb{T}^2$, $\mu_{\infty, x_1, x_2} = \mu_{\Sigma(x_1,x_2)}$ defined by (2.14) satisfies $LSI(\rho)$.

The main theorem is:

Theorem 1. *Let us assume [H1] and [H2]. The following properties then hold:*

(1) The microscopic entropy E_m converges exponentially fast to zero:

$$\exists C > 0, \exists \lambda > 0, \forall t \geq 0, \quad \sqrt{E_m(t)} \leq C e^{-\lambda t}. \tag{2.21}$$

Furthermore, if $\rho \neq 4\pi^2$, then $\lambda = \beta^{-1} \min(\rho, 4\pi^2)$ and $C = \sqrt{E_m(0)} + \dfrac{\gamma}{\beta^{-1}|\rho - 4\pi^2|}\sqrt{\dfrac{I_0}{2\rho}}$. If $\rho = 4\pi^2$, then for all $\lambda < \beta^{-1}\rho$, there exists a positive constant C such that (2.21) *is satisfied.*

(2) $\sqrt{E(t)}$ and $\|\psi(t,.) - \psi_\infty\|_{L^1(\mathbb{T}^n)}$ both converge exponentially fast to zero with rate λ.

(3) The biasing force $\nabla_{x_1^2} A_t$ converges to the mean force $\nabla_{x_1^2} A$ in the following sense:

$$\forall\, t \geq 0, \int_{\mathbb{T}^2} |\nabla_{x_1^2} A_t - \nabla_{x_1^2} A|^2 \psi^\xi(t, x_1, x_2) dx_1 dx_2 \leq \frac{8\gamma^2}{\rho} E_m(t). \tag{2.22}$$

The proofs of the results presented in this section are provided in Section 4.

Remark 6. We would like to emphasize that our arguments hold under the assumption of existence of regular solutions. In particular, we suppose that the density $\psi(t,.)$ is sufficiently regular so that the algebric manipulations in the proofs (see Section 4) are valid.

Remark 7. This remark is devoted to show how the rate of convergence of the dynamics (1.2) is improved thanks to PABF method. First of all, we mention a classical computation to get a rate of convergence for (1.2). Precisely, if one denotes $\varphi(t,.)$ the probability density function of X_t satisfying (1.2), and $\varphi_\infty = Z_\mu^{-1} \mathrm{e}^{-\beta V}$ its longtime limit, then by standard computations (see for example [2]), one obtains:

$$\frac{d}{dt} H(\varphi(t,.)|\varphi_\infty) = -\beta^{-1} I(\varphi(t,.)|\varphi_\infty).$$

Therefore, if φ_∞ satisfies $LSI(R)$, then one obtains the estimate

$$\exists R > 0, \forall t > 0, \quad H(\varphi(t,.)|\varphi_\infty) \leq H(\varphi_0|\varphi_\infty)\, \mathrm{e}^{-2\beta^{-1} R t}. \tag{2.23}$$

From (2.15), we obtain that $\|\varphi(t,.) - \varphi_\infty\|_{L^1(\mathbb{T}^n)}$ converges exponentially fast to zero with rate $\beta^{-1}R$. The constant R is known to be small if the metastable states

are separated by large energy barriers or if high probability regions for μ are separated by large regions with small probability (namely μ is a multimodal measure). Second, by Theorem 1, one can show that ∇A_t converges exponentially fast to ∇A in $L^2(\psi^\xi_\infty(x_1,x_2)dx_1dx_2)-$norm at rate $\lambda = \beta^{-1}\min(\rho, 4\pi^2)$. Indeed, since $\psi^\xi_\infty = 1$,

$$\begin{aligned}\int_{\mathbb{T}^2} |\nabla A_t - \nabla A|^2 dx_1dx_2 &= \int_{\mathbb{T}^2} |\nabla A_t - \nabla A|^2 \frac{\psi^\xi(t,x_1,x_2)}{\psi^\xi(t,x_1,x_2)} dx_1dx_2 \\ &\leq \frac{8\gamma^2}{(1-\varepsilon)\rho} E_m(t),\end{aligned}$$

where $\varepsilon > 0$ such that $\psi^\xi(t,x_1,x_2) \geq 1-\varepsilon$ (for more details refer to the proof of Corollary 1 in Section 4). This result must be compared with (2.23). More precisely, λ is related to the rate of convergence $4\pi^2$ at the macroscopic level, for equation (2.20) satisfied by ψ^ξ, and the rate of convergence ρ at the microscopic level, coming from the logarithmic Sobolev inequalities satisfied by the conditional measures μ_{∞,x_1,x_2}. Of course, ρ depends on the choice of the reaction coordinate. In our framework, we could state that a "good reaction coordinate" is such that ρ is as large as possible. Typically, for good choices of ξ, $\lambda \gg R$, and the PABF dynamics converges to equilibrium much faster than the original dynamics (1.2). This is typically the case if the conditional measures μ_{∞,x_1,x_2} are less multimodal than the original measure μ.

Remark 8. (Extension to other geometric setting)
The results of Theorem 1 are easily generalized to the following setting:

If $\mathcal{D} = \mathbb{R}^n$, $\xi(x) = (x_1, x_2)$ and $\mathcal{M}$ is a compact subspace of $\mathbb{R}^n$, then choose a confining potential W (defined in (1.11)) such that $Z_W = \int \mathrm{e}^{-\beta W} < +\infty$, $Z_W^{-1}\mathrm{e}^{-\beta W}$ satisfies $LSI(r^*)$ (for some $r^* > 0$) and W is convex potential, then Corollary 1 is satisfied with rate $2\beta^{-1}(r^* - \varepsilon)$, for any $\varepsilon \in (0, r^*)$ (refer to Corollary 1 in [9] for further details). In this case, Neumann boundary conditions are needed to solve the Poisson problem (2.5):

$$\begin{cases} \mathrm{div}(\nabla A_t \psi^\xi(t,.)) = \mathrm{div}(F_t \psi^\xi(t,.)) & \text{in } \mathcal{M}, \\ \dfrac{\partial A_t}{\partial n} = F_t.n & \text{on } \partial\mathcal{M}, \end{cases} \tag{2.24}$$

where n denotes the unit normal outward to $\mathcal{M}$. The convergence rate λ of Theorem 1 becomes $\beta^{-1}\min(\rho, r^* - \varepsilon)$. Neumann boundary conditions come from the minimization problem (2.6) associated to the Euler-Lagrange equation. The numerical applications in Section 3 are performed in this setting.

Remark 9. (Extension to more general reaction coordinates)
In this section, we have chosen $\xi(x_1, ..., x_n) = (x_1, x_2)$. The results can be extended to the following settings:

(1) In dimension one, the Helmholtz projection has obviously no sense. If $\mathcal{D} = \mathbb{T}^n$ and $\xi(x) = x_1$, then F_t converges to A', which is a derivative of a periodic

function and thus $\int_{\mathbb{T}} A' = 0$. Since $\int_{\mathbb{T}} F_t$ is not necessary equal to zero, one can therefore take a new approximation $A_t' = F_t - \int_{\mathbb{T}} F_t$, which approximates A' at any time t. The convergence results of this section can be extended to this setting, to show that A_t' converges exponentially fast to A'.

(2) More generally, for a reaction coordinate with values in $\mathbb{T}^m$, the convergence results presented in this paper still hold under the following orthogonality condition:

$$\forall i \neq j,\ \nabla\xi_i \cdot \nabla\xi_j = 0. \tag{2.25}$$

In the case when (2.25) does not hold, it is possible to resort to the following trick used for example in metadynamics (refer to [3, 8]). The idea is to introduce an additional variable z of dimension m, and an extended potential $V_\xi(x,z) = V(x) + \frac{\kappa}{2}|z - \xi(x)|^2$, where κ is a penalty constant. The reaction coordinate is then chosen as $\xi_{meta}(x,z) = z$, so that the associated free energy is:

$$A_\xi(z) = -\beta^{-1} \ln \int_{\mathcal{D}} e^{-\beta V_\xi(x,z)} dx,$$

which converges to $A(z)$ (defined in (1.5)) when κ goes to infinity. The extended PABF dynamics can be written as:

$$\begin{cases} dX_t = -\left(\nabla V(X_t) + \kappa \sum_{i=1}^{m} (\xi_i(X_t) - Z_{i,t}) \nabla\xi_i(X_t)\right) dt + \sqrt{2\beta^{-1}} dW_t, \\ dZ_t = \kappa(\xi(X_t) - \nabla E_t(Z_t))\, dt + \sqrt{2\beta^{-1}} d\overline{W}_t, \\ \nabla E_t = \mathcal{P}_{\psi^{\xi_{meta}}}(G_t), \\ G_t(z) = \mathbb{E}(\xi(X_t) | Z_t = z), \end{cases}$$

where $\overline{W}_t$ is a $m-$dimensional Brownian motion independent of W_t. The results of Theorem 1 apply to this extended PABF dynamics.

