Adaptive force biasing algorithms: new convergence results and tensor approximations of the bias

Virginie Ehrlacher, Tony Lelièvre, Pierre Monmarché

Adaptive force biasing algorithms: new convergence results and tensor approximations of the bias

Virginie Ehrlacher[1], Tony Lelièvre[1], and Pierre Monmarché[2]

[1] *Université Paris-Est - CERMICS (ENPC) - INRIA*
[2] *Sorbonne Université -LJLL - LCT*

November 19, 2021

Abstract

We analyze and propose variants of the Adaptive Biasing Force method. First, we prove the convergence of a version of the algorithm where the biasing force is estimated using a weighted occupation measure, with an explicit asymptotic variance. Second, we propose a new flavour of the algorithm adapted to high dimensional reaction coordinates, for which the standard approaches suffer from the curse of dimensionality. More precisely, the free energy is approximated by a sum of tensor products of one-dimensional functions. The consistency of the tensor approximation is established. Numerical experiments on 5-dimensional reaction coordinates demonstrate that the method is indeed able to capture correlations between them.

keywords: Monte Carlo methods ; tensor ; free energy ; importance sampling ; molecular dynamics.
MSC class (2010): 65C05 ; 65N12.

1 Introduction

Consider $x \in \mathbb{T}^D$ a vector representing the positions of particles with periodic boundary conditions ($\mathbb{T} = \mathbb{R}/\mathbb{Z}$), and a potential energy $V \in \mathcal{C}^\infty\left(\mathbb{T}^D\right)$. Our first motivation is the computation, in multimodal cases, of expectations of the form

$$\frac{1}{\int_{\mathbb{T}^D} e^{-\beta V(x)} \mathrm{d}x} \int_{\mathbb{T}^D} \varphi(x) e^{-\beta V(x)} \mathrm{d}x \quad =: \quad \int_{\mathbb{T}^D} \varphi \mathrm{d}\mu_{V,\beta} \tag{1}$$

where $\varphi : \mathbb{T}^D \to \mathbb{R}$ is called an observable and $\mathrm{d}\mu_{V,\beta} := e^{-\beta V(x)} \mathrm{d}x$ is the Gibbs law with potential V and inverse temperature $\beta > 0$. Our second motivation is the estimation of the free energy associated to this Gibbs law and a given high-dimensional reaction coordinate, as will be explained below. Our main contributions are, first, the long-time convergence of an adaptive Importance Sampling method for estimating (1) (Theorem 2) with an explicit asymptotic variance (Theorem 3) and, second, the definition of a variant of this algorithm based on a tensor approximation (Algorithms 2 and 3) to tackle high-dimensional reaction coordinates and the proof of its consistency (Theorem 5).

The large dimension D is so significant that, in practice, the quantities (1) have to be computed with Markov Chain Monte Carlo (MCMC) algorithms, which consist in approximating

the average of φ with respect to $\mu_{V,\beta}$ along dynamics that are ergodic with respect to $\mu_{V,\beta}$. A typical sampler is the overdamped Langevin dynamics

$$\mathrm{d}X_t \;=\; -\nabla V(X_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t$$

where $(B_t)_{t\geqslant 0}$ is a Brownian motion over $\mathbb{T}^D$. It is ergodic with invariant measure $\mu_{V,\beta}$, so that

$$\frac{1}{t}\int_0^t \varphi(X_s)\mathrm{d}s \;\underset{t\to\infty}{\longrightarrow}\; \int_{\mathbb{T}^D}\varphi \mathrm{d}\mu_{V,\beta} \qquad a.s.$$

for all measurable bounded φ, see e.g. [37] and references therein. Nevertheless, the convergence of the process (or, in practice, of any alternative Markov process with invariant measure $\mu_{V,\beta}$) toward its equilibrium in the long-time limit may be very slow. This is due to the so-called metastability phenomenon, according to which the process remains for long times in some region of the space, with very rare transitions from one of these metastable regions to another. This is related to the multi-modality of the Gibbs measure and the fact MCMC algorithms typically perform local moves, so that leaving a mode of the target measure $\mu_{V,\beta}$ is a rare event. We refer to [34] for more details on this topic. For this reason, several adaptive methods have been developed in order to force the process to leave the metastable traps faster. Among those, we focus on the adaptive biasing force (ABF) algorithm, which may be seen as a particular Importance Sampling method. The general idea is to run a biased process

$$\mathrm{d}\tilde{X}_t \;=\; -\nabla V(\tilde{X}_t)\mathrm{d}t + \nabla V_{bias,t}(\tilde{X}_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t \tag{2}$$

where the biasing potential $V_{bias,t}$ is adaptively constructed from the past trajectory $(\tilde{X}_s)_{s\in[0,t]}$ in such a way that it is expected to converge to some $V_{bias,\infty}$. Expectations with respect to $\mu_{V,\beta}$ are then recovered through a reweighting step, assuming that ergodicity still holds:

$$\frac{\frac{1}{t}\int_0^t \varphi(\tilde{X}_s)e^{-\beta V_{bias,s}(\tilde{X}_s)}\mathrm{d}s}{\frac{1}{t}\int_0^t e^{-\beta V_{bias,s}(\tilde{X}_s)}\mathrm{d}s} \;\underset{t\to\infty}{\longrightarrow}\; \frac{\int_{\mathbb{T}^D}\varphi e^{-\beta V_{bias,\infty}}\mathrm{d}\mu_{V-V_{bias,\infty},\beta}}{\int_{\mathbb{T}^D} e^{-\beta V_{bias,\infty}}\mathrm{d}\mu_{V-V_{bias,\infty},\beta}} \;=\; \int_{\mathbb{T}^D}\varphi\mathrm{d}\mu_{V,\beta}\,. \tag{3}$$

Classically, in such an Importance Sampling scheme, the aim is to design a target bias $V_{bias,\infty}$ such that two conditions are met: 1) sampling the biased equilibrium $\mu_{V-V_{bias,\infty},\beta}$ is simpler than the initial problem (i.e. the corresponding overdamped Langevin process is less metastable) and 2) the biased equilibrium is not too far from the initial target so that the exponential weights in (3) do not cause the asymptotical variance of the estimator to skyrocket.