3. Numerical experiments

3.1. *Presentation of the model*

We consider a system composed of N particles $(q_i)_{0 \leq i \leq N-1}$ in a two-dimensional periodic box of side length L. Among these particles, three particles (numbered 0, 1 and 2 in the following) are designated to form a trimer, while the others are solvent particles. In this model, a trimer is a molecule composed of three identical particles linked together by two bonds (see Figure 1).

3.1.1. *Potential functions*

All particles, except the three particles forming the trimer, interact through the purely repulsive WCA pair potential, which is the Lennard-Jones (LJ) potential

truncated at the LJ potential minimum:

$$V_{WCA}(d) = \begin{cases} \varepsilon + 4\varepsilon \left[\left(\frac{\sigma}{d}\right)^{12} - \left(\frac{\sigma}{d}\right)^{6} \right] & \text{if } d \leq d_0, \\ 0 & \text{if } d \geq d_0, \end{cases}$$

where d denotes the distance between two particles, ε and σ are two positive parameters and $d_0 = 2^{1/6}\sigma$.

A particle of the solvent and a particle of the trimer also interact through the potential V_{WCA}. The interaction potential between two particles of the trimer (q_0/q_1 or q_1/q_2) is a double-well potential (see Figure 2):

$$V_S(d) = h \left[1 - \frac{(d - d_1 - \omega)^2}{\omega^2} \right]^2, \tag{3.1}$$

where d_1, h and ω are positive parameters.
The potential V_S has two energy minima. The first one, at $d = d_1$, corresponds to the compact bond. The second one, at $d = d_1 + 2\omega$, corresponds to the stretched bond. The height of the energy barrier separating the two states is h.

In addition, the interaction potential between q_0 and q_2 is a Lennard-Jones potential:

$$V_{LJ}(d) = 4\varepsilon' \left[\left(\frac{\sigma'}{d}\right)^{12} - \left(\frac{\sigma'}{d}\right)^{6} \right],$$

where ε' and σ' are two positive parameters.

Finally, the three particles of the trimer also interact through the following potential function on the angle θ formed by the vectors $\overrightarrow{q_1 q_0}$ and $\overrightarrow{q_1 q_2}$:

$$V_{\theta_0}(\theta) = \frac{k_\theta}{2} \left(\cos(\theta) - \cos(\theta_0)\right)^2,$$

where θ_0 is the equilibrium angle and k_θ is the angular stiffness. Figure 1 presents a schematic view of the system.

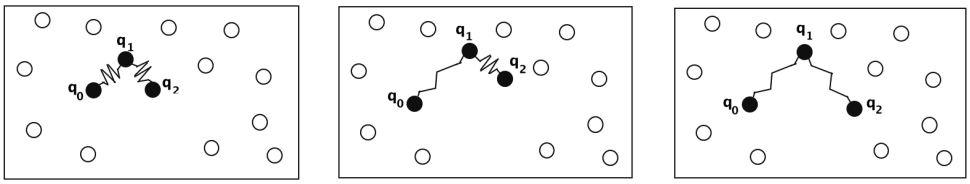

Fig. 1. Trimer (q_0, q_1, q_2). Left: compact state; Center: mixed state; Right: stretched state.

The total energy of the system is therefore, for $q \in (L\mathbb{T})^{2N}$:

$$V(q) = \sum_{3 \le i < j \le N-1} V_{WCA}(|q_i - q_j|) + \sum_{i=0}^{2} \sum_{j=3}^{N-1} V_{WCA}(|q_i - q_j|) + \sum_{i=0}^{1} V_S(|q_i - q_{i+1}|) + V_{LJ}(|q_0 - q_2|) + V_{\theta_0}(\theta).$$

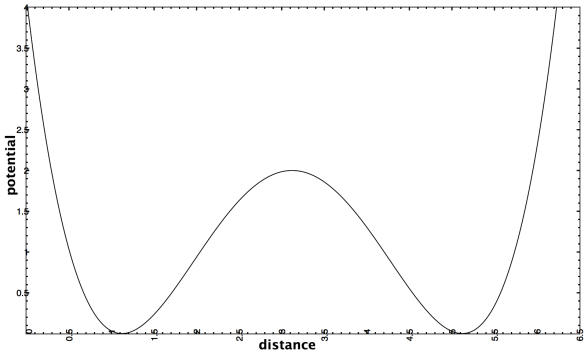

Fig. 2. Double-well potential (3.1), with $d_1 = 2^{1/6}$, $\omega = 2$ and $h = 2$.

3.1.2. *Reaction coordinate and physical parameters*

The reaction coordinate describes the transition from compact to stretched state in each bond. It is the normalised bond length of each bond of the trimer molecule. More precisely, the reaction coordinate is $\xi = (\xi_1, \xi_2)$, with $\xi_1(q) = \frac{|q_0 - q_1| - d_0}{2\omega}$ and $\xi_2(q) = \frac{|q_1 - q_2| - d_0}{2\omega}$. For $i = 1, 2$, the value $\xi_i = 0$ refers to the compact state (i.e. $d = d_0$) and the value $\xi_i = 1$ corresponds to the stretched state (i.e. $d = d_0 + 2\omega$).

We apply ABF and PABF dynamics to the trimer problem described above. The inverse temperature is $\beta = 1$, we use $N = 100$ particles ($N - 3$ solvent particles and the trimer) and the box side length is $L = 15$. The parameters describing the WCA and the Lennard-Jones interactions are set to $\sigma = 1$, $\varepsilon = 1$, $\sigma' = 1$, $\varepsilon' = 0.1$, $d_0 = 2^{1/6}$, $d_1 = 2^{1/6}$ and the additional parameters for the trimer are $\omega = 2$ and $h = 2$. The parameters describing the angle potential are: θ_0 such that $\cos(\theta_0) = 1/3$ and $k_\theta = 1$ (we refer to [12], Section 10.4.2, for the choice of such parameters). The initial condition on the trimer is as follows: Both bonds $q_0 q_1$ and $q_1 q_2$ are in compact state, which means that the distance between q_0 and q_1 and the distance between q_1 and q_2 are equal to d_0. Moreover, the initial bond angle is θ_0.

3.1.3. *Numerical methods and numerical parameters*

Standard and projected ABF methods are used with $N_{replicas} = 100$ replicas of the system evolving according to the overdamped Langevin dynamics discretized with a time-step $\Delta t = 2.5 \times 10^{-4}$. The reaction coordinate space of interest is taken of the form $\mathcal{M} = [\xi_{min}, \xi_{max}] \times [\xi_{min}, \xi_{max}]$, where $\xi_{min} = -0.2$ and $\xi_{max} = 1.2$. $\mathcal{M}$ is discretized into $N_{bins} \times N_{bins} = 50 \times 50 = 2500$ bins of equal sizes and $\delta = \delta_x = \delta_y = \frac{\xi_{max} - \xi_{min}}{N_{bins}} = 0.028$ denotes the size of each bin along both axes.

To implement the ABF and PABF method, one needs to approximate $F_t^i(x,y) = \mathbb{E}[f_i(x,y)|\xi(x,y) = (\xi_1(x,y), \xi_2(x,y))]$, $i = 1, 2$. The mean force F_t is estimated in each bin as a combination of plain trajectorial averages and averages over replicas. It is calculated at each time as an average of the local mean force in the bin over the total number of visits in this bin. More precisely, at time t and for $l = 1, 2$, the value of the mean force in the $(i,j)^{th}$ bin is:

$$F_t^l(i,j) = \frac{\displaystyle\sum_{t' \leq t} \sum_{k=1}^{N_{replicas}} f_l(q_{k,t'}) 1_{\{indx(\xi(q_{k,t'}))=(i,j)\}}}{\displaystyle\sum_{t' \leq t} \sum_{k=1}^{N_{replicas}} 1_{\{indx(\xi(q_{k,t'}))=(i,j)\}}}, \tag{3.2}$$

where $q_{k,t}$ denotes the position (x_k, y_k) at time t, $f = (f_1, f_2)$ is defined in (1.8) and $indx(\xi(q_{k,t'}))$ denotes the number of the bin where $\xi(q_{k,t'})$ lives, i.e.

$$indx(\xi(q)) = \left(\left[\frac{\xi_1(q) - \xi_{min}}{\delta} \right]_+, \left[\frac{\xi_2(q) - \xi_{min}}{\delta} \right]_+ \right), \forall q \in \mathcal{M}.$$

If the the components of the index function (i.e. $indx$) are either equal to -1 or to N_{bins}, it means that we are outside $\mathcal{M}$ and then the confining potential is non zero, and defined as:

$$W(\xi(q_{k,t})) = \sum_{i=1}^{2} \left[1_{\{\xi_i(q_{k,t}) \geq \xi_{max}\}} (\xi_i(q_{k,t}) - \xi_{max})^2 + 1_{\{\xi_i(q_{k,t}) \leq \xi_{min}\}} (\xi_i(q_{k,t}) - \xi_{min})^2 \right].$$

To construct the PABF method, the solution to the following Poisson problem with Neumann boundary conditions are approximated:

$$\begin{cases} \Delta A = \mathrm{div} F & \text{in } \mathcal{M} = [\xi_{min}, \xi_{max}] \times [\xi_{min}, \xi_{max}], \\ \frac{\partial A}{\partial n} = F \cdot n & \text{on } \partial\mathcal{M}, \end{cases} \tag{3.3}$$

where n denotes the unit normal outward to $\mathcal{M}$. We use for simplicity in the numerical experiments the standard Helmholtz problem, without the weight ψ^{ξ}. Problem (3.3) is solved using finite element method of type Q^1 on the quadrilateral mesh defined above, with nodes (x_i, y_j), where $x_i = \xi_{min} + i\delta$ and $y_j = \xi_{min} + j\delta$, for $i, j = 0, .., N_{bins}$. The space $\mathcal{M}$ is thus discretized into $N_T = N_{bins}^2$ squares, with $N_s = (N_{bins} + 1)^2$ nodes. The associated variational formulation is the following:

$$\begin{cases} \text{Find } A \in H^1(\mathcal{M})/\mathbb{R} \text{ such that} \\ \displaystyle\int_{\mathcal{M}} \nabla A \cdot \nabla v = \int_{\mathcal{M}} F \cdot \nabla v, \, \forall\, v \in H^1(\mathcal{M})/\mathbb{R}. \end{cases}$$

3.2. *Comparison of the methods*

In this section, we compare results obtained with three different simulations: without ABF, with ABF and with projected ABF (PABF). First, it is observed numerically that both ABF methods overcome metastable states. Second, it is illustrated how PABF method reduces the variance of the estimated mean force compared to ABF method. As a consequence of this variance reduction, we observe that the convergence of ∇A_t to ∇A with the PABF method is faster than the convergence of F_t to ∇A with the ABF method.

3.2.1. *Metastability*

To illustrate numerically the fact that ABF methods improve the sampling for metastable processes, we observe the variation, as a function of time, of the two metastable distances (i.e. the distance between q_0 and q_1, and the distance between q_1 and q_2). On Figures 3 and 4, the distance between q_1 and q_2 is plotted as a function of time for three dynamics: without ABF, ABF and PABF.

Both ABF methods allow to switch faster between the compact and stretched bond and thus to better explore the set of configurations. Without adding the biasing term, the system remains trapped in a neighborhood of the first potential minimum (i.e. $d_0 \simeq 1.12$) region for 20 units of time at least (see Figure 3), while when the biasing term is added in the dynamics, many jumps between the two local minima are observed (see Figure 4).

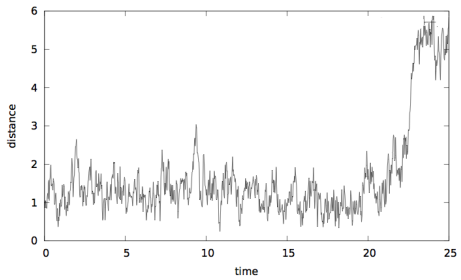

Fig. 3. Without ABF.

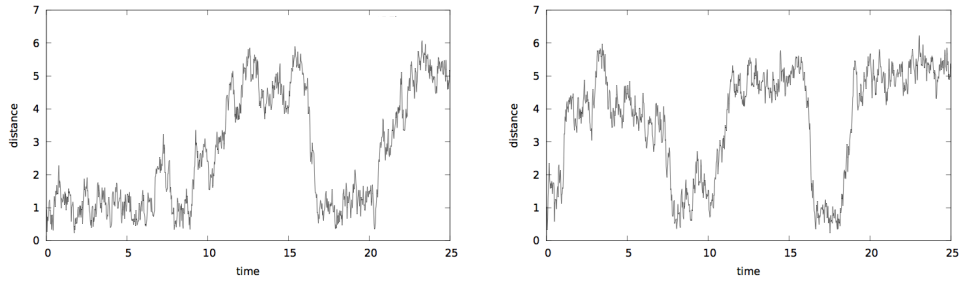

Fig. 4. Left: ABF; Right: PABF.

3.2.2. *Variance reduction*

Since we use Monte-Carlo methods to approximate ∇A_t, the variance is an important quantity to assess the quality of the result. The following general proposition shows that projection reduces the variance.

Proposition 3. *Let F be a random function from $\mathbb{T}^2$ into $\mathbb{R}^2$ and belongs to $H(\mathrm{div}, \mathbb{T}^2)$, and define $\mathcal{P} = \mathcal{P}_1$ (i.e. without weight) the projection on gradient vector fields defined in Section 2.1.2. Then, the variance of $\mathcal{P}(F)$ is smaller than the variance of F:*

$$\int_{\mathcal{M}} \mathrm{Var}(\mathcal{P}(F)) \leq \int_{\mathcal{M}} \mathrm{Var}(F),$$

where, for any vector field F, $\mathrm{Var}(F) = \mathbb{E}(|F|^2) - \mathbb{E}(|F|)^2$ and $|F|$ being the Euclidian norm.

Proof. Let F be a random vector field of $H(\mathrm{div}, \mathbb{T}^2)$. Let us introduce $\mathcal{P}(F) \in H^1(\mathbb{T}^2) \times H^1(\mathbb{T}^2)$ its projection. Notice that by the linearity of the projection $\mathcal{P}(\mathbb{E}(F)) = \mathbb{E}(\mathcal{P}(F))$. By definition of $\mathcal{P}(F)$, one gets:

$$\int_{\mathbb{T}^2} (F - \mathcal{P}(F)) \cdot \nabla h = 0, \, \forall h \in H^1(\mathbb{T}^2).$$

Therefore, using Pythagoras and the fact that $\mathcal{P}(F)$ is agradient,

$$\int_{\mathbb{T}^2} |F|^2 = \int_{\mathbb{T}^2} |F - \mathcal{P}(F)|^2 + \int_{\mathbb{T}^2} |\mathcal{P}(F)|^2$$

and

$$\int_{\mathbb{T}^2} |F - \mathbb{E}(F)|^2 = \int_{\mathbb{T}^2} |F - \mathbb{E}(F) - \mathcal{P}(F - \mathbb{E}(F))|^2 + \int_{\mathbb{T}^2} |\mathcal{P}(F - \mathbb{E}(F))|^2.$$

Using the linearity of $\mathcal{P}$, we thus obtain

$$\int_{\mathbb{T}^2} \mathrm{Var}(F) = \int_{\mathbb{T}^2} \mathrm{Var}(F - \mathcal{P}(F)) + \int_{\mathbb{T}^2} \mathrm{Var}(\mathcal{P}(F)),$$

which concludes the proof. □

We illustrate the improvement of the projected method in terms of the variances of the biasing forces, by comparing $\mathrm{Var}(\nabla A_t) = \mathrm{Var}(\partial_1 A_t) + \mathrm{Var}(\partial_2 A_t)$ (for the PABF method) and $\mathrm{Var}(F_t) = \mathrm{Var}(F_t^1) + \mathrm{Var}(F_t^2)$ (for the ABF method). Figure 5 shows that the variance for the projected ABF method is smaller than for the standard ABF method.