In the ABF algorithm, this issue is addressed with the use of so-called reaction coordinates (or collective variables) and the associated free energy as a bias. Reaction coordinates consist of a small number $d \ll D$ of macroscopic coordinates of the whole microscopic system $x \in \mathbb{T}^D$. These coordinates are defined through a map $\xi : \mathbb{T}^D \to \mathcal{M}$ where $\mathcal{M}$ is a manifold of dimension d. In molecular dynamics, for example, $x \in \mathbb{T}^D$ is a vector which gathers the positions of all the different atoms of the system of interest, and $\xi(x)$ typically represents some distances between particular pairs of atoms, or angles formed by some triplets of atoms. These reaction coordinates should be chosen to capture the main causes of the metastability of the system. More precisely, $\xi(X_t)$ should converge to equilibrium as slowly as X_t, while the conditional laws $\mathcal{L}(X \mid \xi(X) = z)$ for fixed $z \in \mathcal{M}$ when $X \sim \mu_{V,\beta}$ should be easier to sample (see [35] or Section 2.4 for more detailed considerations). In other words, $\xi(x)$ should be a low-dimensional representation of x that captures the slow variables of the system.

To these reaction coordinates ξ is associated the corresponding free energy $A : \mathcal{M} \to \mathbb{R}$, given by

$$A(z) \;=\; -\frac{1}{\beta}\ln \int_{\{x\in\mathbb{T}^D,\, \xi(x)=z\}} e^{-\beta V(x)}\delta_{\xi(x)-z}(\mathrm{d}x)\,,$$

where $\delta_{\xi(x)-z}$ is the so-called delta measure, which can be defined from the Lebesgue measure on the submanifold $\{x \in \mathbb{T}^D,\ \xi(x) = z\}$ through the co-area formula, see for example [36, Section 3.2.1]. This definition ensures that, if X is a random variable with law $\mu_{V,\beta}$ on $\mathbb{T}^D$, then $\xi(X)$ is a random variable with law $\mu_{A,\beta}$ on $\mathcal{M}$. The heuristic of the ABF algorithm is the following. Suppose that $\mathcal{M}$ is compact. If we were to sample from the process

$$\mathrm{d}Y_t = -\nabla \left(V - A \circ \xi\right)(Y_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t, \tag{4}$$

the equilibrium would be $\mu_{V-A\circ\xi,\beta}$, whose image through ξ, by definition of A, is the uniform measure on $\mathcal{M}$. This means that there would be no more metastability along ξ, since all the regions of $\mathcal{M}$ would be equally visited by $\xi(Y_t)$. Unfortunately, it is not possible to use directly this free-energy biased dynamics in practice, since it would require the knowledge of A and thus the computation of expectations in large dimension. The idea of the ABF method is to learn A on the fly, i.e. to run a process $(\tilde{X}_t)_{t\geqslant 0}$ solving (2) with a biasing potential $V_{bias,t}$ constructed from $(\tilde{X}_s)_{s\in[0,t]}$ and designed to target $A \circ \xi$ in the longtime limit.

As mentioned at the beginning of this introduction, in fact, independently from the question of obtaining MCMC estimations for (1), estimating the free energy A associated to some reaction coordinates ξ can be an objective *per se*. Indeed, in molecular dynamics (for either chemistry, biophysics or material science), it is a key quantity to analyze phenomena such as conformational stability, solvation properties or ligand binding affinity, see e.g. [24, 36, 28, 5] and references within. In these cases, the reaction coordinates work as a small dimensional representation of the macroscopic state of the system. This is also the case in coarse-graining problems, where the goal is to replace numerically intensive high-dimensional molecular simulations of a microscopic system X_t by cheaper simulations of an approximate effective dynamics for the low-dimensional $\xi(X_t)$. In that case, the effective force is usually given by the free energy, which has to be estimated on an all-atom high dimensional simulation, see e.g. [41, 43] and references within.

In practice, the choice of good reaction coordinates, both in term of enhanced sampling and of low-dimensional representation of the system, is a difficult problem. Up to recently, their definition has been based on the knowledge and intuition of specialists. The question of the automatic learning of suitable reaction coordinates is currently a vivid research area, see for instance [10, 11] and the recent review [22]. Moreover, some techniques like the orthogonal space random walk [40] provide a general way to construct new reaction coordinates from previous ones. Due to these recent progresses, one would like to consider a relatively large d. In ABF, $V_{bias,t}$ is a function of the d reaction coordinates. From a numerical point of view, since $V_{bias,t}$ is adaptively learned on the fly, its values have to be kept in memory, which requires a grid whose size typically scales exponentially with d. This limits the application of ABF to small dimensional reaction coordinates ($d \leqslant 4$). The aim of the present work is to lift this limitation by approximating $V_{bias,t}$ using a sum of tensor products of one-dimensional functions, which reduces the size of the memory to $\mathcal{O}(dm)$ where m is the number of tensor terms. Remark that this can in turn help for the definition of good reaction coordinates, by considering as candidates a relatively large number of reaction coordinates and then conduct a statistical study to select or combine some of them. A basic idea would be to conduct a sensitivity analysis of the free energy, computing for instance for each reaction coordinate ξ_i the best approximation in the least square sense of the (estimated) free energy by a function only of the other reaction coordinates $(\xi_j)_{j\neq i}$ (which is easily done for a function given as a sum of tensor products, see [31]) and then discarding the reaction coordinate whose disparition gives the lowest error. Nevertheless, this question exceeds the scope of the present work, in which ξ is supposed to be given.