We have $N_{bins} \times N_{bins} = 2500$ variable for each term (i.e. $\partial_1 A_t$, $\partial_2 A_t$, F_t^1 and F_t^2). The variances are computed using 20 independent realizations as follows:

$$\mathrm{Var}(F_t^1) = \frac{1}{2500}\sum_{i,j=1}^{50}\frac{1}{20}\sum_{k=1}^{20} F_t^{1,k}(x_i,y_j)^2 - \frac{1}{2500}\sum_{i,j=1}^{50}\left(\frac{1}{20}\sum_{k=1}^{20} F_t^{1,k}(x_i,y_j)\right)^2 .$$

Note that four averages are involved in this formula: an average with respect to the space variable, an average over the 20 Monte-Carlo realizations, an average over replicas and a trajectorial average (the last two averages are more explicit in (3.2)). Notice that since the variance of the biasing force is smaller with PABF, one may expect better convergence in time results. This will be investigated in Section 3.2.3 and Section 3.2.4.

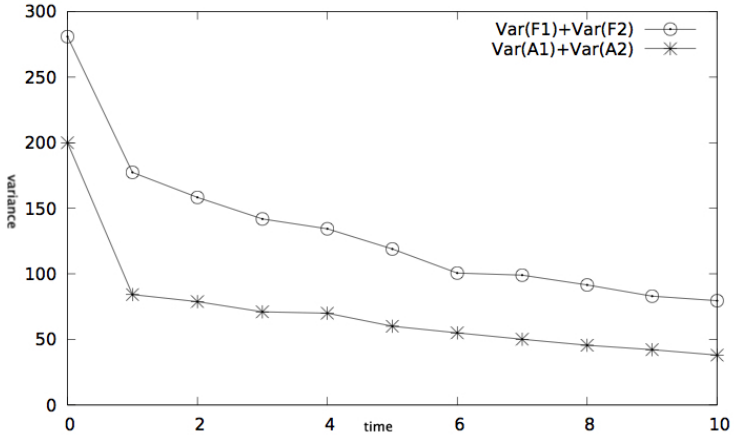

Fig. 5. Variances as a function of time.

3.2.3. *Free energy error*

We now present, the variation, as a function of time, of the normalized L^2- distance between the real free energy and the estimated one, in both cases: ABF and PABF methods. As can be seen in Figure 6, in both methods, the error decreases as time increases. Moreover, this error is always smaller for the projected ABF method than for the ABF method.

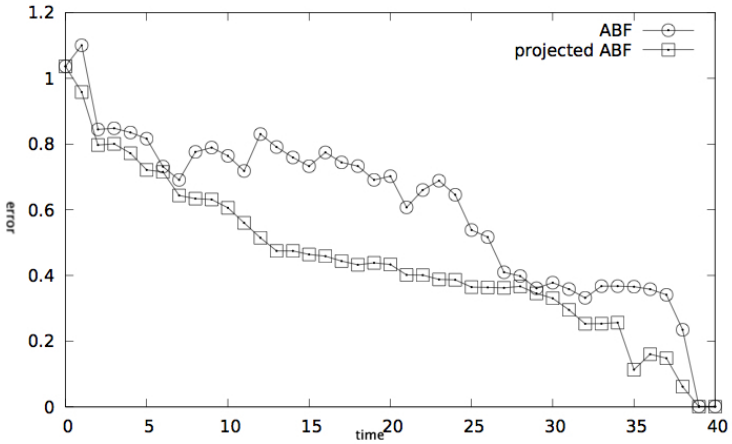

Fig. 6. Free energy error as a function of time.

3.2.4. *Distribution*

Another way to illustrate that the projected ABF method converges faster than the standard ABF method is to plot the density function ψ^{ξ} as a function of time (see Figure 7-11). It is illustrated that, as time increases, the probability of visiting all bins (of the reaction coordinate space $\mathcal{M}$) increases.

It is observed that, for the projected ABF method, the state where both bonds are stretched is visited earlier (at time 5) than for the standard ABF method (at time 20). The convergence to uniform law along (ξ_1, ξ_2) is faster with the projected ABF method.

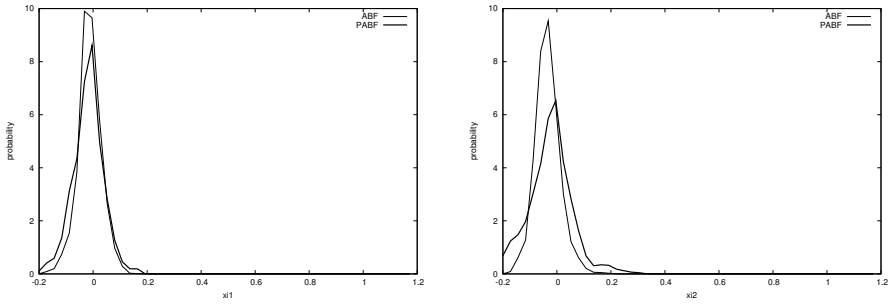

Fig. 7. At time 0.025. Left: $\int \psi^{\xi}(x_1, x_2)dx_2$; Right: $\int \psi^{\xi}(x_1, x_2)dx_1$.

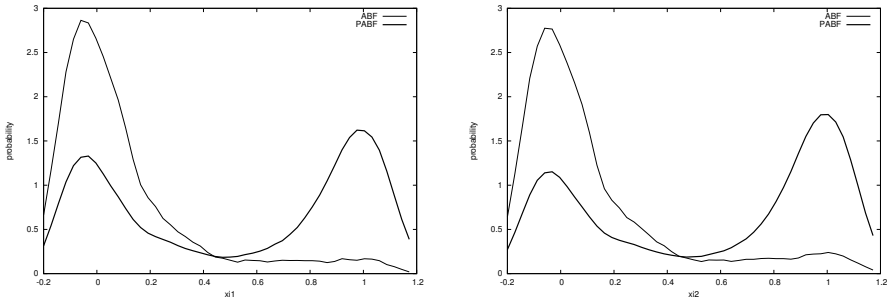

Fig. 8. At time 5. Left: $\int \psi^{\xi}(x_1, x_2)dx_2$; Right: $\int \psi^{\xi}(x_1, x_2)dx_1$.

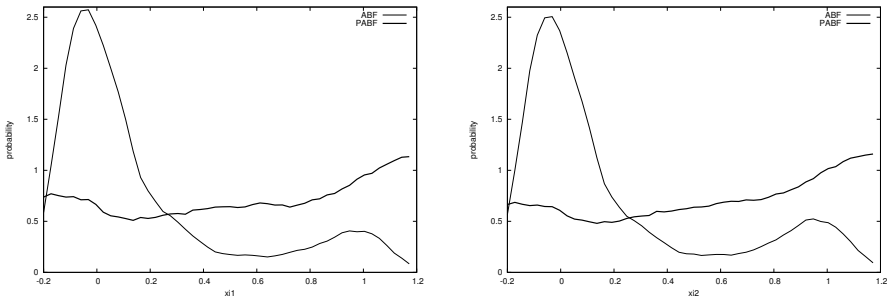

Fig. 9. At time 10. Left: $\int \psi^{\xi}(x_1, x_2)dx_2$; Right: $\int \psi^{\xi}(x_1, x_2)dx_1$.

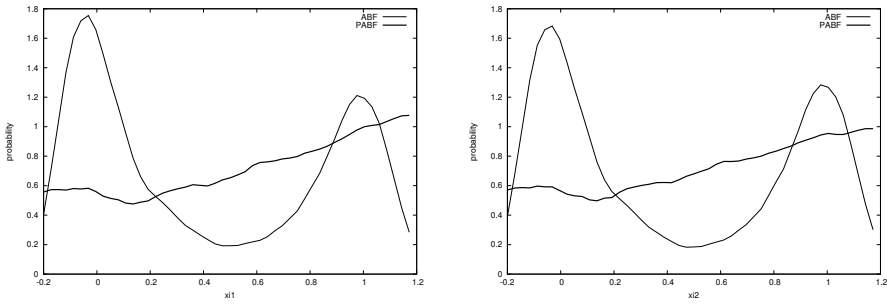

Fig. 10. At time 20. Left: $\int \psi^{\xi}(x_1, x_2)dx_2$; Right: $\int \psi^{\xi}(x_1, x_2)dx_1$.

4. Proofs

The proofs are inspired from [9]. One may assume that $\beta = 1$ up to the following change of variable: $\tilde{t} = \beta^{-1}t$, $\tilde{\psi}(\tilde{t}, x) = \psi(t, x)$, $\tilde{V}(x) = \beta V(x)$. Recall, that we work in $\mathcal{D} = \mathbb{T}^n$, $\mathcal{M} = \mathbb{T}^2$ and $\forall x = (x_1, ..., x_n) \in \mathbb{T}^n$, $\xi(x) = (x_1, x_2)$.

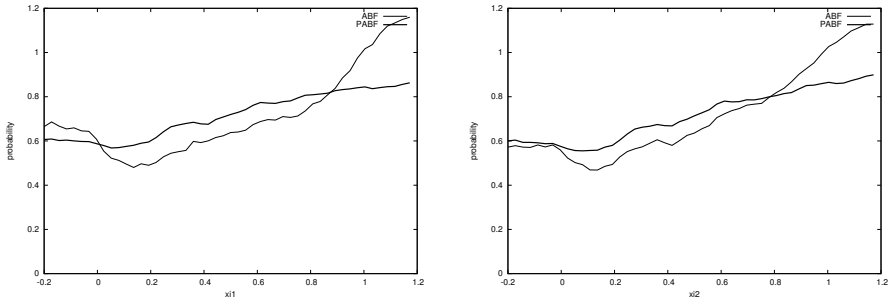

Fig. 11. At time 25. Left: $\int \psi^{\xi}(x_1,x_2)dx_2$; Right: $\int \psi^{\xi}(x_1,x_2)dx_1$.

4.1. *Proof of Proposition 2*

Let $g : \mathbb{T}^2 \to \mathbb{R}$ be a function in $H^1(\mathbb{T}^2)$.

$$\begin{aligned}
\frac{d}{dt}\int_{\mathbb{T}^2} \psi^{\xi} g dx_1 dx_2 &= \frac{d}{dt}\int_{\mathbb{T}^n} \psi g \circ \xi dx_1^n \\
&= \int_{\mathbb{T}^n} \mathrm{div}[(\nabla V - \sum_{i=1}^{2} \partial_i A_t \circ \xi \nabla \xi_i)\psi + \nabla\psi] g \circ \xi dx_1^n \\
&= -\int_{\mathbb{T}^n} \sum_{j=1}^{2} [(\nabla V - \sum_{i=1}^{2} \partial_i A_t \circ \xi \nabla \xi_i)\psi + \nabla\psi].\nabla\xi_j \partial_j g \circ \xi dx_1^n \\
&= -\sum_{i=1}^{2} \int_{\mathbb{T}^n} [(\nabla V.\nabla\xi_i \psi + \nabla\psi.\nabla\xi_i]\partial_i g \circ \xi dx_1^n \\
&\quad + \sum_{i=1}^{2} \int_{\mathbb{T}^n} \partial_i A_t \circ \xi \psi \partial_i g \circ \xi dx_1^n.
\end{aligned}$$

Applying Fubini, it holds:

$$\begin{aligned}
\frac{d}{dt}\int_{\mathbb{T}^2} \psi^{\xi} g dx_1 dx_2 &= -\sum_{i=1}^{2} \int_{\mathbb{T}^2} \int_{\Sigma_{(x_1,x_2)}} [\partial_i V \psi + \partial_i \psi] dx_3^n \partial_i g(x_1,x_2) dx_1 dx_2 \\
&\quad + \sum_{i=1}^{2} \int_{\mathbb{T}^2} \int_{\Sigma_{(x_1,x_2)}} \partial_i A_t(x_1,x_2)\psi dx_3^n \partial_i g(x_1,x_2) dx_1 dx_2 \\
&= -\sum_{i=1}^{2} \int_{\mathbb{T}^2} F_t^i \psi^{\xi} \partial_i g(x_1,x_2) dx_1 dx_2 - \sum_{i=1}^{2} \int_{\mathbb{T}^2} \partial_i \psi^{\xi} \partial_i g(x_1,x_2) dx_1 dx_2 \\
&\quad + \sum_{i=1}^{2} \int_{\mathbb{T}^2} \partial_i A_t(x_1,x_2)\psi^{\xi} \partial_i g(x_1,x_2) dx_1 dx_2 \\
&= \int_{\mathbb{T}^2} \Delta\psi^{\xi} g(x_1,x_2) dx_1 dx_2,
\end{aligned}$$

where we used (2.4) with $\varphi = \psi^\xi(t,.)$. This is the weak formulation of:

$$\partial_t \psi^\xi = \Delta \psi^\xi, \text{ on } [0,\infty[\times\mathbb{T}^2.$$

Remark 10. The reason why we consider the weighted Helmholtz decomposition (2.3) with $\varphi = \psi^\xi(t,.)$ in the PABF dynamics (2.9) instead of the standard one (1.14) is precisely to obtain this simple diffusion equation on the function ψ^ξ. This is will also be useful in the proof of Lemma 6 below.

4.2. *Proof of Corollary 1*

Let $\phi = \psi^\xi$ and $\phi_\infty = \psi^\xi_\infty = 1$. It is known that $\forall t \geq 0$ and $\forall (x_1,x_2) \in \mathbb{T}^2$, ϕ satisfies:

$$\partial_t \phi = \Delta_{x_1^2} \phi. \tag{4.1}$$

Moreover (See Remark 4), it is assumed that and is such that

$$\int_{\mathbb{T}^2} \phi(0,.) = 1 \,\text{and}\, \phi(0,.) \geq 0.$$

Let us show that $\forall\, t \geq 0, \forall k > 0$, $\|\phi(t,.)-1\|_{H^k(\mathbb{T}^2)} \leq \|\phi(0,.)-1\|_{H^k(\mathbb{T}^2)} e^{-8\pi^2 t}$.

First, we prove that ϕ converges to 1 in $L^2(\mathbb{T}^2)$,

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\int_{\mathbb{T}^2} |\phi-1|^2 &= \int_{\mathbb{T}^2} \partial_t\phi(\phi-1) \\
&= \int_{\mathbb{T}^2} \Delta\phi(\phi-1) \\
&= -\int_{\mathbb{T}^2} \nabla\phi\nabla(\phi-1) \\
&= -\int_{\mathbb{T}^2} |\nabla\phi|^2 \\
&\leq -4\pi^2 \int_{\mathbb{T}^2} |\phi-1|^2,
\end{aligned}$$

where we have used the Poincaré-Wirtinger inequality on the torus $\mathbb{T}^2$, applied to ϕ: for any function $f \in H^1(\mathbb{T}^2)$,

$$\int_{\mathbb{T}^2} \left(f - \int_{\mathbb{T}^2} f\right)^2 \leq \frac{1}{4\pi^2}\int_{\mathbb{T}^2} |\nabla f|^2.$$

We therefore obtain, $\|\phi(t,.)-1\|^2_{L^2(\mathbb{T}^2)} \leq \|\phi(0,.)-1\|^2_{L^2(\mathbb{T}^2)} e^{-8\pi^2 t}$.

Second, we prove that $\partial_i\phi$ converges to 0 in $L^2(\mathbb{T}^2)$. For $i = 1,2$, $\partial_i\phi$ satis-

fies (4.1): $\partial_t(\partial_i\phi) = \Delta_{x_1^2}(\partial_i\phi)$, with periodic boundary conditions. As above,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\int_{\mathbb{T}^2}|\partial_i\phi|^2 &= \int_{\mathbb{T}^2}\partial_t(\partial_i\phi)\partial_i\phi \\ &= \int_{\mathbb{T}^2}\Delta(\partial_i\phi)\partial_i\phi \\ &= -\int_{\mathbb{T}^2}|\nabla(\partial_i\phi)|^2.\end{aligned}$$

Using again Poincaré-Wirtinger inequality on $\partial_i\phi$,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\int_{\mathbb{T}^2}|\partial_i\phi|^2 &\leq -4\pi^2\int_{\mathbb{T}^2}\left(\partial_i\phi - \int_{\mathbb{T}^2}\partial_i\phi\right)^2 \\ &= -4\pi^2\int_{\mathbb{T}^2}|\partial_i\phi|^2.\end{aligned}$$

Where we used $\int_{\mathbb{T}^2}\partial_i\phi = 0$, since ϕ is periodic on $\mathbb{T}^2$. Therefore, it holds $\|\partial_i\phi(t,.)\|^2_{L^2(\mathbb{T}^2)} \leq \|\partial_i\phi(0,.)\|^2_{L^2(\mathbb{T}^2)}e^{-8\pi^2 t}$.