Note that the question of increasing the number of reaction coordinates in adaptive biasing

algorithms has also been considered in the Bias-Exchange algorithm introduced in [48], where several replicas of the system are run in parallel, each associated with a one-dimensional reaction coordinate. The replicas exchange their bias according to some Metropolis-Hastings probability, so that each replica eventually feels the bias in all the different directions of the reaction coordinates. Nevertheless, in this case where one-dimensional reaction coordinates are treated independently one from the others, correlations between reaction coordinates cannot be measured (the same goes for the generalized ABF introduced in [51]), contrary to the algorithm introduced in the present work.

Besides, let us mention that numerical methods involving both tensor approximation and Monte Carlo methods for molecular dynamics are also introduced in [30, 45] for other purposes.

The rest of the paper is organized as follows. The algorithms are introduced in Section 2, where our main results are stated and discussed. Section 3 is devoted to the proofs of Theorems 2 and 3 which are related to the long-time convergence of the ABF algorithm. In Section 4, we prove Theorem 5, which concerns the tensor approximation scheme. Section 5 provides a detailed discussion on practical considerations and possible variations of the algorithm. Finally, some numerical experiments with the tensor ABF algorithm are reported in Section 6.

2 Algorithms and results

In this section we provide a presentation of the ABF algorithm which is considered in this work in a simple framework, and refer to Section 5 for generalizations. The presentation is divided into two parts. In Section 2.1, we present the reference ABF algorithm that we consider, without the tensor-product approximation, and state a long-time convergence result and a CLT. In Section 2.2, we introduce the tensor-product approximation of the bias and state its consistency. These two ingredients are then combined to yield the Tensor-ABF algorithm in Section 2.3. The two algorithms and associated convergence proofs of the reference ABF algorithm and of the tensor-product approximation are presented separately since they have their own interest.

2.1 Free energy and the ABF algorithm

Let us first present the ABF algorithm in a simple framework (see [26, 17, 35] for more general settings). From now on, we write

$$\mu \;=\; \mu_{V,\beta}\,,$$

seen both as a probability law and as the density of the latter with respect to the Lebesgue measure.

Let us assume that $\mathcal{M} = \mathbb{T}^d$ and that, for all $x = (q, z) \in \mathbb{T}^D = \mathbb{T}^p \times \mathbb{T}^d$, $\xi(x) = \xi(q, z) = z$ where $p = D - d$.

At first sight, this may seem a very restrictive choice of reaction coordinates. But, using extended variables (see [21]), this can actually be applied in very general contexts. We refer the reader to Section 5 for more details on this point.

The associated free energy for $z \in \mathbb{T}^d$ is then

$$A(z) \;=\; -\frac{1}{\beta}\ln\int_{\mathbb{T}^p} e^{-\beta V(q,z)}\mathrm{d}q\,.$$

Following the previous discussion, our aim is then to define for all time $t \geqslant 0$ a function A_t on $\mathbb{T}^d$ and to sample the process

$$\left\{ \begin{array}{rcl} \mathrm{d}Q_t &=& -\nabla_q V(Q_t, Z_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t^1 \\ \mathrm{d}Z_t &=& -\nabla_z V(Q_t, Z_t)\mathrm{d}t + \nabla_z A_t(Z_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t^2 \,, \end{array} \right. \tag{5}$$

where B^1 and B^2 are independent Brownian motions respectively of dimension p and d, in such a way that A_t gets close to A in large time.

Note that the free energy A satisfies

$$\nabla_z A(z) \;=\; \frac{\int_{\mathbb{T}^p} \nabla_z V(q,z) e^{-\beta V(q,z)}\mathrm{d}q}{\int_{\mathbb{T}^p} e^{-\beta V(q,z)}\mathrm{d}q} \;=\; \mathbb{E}_\mu\left[\nabla_z V(Q,Z) \mid Z = z\right].$$

The following alternative equivalent characterization of A will be useful in the sequel. Denoting by $H^1(\mathbb{T})$ the set of functions of $L^2(\mathbb{T}^d)$ with a weak gradient in $L^2(\mathbb{T}^d)$, define

$$H \;:=\; \left\{ f \in H^1(\mathbb{T}^d) \,:\; \int_{\mathbb{T}^d} f(z)\mathrm{d}z = 0 \right\}, \tag{6}$$

and let us denote by $\mathcal{P}(\mathbb{T}^p \times \mathbb{T}^d)$ the set of probability measures on $\mathbb{T}^p \times \mathbb{T}^d = \mathbb{T}^D$. For all $\nu \in \mathcal{P}(\mathbb{T}^p \times \mathbb{T}^d)$ and $f \in H$, let us define

$$\mathcal{E}_\nu(f) := \int_{\mathbb{T}^p \times \mathbb{T}^d} |\nabla_z V(q,z) - \nabla_z f(z)|^2 \mathrm{d}\nu(q,z).$$

As detailed in [1], up to an additive constant (like the potential V, the free energy is in fact always defined up to an additive constant), A is the unique minimizer in H of the functional $\mathcal{E}_\mu$, i.e.

$$A - \int_{\mathbb{T}^d} A(z)\mathrm{d}z = \operatorname*{argmin}_{f \in H} \mathcal{E}_\mu(f). \tag{7}$$

At time $t \geqslant 0$, a trajectory $(Q_s, Z_s)_{s\in[0,t]}$ of (5) is available. Let ν_t be the probability measure on $\mathbb{T}^p \times \mathbb{T}^d = \mathbb{T}^D$ defined as follows: for all $\varphi \in \mathcal{C}(\mathbb{T}^p \times \mathbb{T}^d)$,

$$\int_{\mathbb{T}^p \times \mathbb{T}^d} \varphi \mathrm{d}\nu_t \;=\; \left(\int_0^t e^{-\beta A_s(Z_s)}\mathrm{d}s \right)^{-1} \int_0^t \varphi\left(Q_s, Z_s\right) e^{-\beta A_s(Z_s)}\mathrm{d}s \,. \tag{8}$$

We call ν_t the unbiased occupation distribution of the process. By the ergodic limit (3), ν_t is expected to converge weakly to μ as t goes to infinity almost surely (at least if A_t does not change too fast with t).

However, note that ν_t is a singular probability measure, so that the minimization problem

$$\inf_{f \in H} \mathcal{E}_{\nu_t}(f)$$

is ill-posed. To circumvent this difficulty, one may consider two different alternatives to regularize the problem which we detail hereafter. Consider a smooth symmetric positive density kernel $K \in \mathcal{C}^\infty(\mathbb{T}^d \times \mathbb{T}^d, \mathbb{R}_+)$ with

$$\int_{\mathbb{T}^d} K(y,z)\mathrm{d}z = 1 \qquad \text{and} \qquad K(y,z) = K(z,y) \qquad \forall y, z \in \mathbb{T}^d \,. \tag{9}$$

In practice, $K(y,\cdot)$ should be close to a Dirac mass at y (see Theorem 2 below). For instance, a possible choice for K would be the so-called von-Mises kernel for a given small parameter $\varepsilon > 0$, i.e.

$$K(y,z) \;\propto\; \prod_{i=1}^{d} \exp\left(-\frac{1}{\varepsilon^2/2}\sin^2\left(\frac{z_i - y_i}{2}\right)\right). \tag{10}$$

Now, consider also a regularization parameter $\lambda \geqslant 0$. For all $\nu \in \mathcal{P}(\mathbb{T}^p \times \mathbb{T}^d)$ and all $f \in H$, we define

$$\mathcal{J}_\nu(f) \;:=\; \int_{\mathbb{T}^p\times\mathbb{T}^d\times\mathbb{T}^d} |\nabla_y V(q,y) - \nabla_z f(z)|^2 K(y,z)\mathrm{d}z\mathrm{d}\nu(q,y) + \lambda \int_{\mathbb{T}^d} |\nabla_z f(z)|^2 \mathrm{d}z\,, \tag{11}$$

Note that, as $K(y,\cdot)$ converges weakly toward the Dirac mass at y and λ goes to 0, for all $f \in H$, $\mathcal{J}_\nu(f)$ converges towards $\mathcal{E}_\nu(f)$. The interest of introducing $\mathcal{J}_\nu$ is that, thanks to the regularization, the minimization problem is now well-posed:

Proposition 1. *Assume that either $K > 0$ on $\mathbb{T}^d \times \mathbb{T}^d$ or $\lambda > 0$. Then, for all $\nu \in \mathcal{P}(\mathbb{T}^p \times \mathbb{T}^d)$, $\mathcal{J}_\nu$ admits a unique minimizer in H.*

This is a direct consequence of the strict convexity of $\mathcal{J}$, see Section 4. In summary, in the whole article, we work under the following conditions.

Assumption 1. *$V \in \mathcal{C}^\infty(\mathbb{T}^D)$, $D \geqslant 3$, $\beta > 0$, $\lambda \geqslant 0$ and $K \in \mathcal{C}^\infty(\mathbb{T}^d \times \mathbb{T}^d, \mathbb{R}_+)$ satisfies (9). Moreover, either $K > 0$ or $\lambda > 0$.*

We now have all the elements to define the reference ABF algorithm in this work, see Algorithm 1 below.

Algorithm 1 ABF algorithm

Input:
Initial condition $(q_0, z_0) \in \mathbb{T}^p \times \mathbb{T}^d$
Brownian motion $(B_t^1, B_t^2)_{t\geqslant 0}$ on $\mathbb{T}^p \times \mathbb{T}^d$
Regularization parameters K, λ
Update period $T_{up} > 0$, number of updates $N_{up} \in \mathbb{N}_*$, total simulation time $T_{tot} = T_{up}N_{up}$
Output:
Estimated free energy $A_{T_{tot}} \in H$
Trajectory $(Q_t, Z_t)_{t\in[0,T_{tot}]} \in \mathcal{C}\left([0,T_{tot}], \mathbb{T}^p \times \mathbb{T}^d\right)$
Begin:
Set $(Q_0, Z_0) = (q_0, z_0)$.
Set $A_0(z) = 0$ for all $z \in \mathbb{T}^d$.
Set $t_k = kT_{up}$ for all $k \in [\![0, N_{up}]\!]$.
for $k \in [\![1, N_{up}]\!]$ **do**
    Set $A_t = A_{t_{k-1}}$ for all $t \in [t_{k-1}, t_k)$.
    Set $(Q_t, Z_t)_{t\in[t_{k-1},t_k]}$ to be the solution of (5) with initial condition $(Q_{t_{k-1}}, Z_{t_{k-1}})$ at time t_{k-1}.
    Set A_{t_k} to be the minimizer in H of $\mathcal{J}_{\nu_{t_k}}$ given by (8) and (11).
end for
Return $A_{T_{tot}}$ and $(Q_t, Z_t)_{t\in[0,T_{tot}]}$.

Remark that, contrary to the cases studied in other theoretical works like [35, 1, 6], in Algorithm 1, the bias A_t is piecewise constant in time, with updates at the times t_k, $k \in [\![1, N_{up}]\!]$. This is due to the fact that, as will be detailed in Section 2.2, in addition to the classical case, we are also interested in a case where a bias update is numerically demanding, and thus cannot be performed at each time step.