Third, one can prove by induction that all higher derivatives of ϕ converge exponentially fast to 0, with rate $8\pi^2$ and the following estimation is then proven:

$$\forall\, t \geq 0, \forall k > 0, \quad \|\phi(t,.) - 1\|^2_{H^k(\mathbb{T}^2)} \leq \|\phi(0,.) - 1\|^2_{H^k(\mathbb{T}^2)}e^{-8\pi^2 t}.$$

As $H^k(\mathbb{T}^2) \hookrightarrow L^\infty(\mathbb{T}^2)$, $\forall k > 1$, then $\exists c > 0$,

$$\|\phi - 1\|^2_{L^\infty} \leq c\|\phi - 1\|^2_{H^k} \leq ce^{-8\pi^2 t}.$$

Therefore, $\forall\varepsilon > 0$, $\exists t_0 > 0$, $\forall x \in \mathbb{T}^2$, $\forall t > t_0$, $\phi(t,x) \geq 1 - \varepsilon$. Finally, $\forall t > t_0$

$$I(\psi^\xi|\psi^\xi_\infty) = \int_{\mathbb{T}^2}\frac{|\nabla_{x_1^2}\phi|^2}{\phi} \leq \frac{1}{1-\varepsilon}\int_{\mathbb{T}^2}|\nabla_{x_1^2}\phi|^2 \leq \frac{\|\nabla_{x_1^2}\phi(0,.)\|^2_{L^2(\mathbb{T}^2)}}{1-\varepsilon}e^{-8\pi^2 t}.$$

4.3. *Proof of Corollary 2*

We have that $\psi^\xi_\infty = 1$ satisfies $LSI(r)$, for some $r > 0$ (see Chapter 3, Section 3 in [1]). Referring to Proposition 2 and since ψ^ξ is a probability density function, one gets:

$$\begin{aligned}\frac{d}{dt}E_M &= \int_{\mathbb{T}^2}\partial_t\left(\psi^\xi\ln(\psi^\xi)\right) \\ &= \int_{\mathbb{T}^2}\partial_t\psi^\xi\ln(\psi^\xi) + \int_{\mathbb{T}^2}\partial_t\psi^\xi \\ &= \int_{\mathbb{T}^2}\Delta\psi^\xi\ln(\psi^\xi) \\ &= -\int_{\mathbb{T}^2}|\nabla_{x_1^2}\ln(\psi^\xi)|^2\psi^\xi \\ &= -I(\psi^\xi|\psi^\xi_\infty) \\ &\leq -2rH(\psi^\xi|\psi^\xi_\infty) \\ &= -2rE_M.\end{aligned}$$

Therefore, E_M converges exponentially fast to zero. Referring to Corollary 1 and since E_M converges to zero, we have that for any $t > t_0$,

$$\begin{aligned} -E_M(t) = \int_t^\infty \frac{d}{ds} E_M(s) ds &= -\int_t^\infty I(\psi^\xi | \psi^\xi_\infty) ds \\ &\geq -I_0 \int_t^\infty \mathrm{e}^{-8\pi^2 s} ds \\ &= -\frac{I_0}{8\pi^2} \mathrm{e}^{-8\pi^2 t}, \end{aligned}$$

which yields the desired estimation.

4.4. *Proof of Theorem 1*

To prove our main result, several intermediate lemmas are needed.

Lemma 3. *$\forall t \geq 0$, $\forall (x_1, x_2) \in \mathbb{T}^2$ and for $i = 1, 2$, we have:*

$$(F^i_t - \partial_i A)(x_1, x_2) = \left(\int_{\Sigma_{(x_1,x_2)}} \partial_i \ln(\psi/\psi_\infty) \frac{\psi}{\psi^\xi} dx_3^n \right)(x_1, x_2) - \left(\partial_i \ln(\psi^\xi/\psi^\xi_\infty)\right)(x_1, x_2).$$

Proof.

$$\begin{aligned} &\int_{\Sigma_{(x_1,x_2)}} \partial_i \ln(\psi/\psi_\infty) \frac{\psi}{\psi^\xi} dx_3^n - \partial_i \ln(\psi^\xi/\psi^\xi_\infty) \\ &= \int_{\Sigma_{(x_1,x_2)}} \partial_i \ln(\psi) \frac{\psi}{\psi^\xi} dx_3^n - \int_{\Sigma_{(x_1,x_2)}} \partial_i \ln(\psi_\infty) \frac{\psi}{\psi^\xi} dx_3^n - \partial_i \ln(\psi^\xi) + \partial_i \ln(\psi^\xi_\infty) \\ &= \frac{1}{\psi^\xi} \int_{\Sigma_{(x_1,x_2)}} \partial_i \psi dx_3^n + \int_{\Sigma_{(x_1,x_2)}} (\partial_i V - \nabla \xi_i \partial_i A \circ \xi) \frac{\psi}{\psi^\xi} dx_3^n - \partial_i \ln(\psi^\xi) \\ &= \frac{\partial_i \psi^\xi}{\psi^\xi} + F^i_t - \partial_i A - \frac{\partial_i \psi^\xi}{\psi^\xi} \\ &= F^i_t - \partial_i A. \end{aligned}$$

□

Lemma 4. *Suppose that [H1] and [H2] hold, then for all $t \geq 0$, for all $(x_1, x_2) \in \mathbb{T}^2$ and for $i = 1, 2$, we have:*

$$|F^i_t(x_1, x_2) - \partial_i A(x_1, x_2)| \leq \gamma \sqrt{\frac{2}{\rho} e_m(t, x_1, x_2)}.$$

Proof. For any coupling measure $\pi \in \prod(\mu_{t,x_1,x_2}, \mu_{\infty,x_1,x_2})$ defined on $\Sigma_{(x_1,x_2)} \times \Sigma_{(x_1,x_2)}$, it holds:

$$\begin{aligned} |F^i_t - \partial_i A| &= \left| \int_{\Sigma_{(x_1,x_2)} \times \Sigma_{(x_1,x_2)}} (\partial_i V(x) - \partial_i V(x')) \pi(dx, dx') \right| \\ &= \|\nabla_{x_3^n} \partial_i V\|_{L^\infty} \sqrt{\int_{\Sigma_{(x_1,x_2)} \times \Sigma_{(x_1,x_2)}} d_{\Sigma_{(x_1,x_2)}}(x, x')^2 \pi(dx, dx')}. \end{aligned}$$

Taking now the infimum over all $\pi \in \prod(\mu(t,.|(x_1,x_2)), \mu(\infty,.|(x_1,x_2)))$ and using Lemma 1, we obtain

$$\begin{aligned}|F_t^i - \partial_i A| &\leq \gamma W(\mu(t,.|(x_1,x_2)), \mu^\xi(\infty,.|(x_1,x_2))) \\ &\leq \gamma\sqrt{\frac{2}{\rho}H(\mu^\xi(t,.|(x_1,x_2)), \mu^\xi(\infty,.|(x_1,x_2)))} \\ &= \gamma\sqrt{\frac{2}{\rho}e_m(t,(x_1,x_2))}.\end{aligned}$$

□

Lemma 5. *Suppose that [H2] holds, then for all $t \geq 0$,*

$$E_m(t) \leq \frac{1}{2\rho}\int_{\mathbb{T}^n} |\nabla_{x_3^n} \ln(\psi(t,.)/\psi_\infty)|^2 \psi.$$

Proof. Using [H2],

$$\begin{aligned}E_m &= \int_{\mathbb{T}^2} e_m \psi^\xi dx_1 dx_2 \\ &= \int_{\mathbb{T}^2} H(\mu(t,.|(x_1,x_2))|\mu(\infty,.|(x_1,x_2)))\psi^\xi dx_1 dx_2 \\ &\leq \int_{\mathbb{T}^2} \frac{1}{2\rho}\int_{\Sigma_{x_1^2}} |\nabla_{x_3^n} \ln(\psi(t,.)/\psi_\infty)|^2 dx_3^n \frac{\psi(t,.)}{\psi^\xi(t,x_1,x_2)} dx_1 dx_2 \\ &= \frac{1}{2\rho}\int_{\mathbb{T}^n} |\nabla_{x_3^n} \ln(\psi(t,.)/\psi_\infty)|^2 \psi dx_1^n.\end{aligned}$$

□

Lemma 6. *It holds for all $t \geq 0$,*

$$\int_{\mathbb{T}^n} (\partial_1 A_t - F_t^1)[\partial_1 \ln(\psi/\psi_\infty)]\psi + \int_{\mathbb{T}^n} (\partial_2 A_t - F_t^2)[\partial_2 \ln(\psi/\psi_\infty)]\psi \leq 0.$$

Proof. Using Fubini,

$$\begin{aligned}&\int_{\mathbb{T}^n} (\partial_1 A_t - F_t^1)[\partial_1 \ln(\psi/\psi_\infty)]\psi + \int_{\mathbb{T}^n} (\partial_2 A_t - F_t^2)[\partial_2 \ln(\psi/\psi_\infty)]\psi \\ &= \int_{\mathbb{T}^2} (\partial_1 A_t - F_t^1) \int_{\Sigma_{(x_1,x_2)}} [\partial_1 \ln(\psi/\psi_\infty)]\psi + \int_{\mathbb{T}^2} (\partial_2 A_t - F_t^2) \int_{\Sigma_{(x_1,x_2)}} [\partial_2 \ln(\psi/\psi_\infty)]\psi.\end{aligned}$$