We prove in Section 3 the long-time convergence of Algorithm 1:

Theorem 2. *Under Assumption 1, let $(Q_t, Z_t, A_t)_{t\geqslant 0}$ be given by Algorithm 1 (with a fixed $T_{up} > 0$ and $N_{up} = +\infty$ so that $T_{tot} = +\infty$ and the process is defined for all positive times). Then, as $t \to +\infty$, almost surely, ν_t given by (8) weakly converges toward μ and*

$$\|\nabla A_t - \nabla A_*\|_\infty \underset{t\to\infty}{\longrightarrow} 0,$$

where A_* is the unique minimizer in H of $\mathcal{J}_\mu$. Moreover, A_* satisfies

$$\int_{\mathbb{T}^d} |\nabla A(z) - \nabla A_*(z)|^2 \left(\int_{\mathbb{T}^p\times\mathbb{T}^d} K(y,z)\mu(q,y)\,dq\,dy \right) dz \\ \leqslant 4\|\nabla^2 A\|_\infty \sup_{y\in\mathbb{T}^d} \int_{\mathbb{T}^d} |y-z|^2 K(y,z)\,dz + 2\lambda \int_{\mathbb{T}^d} |\nabla A(z)|^2 dz\,. \quad (12)$$

Note that (12) implies that, as λ and $\sup_{y\in\mathbb{T}^d} \int_{\mathbb{T}^d} |y-z|^2 K(y,z)\mathrm{d}z$ go to zero, A_* converges in H to $A - \int A$ (which corresponds to $\lambda = 0$ and $K(y,z) = \delta_y(z)$).

The almost sure weak convergence of ν_t toward μ implies, of course, the almost sure convergence of the importance sampling estimator $\int_{\mathbb{T}^D} \varphi \mathrm{d}\nu_t$ toward the target $\int_{\mathbb{T}^D} \varphi \mathrm{d}\mu$ for all continuous observable φ.

The long-time convergence of a similar ABF algorithm has been established in [1] but in a case where, instead of its occupation measure, the process interacts with its law at time t. Rather than a self-interacting process (i.e. a single trajectory with memory), this corresponds to a system of N interacting particles (with no memory), and more precisely to the mean-field limit as N goes to infinity of this system. The techniques to study such a non-linear process is completely different from our non-Markovian case. Moreover, a result similar to Theorem 2 has been established in [6] for a closely related self-interacting process, the adaptive biasing potential algorithm. In addition, in the recent work [8], a similar result is established for the ABF algorithm but when the occupation measure is not unbiased (see the discussion in Section 5.3).

The previous qualitative result states that the algorithm is consistent, but gives no information on its efficiency. We now state that the asymptotic variance of the estimators obtained from the ABF algorithm is the same as in the case of a process with constant biasing potential equal to A_*. More precisely, consider $X^* = (Q^*, Z^*)$ the solution of

$$\left\{ \begin{array}{rcl} \mathrm{d}Q_t^* &=& -\nabla_q V(Q_t^*, Z_t^*)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t^1 \\ \mathrm{d}Z_t^* &=& -\nabla_z V(Q_t^*, Z_t^*)\mathrm{d}t + \nabla_z A_*(Z_t^*)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t^2\,, \end{array} \right.$$

where A_* is given by Theorem 2, and let

$$\nu_t^* = \left(\int_0^t e^{-\beta A_*(Z_s)}\mathrm{d}s \right)^{-1} \int_0^t \delta_{X_s^*} e^{-\beta A_*(Z_s)}\mathrm{d}s\,.$$

Theorem 3. *Under the settings of Theorem 2, there exist $C > 0$ such that for all $t \geqslant 0$ and all $\varphi \in \mathcal{C}\left(\mathbb{T}^D\right)$,*

$$\mathbb{E}\left(|\nu_t(\varphi) - \mu(\varphi)|^2\right) \leqslant \frac{C}{t}\|\varphi\|_\infty^2\,.$$

Moreover, $t\mathbb{E}\left(|\nu_t(\varphi)-\mu(\varphi)|^2\right)$ *converges as* $t\to+\infty$ *to a limit* $\sigma^2_\infty(\varphi)\in\mathbb{R}_+$*, which is also the limit of* $t\mathbb{E}\left(|\nu_t^*(\varphi)-\mu(\varphi)|^2\right)$*.*

As detailed in the proof of Theorem 3 (in Section 3), the asymptotic variance is given as follows:

$$\sigma^2_\infty(\varphi) \;:=\; \frac{2}{\beta}\int_{\mathbb{T}^D} e^{\beta A_*\circ\xi}\,|\nabla\psi|^2\,\mathrm{d}\mu\int_{\mathbb{T}^D} e^{\beta A_*\circ\xi}\mathrm{d}\mu$$

where ψ solves the Poisson equation

$$\left(\frac{1}{\beta}\Delta-\nabla\left(V-A_*\circ\xi\right)\nabla\right)\psi \;=\; e^{-\beta A_*\circ\xi}\left(\varphi-\int_{\mathbb{T}^D}\varphi\mathrm{d}\mu\right).$$

The consequences of Theorem 3 in term of efficiency of the algorithm are discussed in Section 2.4.

2.2 Tensor approximation

This section focuses on the minimization step of Algorithm 1. Assumption 1 is enforced. Fix $\nu\in\mathcal{P}(\mathbb{T}^p\times\mathbb{T}^d)$. For all $f\in H$, the cost function $\mathcal{J}_\nu(f)$ defined by (11) is equal to

$$\mathcal{J}_\nu(f) \;=\; C_\nu+(1+\lambda)\int_{\mathbb{T}^d}|F_\nu(z)-\nabla f(z)|^2\theta_\nu(z)\mathrm{d}z \tag{13}$$

with some constant C_ν independent from f and where, for all $z\in\mathbb{T}^d$,

$$\theta_\nu(z) \;:=\; \frac{1}{\lambda+1}\left(\lambda+\int_{\mathbb{T}^p\times\mathbb{T}^d}K(y,z)\mathrm{d}\nu(q,y)\right) \tag{14}$$

$$F_\nu(z) \;:=\; \frac{1}{(\lambda+1)\theta_\nu(z)}\int_{\mathbb{T}^p\times\mathbb{T}^d}\nabla_y V(q,y)K(y,z)\mathrm{d}\nu(q,y)\,. \tag{15}$$

Note that, under Assumption 1, θ_ν is the density of a probability measure, bounded from below by $(\lambda+\min K)/(1+\lambda)>0$. Moreover, since K is smooth and bounded, so are θ_ν and F_ν. Note that neither the additive constant C_ν nor the multiplication by $1+\lambda$ affect the problem of minimizing $\mathcal{J}_\nu$. As a consequence, the unique minimizer f_* of $\mathcal{J}_\nu$ on H (see Proposition 1) is equivalently the unique minimizer of

$$H\ni f\;\mapsto\;\widetilde{\mathcal{J}}_\nu(f)\;:=\;\int_{\mathbb{T}^d}|F_\nu(z)-\nabla f(z)|^2\theta_\nu(z)\mathrm{d}z. \tag{16}$$

The gradient of $\widetilde{\mathcal{J}}_\nu$ at f_* is the Helmholtz projection in $L^2(\theta_\nu)$ of F_ν. The Euler-Lagrange equation associated to the minimization problem of $\widetilde{\mathcal{J}}_\nu$ over H is

$$\nabla\cdot\left(\theta_\nu\left(\nabla f_*-F_\nu\right)\right) \;=\; 0\,, \tag{17}$$

where $\nabla\cdot$ denotes the divergence operator. When d is small ($d=2$ in [1]), as t increases, the functions θ_{ν_t} and $\theta_{\nu_t}F_{\nu_t}$ are updated and kept in memory on a discrete grid of dimension M^d for some $M\in\mathbb{N}^*$, and the Euler equation is solved with standard PDE techniques. However this is not sustainable if one wants to consider a larger number of reaction coordinates. For this reason, we now present a method to approximate f_* by a sum of tensor products, namely by a function $f_m\in H$ which reads as follows

$$\forall z:=(z_1,\cdots,z_d)\in\mathbb{T}^d,\quad f_m(z)\;=\;\sum_{k=1}^m\prod_{j=1}^d r_{k,j}(z_j)$$

for some $m \in \mathbb{N}_*$ and some functions $r_{k,j} : \mathbb{T} \to \mathbb{R}$ for $1 \leq j \leq d$ and $1 \leq k \leq m$. See [23] for a general overview on tensor methods.

Let g be a simple tensor product function, i.e. a function such that for all $z = (z_1, \cdots, z_d) \in \mathbb{T}^d$, $g(z) = \prod_{j=1}^d r_j(z_j)$ for some $r_1, \cdots, r_d \in H^1(\mathbb{T})$. Such a simple tensor product function will be denoted herefater by $g = \bigotimes_{j=1}^d r_j$.

If g belongs to H, its (Lebesgue) integral vanishes, which is equivalent to the fact there exists $i \in [\![1, d]\!]$ such that the (Lebesgue) integral of r_i vanishes. This motivates the introduction of the following subspaces of H: for $i \in [\![1, d]\!]$, define

$$\Sigma_i \;:=\; \left\{ g \in H,\; g = \bigotimes_{j=1}^{d} r_j \text{ with } r_j \in H^1(\mathbb{T}) \text{ for all } j \in [\![1, d]\!] \text{ and } \int_{\mathbb{T}} r_i(z_i) \mathrm{d}z_i = 0 \right\}.$$

Proposition 4. *Under Assumption 1, for all $\nu \in \mathcal{P}(\mathbb{T}^p \times \mathbb{T}^d)$, $i \in [\![1, d]\!]$ and $f \in H$, there always exists at least one minimizer in Σ_i to the optimization problem*

$$\min_{g \in \Sigma_i} \mathcal{J}_\nu(f + g). \tag{18}$$

This is proven in Section 4. From Proposition 4, the greedy algorithm described in Algorithm 2 below is well-defined.

Algorithm 2 Greedy(ν, f_0, m)

Input:
Probability measure $\nu \in \mathcal{P}(\mathbb{T}^p \times \mathbb{T}^d)$
Initial guess $f_0 \in H$
number of tensor terms $m \in \mathbb{N}_*$

Output:
$f_m \in H$.

Begin:
$n = 0$
while $n < m$ **do**
    for $i \in [\![1, d]\!]$ **do**
        Find $g_n := \bigotimes_{j=1}^d r_{n,j}$ a minimizer of $g \mapsto \mathcal{J}_\nu(f_n + g)$ over $g \in \Sigma_i$ (i.e. with $r_{n,j} \in H^1(\mathbb{T})$ for all $1 \leq j \leq d$ and $\int_{\mathbb{T}} r_{n,i} = 0$)
        Set $f_{n+1} = f_n + g_n$.
        Increment $n \leftarrow n + 1$.
    end for
end while

Return f_m.