For the first term, we have

$$\begin{aligned}\int_{\Sigma_{(x_1,x_2)}} [\partial_1 \ln(\psi/\psi_\infty)]\psi &= \int_{\Sigma_{(x_1,x_2)}} (\partial_1 \ln \psi)\psi - \int_{\Sigma_{(x_1,x_2)}} (\partial_1 \ln \psi_\infty)\psi \\ &= \partial_1 \psi^\xi + \int_{\Sigma_{(x_1,x_2)}} \partial_1(V - A)\psi \\ &= (\partial_1 \ln \psi^\xi)\psi^\xi + F_t^1 \psi^\xi - \partial_1 A \psi^\xi.\end{aligned}$$

Similarly, we have

$$\int_{\Sigma_{(x_1,x_2)}} [\partial_2 \ln(\psi/\psi_\infty)]\psi = (\partial_2 \ln \psi^\xi)\psi^\xi + F_t^2\psi^\xi - \partial_2 A\psi^\xi.$$

Therefore, one gets

$$\begin{aligned}
&\int_{\mathbb{T}^n} (\partial_1 A_t - F_t^1)[\partial_1 \ln(\psi/\psi_\infty)]\psi + \int_{\mathbb{T}^n} (\partial_2 A_t - F_t^2)[\partial_2 \ln(\psi/\psi_\infty)]\psi \\
&= \int_{\mathbb{T}^2} (\partial_1 A_t - F_t^1)(\partial_1 \ln \psi^\xi)\psi^\xi + \int_{\mathbb{T}^2} (\partial_2 A_t - F_t^2)(\partial_2 \ln \psi^\xi)\psi^\xi \\
&\quad - \int_{\mathbb{T}^2} (\partial_1 A_t - F_t^1)^2\psi^\xi - \int_{\mathbb{T}^2} (\partial_2 A_t - F_t^2)^2\psi^\xi
\end{aligned}$$

which concludes the assertion since the first line is equal to zero (by (2.4) with $\varphi = \psi^\xi(t,.)$) and the second line is non positive. Again the weighted Helmholtz decomposition helps in simplifying terms. □

Proof of Theorem 1:

Now we will prove the exponentially convergence of $E_m(t)$. Recall (2.13):

$$\partial_t \psi = \operatorname{div}(\nabla V \psi + \nabla \psi) - \partial_1((\partial_1 A_t)\psi) - \partial_2((\partial_2 A_t)\psi),$$

which is equivalent to

$$\partial_t \psi = \operatorname{div}\left(\psi_\infty \nabla\left(\frac{\psi}{\psi_\infty}\right)\right) + \partial_1[(\partial_1 A - \partial_1 A_t)\psi] + \partial_2[(\partial_2 A - \partial_2 A_t)\psi].$$

Using (2.17), (2.18) and (4.1), one obtains

$$\begin{aligned}
\frac{dE}{dt} &= -\int_{\mathbb{T}^n} |\nabla \ln(\psi/\psi_\infty)|^2\psi + \int_{\mathbb{T}^n} (\partial_1 A_t - \partial_1 A)[\partial_1 \ln(\psi/\psi_\infty)]\psi \\
&\quad + \int_{\mathbb{T}^n} (\partial_2 A_t - \partial_2 A)[\partial_2 \ln(\psi/\psi_\infty)]\psi, \\
\frac{dE_M}{dt} &= -\int_{\mathbb{T}^2} |\nabla \ln(\psi^\xi)|^2\psi^\xi.
\end{aligned}$$

Using then Lemma 2 and Lemma 3, one gets

$$\begin{aligned}
\frac{dE_m}{dt} &= \frac{dE}{dt} - \frac{dE_M}{dt} \\
&= -\int_{\mathbb{T}^n} |\nabla \ln(\psi/\psi_\infty)|^2 \psi + \int_{\mathbb{T}^n} (\partial_1 A_t - \partial_1 A)\partial_1 \ln(\psi/\psi_\infty)\psi \\
&\quad + \int_{\mathbb{T}^n} (\partial_2 A_t - \partial_2 A)\partial_2 \ln(\psi/\psi_\infty)\psi + \int_{\mathbb{T}^2} |\partial_1 \ln \psi^\xi|^2 \psi^\xi + \int_{\mathbb{T}^2} |\partial_2 \ln \psi^\xi|^2 \psi^\xi \\
&= -\int_{\mathbb{T}^n} |\nabla \ln(\psi/\psi_\infty)|^2 \psi \\
&\quad + \int_{\mathbb{T}^n} (\partial_1 A_t - F_t^1)[\partial_1 \ln(\psi/\psi_\infty)]\psi + \int_{\mathbb{T}^n} (F_t^1 - \partial_1 A)[\partial_1 \ln(\psi/\psi_\infty)]\psi \\
&\quad + \int_{\mathbb{T}^n} (\partial_2 A_t - F_t^2)[\partial_2 \ln(\psi/\psi_\infty)]\psi + \int_{\mathbb{T}^n} (F_t^2 - \partial_2 A)[\partial_2 \ln(\psi/\psi_\infty)]\psi \\
&\quad + \int_{\mathbb{T}^2} |\partial_1 \ln \psi^\xi|^2 \psi^\xi + \int_{\mathbb{T}^2} |\partial_2 \ln \psi^\xi|^2 \psi^\xi.
\end{aligned}$$

Lemma 6 then yields

$$\begin{aligned}
\frac{dE_m}{dt} \leq &-\int_{\mathbb{T}^n} |\nabla \ln(\psi/\psi_\infty)|^2 \psi \\
&+ \int_{\mathbb{T}^n} (F_t^1 - \partial_1 A)\partial_1 \ln(\psi/\psi_\infty)\psi + \int_{\mathbb{T}^n} (F_t^2 - \partial_2 A)\partial_2 \ln(\psi/\psi_\infty)\psi \\
&+ \int_{\mathbb{T}^2} |\partial_1 \ln \psi^\xi|^2 \psi^\xi + \int_{\mathbb{T}^2} |\partial_2 \ln \psi^\xi|^2 \psi^\xi.
\end{aligned}$$

Using lemma 3 and Fubini, one then obtains

$$\begin{aligned}
\frac{dE_m}{dt} \leq &-\int_{\mathbb{T}^n} |\nabla \ln(\psi/\psi_\infty)|^2 \psi \\
&+ \int_{\mathbb{T}^2} \left[\int_{\Sigma_{(x_1,x_2)}} \partial_1 \ln(\psi/\psi_\infty)\frac{\psi}{\psi^\xi}\right] \int_{\Sigma_{(x_1,x_2)}} \partial_1 \ln(\psi/\psi_\infty)\psi - \int_{\mathbb{T}^n} \partial_1 \ln(\psi^\xi)\partial_1 \ln(\psi/\psi_\infty)\psi \\
&+ \int_{\mathbb{T}^2} \left[\int_{\Sigma_{(x_1,x_2)}} \partial_2 \ln(\psi/\psi_\infty)\frac{\psi}{\psi^\xi}\right] \int_{\Sigma_{(x_1,x_2)}} \partial_2 \ln(\psi/\psi_\infty)\psi - \int_{\mathbb{T}^n} \partial_2 \ln(\psi^\xi)\partial_2 \ln(\psi/\psi_\infty)\psi \\
&+ \int_{\mathbb{T}^2} |\partial_1 \ln \psi^\xi|^2 \psi^\xi + \int_{\mathbb{T}^2} |\partial_2 \ln \psi^\xi|^2 \psi^\xi \\
\leq &-\int_{\mathbb{T}^n} |\nabla \ln(\psi/\psi_\infty)|^2 \psi \\
&+ \int_{\mathbb{T}^2} \left[\int_{\Sigma_{(x_1,x_2)}} \partial_1 \ln(\psi/\psi_\infty)\psi\right]^2 \frac{1}{\psi^\xi} - \int_{\mathbb{T}^n} \partial_1 \ln(\psi^\xi)\partial_1 \ln(\psi/\psi_\infty)\psi \\
&+ \int_{\mathbb{T}^2} \left[\int_{\Sigma_{(x_1,x_2)}} \partial_2 \ln(\psi/\psi_\infty)\psi\right]^2 \frac{1}{\psi^\xi} - \int_{\mathbb{T}^n} \partial_2 \ln(\psi^\xi)\partial_2 \ln(\psi/\psi_\infty)\psi \\
&+ \int_{\mathbb{T}^2} |\partial_1 \ln \psi^\xi|^2 \psi^\xi + \int_{\mathbb{T}^2} |\partial_2 \ln \psi^\xi|^2 \psi^\xi.
\end{aligned}$$