In Section 4 is established the following:

Theorem 5. *Under Assumption 1, let f_* be the minimizer of $\mathcal{J}_\nu$ in H and $f_m = \mathrm{Greedy}(\nu, f_0, m)$ as given by Algorithm 2 for some $\nu \in \mathcal{P}(\mathbb{T}^p \times \mathbb{T}^d)$, $f_0 \in H$ and $m \in \mathbb{N}$. Then*

$$\|f_m - f_*\|_{H^1} \underset{m \to +\infty}{\longrightarrow} 0\,.$$

The interest of Algorithm 2 is that at each iteration, one only has to compute d one-dimensional functions, which makes it possible to implement even if d is relatively large (say

$4 < d < 10$). Notice that the price to pay when going from the original problem of minimizing $\mathcal{J}_\nu$ over H to the problem (18) is that the Euler-Lagrange equations associated to the initial problem are linear (since $\mathcal{J}_\nu$ is a quadratic functional) whereas the Euler-Lagrange equations associated to (18) are nonlinear. This is due to the fact that the quadratic functional is minimized over a *non-linear* space in (18).

In practice, a minimizer of $\mathcal{J}_\nu(f+g)$ over $g = \bigotimes_{j=1}^d r_j \in \Sigma_i$ is approximated through the Alternating Least Square method [19], which is a fixed point procedure on the Euler-Lagrange equation (17): the r_j's are optimized one after the other, the others being fixed, repeatedly. This amounts to solving a system of one-dimensional elliptic PDEs of the form

$$\partial_{z_j}\left(a_j \partial_{z_j} r_j\right)(z_j) - b_j(z_j) r_j(z_j) = c_j(z_j) \tag{19}$$

with

$$\begin{aligned}
a_j(z_j) &= \int_{\mathbb{T}^{d-1}} \left(\prod_{l\neq j} r_l(z_l)\right)^2 \theta_\nu(z) \mathrm{d}z_{\neq j} \\
b_j(z_j) &= \sum_{h\neq j} \int_{\mathbb{T}^{d-1}} \left|\partial_{z_h} \prod_{l\neq j} r_l(z_l)\right|^2 \theta_\nu(z) \mathrm{d}z_{\neq j} \\
c_j(z_j) &= \int_{\mathbb{T}^{d-1}} \left(\prod_{l\neq j} r_l(z_l)\right) \partial_{z_j}\left(F_{\nu,j}\theta_\nu\right)(z) \mathrm{d}z_{\neq j} - \sum_{h\neq j} \int_{\mathbb{T}^{d-1}} \partial_{z_h}\left(\prod_{l\neq j} r_l(z_l)\right) F_{\nu,h}(z)\theta_\nu(z)\mathrm{d}z_{\neq j}\,,
\end{aligned}$$

where $\mathrm{d}z_{\neq j}$ means that all variables except the j^{th} are integrated and $F_{\nu,j}$ denotes the j^{th} component of F_ν. If for all $y = (y_1, \cdots, y_d), z = (z_1, \cdots, z_d) \in \mathbb{T}^d$, $K(y,z) = \Pi_{i=1}^d K_i(y_i, z_i)$ for some functions $K_i : \mathbb{T}\times\mathbb{T} \to \mathbb{R}$ for all $1 \leq i \leq d$ (like the kernel (10)), for $\nu = \nu_t$ given by (8),

$$(1+\lambda)a_j(z_j) = \lambda \prod_{l\neq j} \|r_l\|^2_{L^2(\mathbb{T})} + \int_{\mathbb{T}^D} \left(\prod_{l\neq j} \int_{\mathbb{T}} r_l^2(z_l) K_l(y_l, z_l)\mathrm{d}z_l\right) K_j(y_j, z_j)\mathrm{d}\nu_t(q,y)\,,$$

which can be computed without computing $F_{\nu_t}(z)$ and $\theta_{\nu_t}(z)$ for all $z \in \mathbb{T}^d$ (which would be impossible in practice). The same holds for b_j and c_j.

2.3 The tensor ABF algorithm

As already explained above, the main objective of this work is to introduce a new algorithm to adapt the standard ABF approach to multi-dimensional reaction coordinates. Combining Algorithms 1 and 2, the Tensor ABF (TABF) algorithm is described in Algorithm 3 below. Note that, for the sake of clarity, it has been kept relatively simple. In particular, we haven't addressed here the question of time and space discretization.

Moreover, the proofs of convergence of Algorithm 1 and Algorithm 2 also have their own interest. The convergence of Algorithm 1 is based on the so-called ordinary differential equation method [9], and requires specific contractivity bounds. The convergence of Algorithm 2 is an adaptation of the proof of convergence of greedy algorithms [12], the main difficulty being to deal with the zero average constraint in H.

2.4 Discussion on the results and efficiency

In view of our theoretical results, we now discuss separately the efficiency of the three algorithms: first, the usual ABF algorithm (Algorithm 1) associated with Theorems 2 and 3;

Algorithm 3 TABF algorithm

Input:
Initial condition $(q_0, z_0) \in \mathbb{T}^p \times \mathbb{T}^d$
Brownian motion $(B_t^1, B_t^2)_{t \geqslant 0}$ on $\mathbb{T}^p \times \mathbb{T}^d$
Regularization parameters K, λ
Update period $T_{up} > 0$, number of updates $N_{up} \in \mathbb{N}_*$, total simulation time $T_{tot} = T_{up} N_{up}$
Number of tensor terms by update $m \in \mathbb{N}_*$

Output:
Estimated free energy $A_{T_{tot}} \in H$
Trajectory $(Q_t, Z_t)_{t \in [0, T_{tot}]} \in \mathcal{C}\left([0, T_{tot}], \mathbb{T}^p \times \mathbb{T}^d\right)$

Begin:
Set $(Q_0, Z_0) = (q_0, z_0)$.
Set $A_0(z) = 0$ for all $z \in \mathbb{T}^d$.
Set $t_k = kT_{up}$ for all $k \in [\![0, N_{up}]\!]$.
for $k \in [\![1, N_{up}]\!]$ **do**
    Set $A_t = A_{t_{k-1}}$ for all $t \in [t_{k-1}, t_k)$.
    Set $(Q_t, Z_t)_{t \in [t_{k-1}, t_k]}$ to be the solution of (5) with value $(Q_{t_{k-1}}, Z_{t_{k-1}})$ at time t_{k-1}.
    Set $f_m = \text{Greedy}(\nu_{t_k}, A_{t_{k-1}}, m)$ given by Algorithm 2 where ν_{t_k} is given by (8).
    Set $A_{t_k} = f_m$.