Applying Cauchy-Schwarz on the first terms of the second and third lines, we obtain

$$\begin{aligned}\frac{dE_m}{dt} \leq & -\int_{\mathbb{T}^n} |\nabla_{x_3^n} \ln(\psi/\psi_\infty)|^2 \psi \\ & - \int_{\mathbb{T}^n} \partial_1 \ln(\psi^\xi) \partial_1 \ln(\psi/\psi_\infty)\psi - \int_{\mathbb{T}^n} \partial_2 \ln(\psi^\xi)\partial_2 \ln(\psi/\psi_\infty)\psi \\ & + \int_{\mathbb{T}^2} |\partial_1 \ln \psi^\xi|^2 \psi^\xi + \int_{\mathbb{T}^2} |\partial_2 \ln \psi^\xi|^2 \psi^\xi \\ \leq & -\int_{\mathbb{T}^n} |\nabla_{x_3^n} \ln(\psi/\psi_\infty)|^2 \psi - \int_{\mathbb{T}^2} \partial_1 \ln(\psi^\xi) \left[\int_{\Sigma_{(x_1,x_2)}} \partial_1 \ln(\psi/\psi_\infty) \frac{\psi}{\psi^\xi} - \partial_1 \ln \psi^\xi \right] \psi^\xi \\ & - \int_{\mathbb{T}^2} \partial_2 \ln(\psi^\xi) \left[\int_{\Sigma_{(x_1,x_2)}} \partial_2 \ln(\psi/\psi_\infty) \frac{\psi}{\psi^\xi} - \partial_2 \ln \psi^\xi \right] \psi^\xi.\end{aligned}$$

Applying Lemma 5, Lemma 3, Cauchy-Schwarz, Lemma 4 and Corollary 1,

$$\begin{aligned}\frac{dE_m}{dt} &\leq -2\rho E_m + \sqrt{\int_{\mathbb{T}^2} |\partial_1 \ln(\psi^\xi)|^2 \psi^\xi} \sqrt{\int_{\mathbb{T}^2} \frac{2}{\rho} e_m(t,(x_1,x_2))\psi^\xi} \\ &\quad + \sqrt{\int_{\mathbb{T}^2} |\partial_2 \ln(\psi^\xi)|^2 \psi^\xi} \sqrt{\int_{\mathbb{T}^2} \frac{2}{\rho} e_m(t,(x_1,x_2))\psi^\xi} \\ &\leq -2\rho E_m + 2\gamma \sqrt{\frac{2}{\rho} E_m} \sqrt{\int_{\mathbb{T}^2} |\nabla_{x_1^2} \ln(\psi^\xi)|^2 \psi^\xi} \\ &\leq -2\rho E_m + 2\gamma \sqrt{\frac{2}{\rho} E_m} \sqrt{I(\psi^\xi/\psi^\xi_\infty)} \\ &\leq -2\rho E_m + 2\gamma \sqrt{\frac{2}{\rho} E_m} \sqrt{I_0} e^{-4\pi^2 t}.\end{aligned}$$

Finally we obtain

$$\frac{d}{dt}\sqrt{E_m(t)} \leq -\rho \sqrt{E_m(t)} + \gamma \sqrt{\frac{I_0}{2\rho}} e^{-4\pi^2 t}.$$

First, if $\rho \neq 4\pi^2$, using Gronwall inequality, one obtains

$$\begin{aligned}\sqrt{E_m(t)} &\leq \sqrt{E_m(0)}\, e^{-\rho t} + \gamma \sqrt{\frac{I_0}{2\rho}} \int_0^t e^{\rho(-t+s)} e^{-4\pi^2 s} ds \\ &\leq \sqrt{E_m(0)}\, e^{-\rho t} + \gamma \sqrt{\frac{I_0}{2\rho}} \frac{e^{-\rho t}}{\rho - 4\pi^2} \left(e^{(\rho - 4\pi^2)t} - 1 \right) \\ &\leq \sqrt{E_m(0)}\, e^{-\rho t} + \gamma \sqrt{\frac{I_0}{2\rho}} \frac{e^{-\rho t}}{|\rho - 4\pi^2|} e^{(\rho - 4\pi^2)t}.\end{aligned}$$

$$\begin{aligned}\sqrt{E_m(t)} &\leq \sqrt{E_m(0)}\, e^{-\rho t} + \gamma\sqrt{\frac{I_0}{2\rho}}\int_0^t e^{-\rho t} ds \\ &\leq \left(\sqrt{E_m(0)} + \gamma\sqrt{\frac{I_0}{2\rho}}t\right) e^{-\rho t},\end{aligned}$$

which leads the desired estimation (2.21).

Using this above convergence, Corollary 2 and Lemma 2, it is then easy to see that E converges exponentially fast to zero. Using (2.15), one obtains the convergence of ψ to ψ_∞ since:

$$\|\psi - \psi_\infty\|_{L^1(\mathbb{T}^n)} \leq \sqrt{2H(\psi|\psi_\infty)} = \sqrt{2E}.$$

The second point of the theorem is checked. Finally, we are now in position to prove the last point of Theorem 1. Using (2.6) and Lemma 4,

$$\begin{aligned}\|\nabla A_t - \nabla A\|^2_{L^2_{\psi\xi}(\mathbb{T}^2)} &\leq 2\|\nabla A_t - F_t\|^2_{L^2_{\psi\xi}(\mathbb{T}^2)} + 2\|F_t - \nabla A\|^2_{L^2_{\psi\xi}(\mathbb{T}^2)} \\ &\leq 4\|F_t - \nabla A\|^2_{L^2_{\psi\xi}(\mathbb{T}^2)} \\ &\leq 8\frac{\gamma^2}{\rho}E_m.\end{aligned}$$

Acknowledgment

The authors thank Gabriel Stoltz for very helpful discussions on the numerical experiments. Houssam Alrachid would like to thank the Ecole des Ponts ParisTech and CNRS Libanais for supporting his PHD thesis. The work of Tony Lelièvre is supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement number 614492.

References

1. C. Ané, S. Blachère, D. Chafaï, P. Fougères, I Gentil , C Roberto and G Scheffer, Sur les inégalités de Sobolev logarithmiques, *Société Mathématiques de France (in French)* (2000).
2. A. Arnold, P. Markowich, G. Toscani and A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations, *Commun. Partial Diff. Eqns* **26** (2001) 43–100.
3. G. Bussi, A. Laio, and M. Parinello, Equilibrium free energies from nonequilibrium metadynamics, *Phys. Rev. Lett. 96*, (2006), 090601.
4. C. Chipot and A. Pohorille (ed), Free energy calculations, *Springer Series in Chemical Physics vol 86(Berlin: Springer)* (2007).
5. E. Darve and A. Pohorille, Calculating free energy using average forces, *J. Chem. Phys.*, **115** (2001) 9169-83.
6. V. Girault and P. Raviart, Finite element methods for Navier-Stokes equations: theory and algorithms, *Springer-Verlag* (1986).
7. J. Hénin and C. Chipot, Overcoming free energy barriers using unconstrained molecular dynamics simulations, *J. Chem. Phys*, **121** (2004) 2904.

8. T. Lelièvre, M. Rousset and G. Stoltz, Computation of free energy profiles with adaptive parallel dynamics, *J. Chem. Phys. 126*, **21** (2007b) 134111.
9. T. Lelièvre, M. Rousset and G. Stoltz, Long-time convergence of an adaptive biasing force method, *Nonlinearity*, **21** (2008) 1155–1181.
10. T. Lelièvre, M. Rousset and G. Stoltz, Free energy computations: A mathematical perspective, *Imperial College Press*, (2010).
11. F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, *J. Funct. Anal.*, **173** (2000) 361-400.
12. D. C. Rapaport, The Art of Molecular Dynamics Simulation, *Cambridge university press*, (1997).
13. C. Villani, Topics in Optimal Transportation, *Graduate Studies in Mathematics vol 58*, (2003).
14. F. Wang and D. P. Landau, Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States, *Phys. Rev. Lett. (American Physical Society)*, (2001).