Return $A_{T_{tot}}$ and $(Q_t, Z_t)_{t \in [0, T_{tot}]}$.

second, the greedy tensor approximation of the free energy (Algorithm 2), associated with Theorem 5; finally, the TABF algorithm (Algorithm 3).

2.4.1 Enhanced sampling with the ABF algorithm

As Algorithm 1 is meant to tackle metastability issues, a natural frame to discuss its efficiency is the low temperature regime $\beta \to +\infty$. For Markov processes, obtaining an equivalent in this regime of the convergence rate of the law of the process toward its equilibrium is a classical topic, but the case of non-Markovian self-interacting dynamics or similar stochastic algorithms is known to be much more difficult, and there are much less results. Theorem 3 states that, in term of asymptotic variance, the efficiency of the adaptive scheme is approximately (as A_* is close to A) the same as the efficiency of the importance sampling scheme based on (4). As discussed earlier, the question is thus whether (4) would be efficient. One way to quantify the interest of using (4) is to discuss the spectral gap of the associated infinitesimal generator: it is indeed known that the larger the spectral gap, the smaller the asymptotic variance, and the faster the convergence to equilibrium. In fact, at low temperature, the spectral gap of the overdamped Langevin process is well known to scale as $\exp(-\beta c_*)$ where c_* is the so-called critical depth of the potential, see [27]. On the other hand, applying the results of [33], we see that the spectral gap of (4) can be obtained, on the one hand, from the Poincaré inequality satisfied by the marginal law of the reaction coordinates and, on the other hand, by the conditional laws for fixed values of the reaction coordinates. The marginal law being uniform on the torus for all β, the scaling in β of the spectral gap of (4) is given by the scaling of the Poincaré inequality of the conditional laws, which means only the "orthogonal" metastability intervenes. The spectral gap of (4) then scales at most as $\exp(-\beta \sup_{z \in \mathbb{T}^d} c_*(z))$ where $c_*(z)$ is the critical depth of $q \mapsto V(q, z)$. This gives a precise criterion (although difficult to use in practice) for selecting reaction coordinates: a reaction coordinate is good if

$\sup_{z\in\mathbb{T}^d} c_*(z) < c_*$. In the toy problem studied in Section 6.1, for instance, $\sup_{z\in\mathbb{T}^d} c_*(z) = 0$ (there is no orthogonal metastability).

A numerical comparison of a classical overdamped Langevin sampler and of a TABF algorithm in this case at low temperature is given in Figure 4. We can see that, in the same physical time, the adaptive process successfully visits the whole space, while the classical sampler remains trapped in its initial well. Of course this is not a fair comparison of the practical algorithms since the numerical cost of the adaptive algorithm is higher, but it illustrates the difference of the sampling rates of the continuous-time processes. In fact, for a reaction coordinate of dimension less than 3, a standard ABF method, for which the free energy is recorded on a grid rather than approximated in a tensor form, has a numerical cost in term of computational time which is very similar to a basic unbiased simulation of an overdamped Langevin process (there is also a small additional memory). This means that, in that case, a comparison at a given computational time between the two algorithms yields a result similar to a comparison at a given physical time (as in our theoretical result of Theorem 3 and in the numerical results displayed in Figure 4), i.e. the ABF method is efficient in practice at a given computational cost. From a numerical point of view, this is a well-known fact, indeed this algorithm has proven to be useful and efficient for nearly 20 years in a large number of empirical studies, see e.g. [17, 26, 21, 16] and references within. For this reason, in the present work, only the TABF variant has been implemented and used for the numerical experiments of Section 6. However, our contribution, i.e. Theorems 2 and 3, gives the first theoretical proof of the consistency and efficiency of the self-interacting ABF method.

2.4.2 Tensor approximation of the free energy

First, it should be noticed that the convergence of the greedy tensor method, Algorithm 2, is proven only in the case where the problem of minimizing $\mathcal{J}_\nu(f+g)$ over single tensor terms $g = \bigotimes_{j=1}^d r_j$ is exactly solved, which is in fact not the case with the Alternating Least Square method (see [19, 50] for convergence results for this algorihm). Moreover, by contrast with Theorem 3 for the long-time convergence, the convergence of the tensor approximation stated in Theorem 5 is only qualitative.

That being said, let us focus on the efficiency of Algorithm 2. Whereas the precise convergence rate is not known from a theoretical point of view, it is numerically observed in numerous applications related to computational mechanics [14, 15, 44], rheology [13], biology [4] or uncertainty quantification [49], that this method, also called *Proper Generalised Decomposition* enables to give very good numerical approximations of the functions to be computed with a low number of tensor terms or iterations. In our context, we provide in Section 6 a numerical experiment where interesting non-trivial results are obtained when approximating a 5-dimensional free energy with $m = 140$ tensor terms, each one-dimensional function being piecewise linear on a grid with $M = 30$ points. Hence, the total memory cost is $5mM$, orders of magnitude smaller than M^5. We refer to the end of Section 6 (in particular Figure 10) for a more detailed numerical discussion on the efficiency of the tensor approximation.

2.4.3 TABF and free energy estimation

Recall that there are two distinct motivations for the TABF algorithm: first, enhancing the sampling and, second, estimating the free energy associated to a high-dimensional reaction coordinate.

Concerning the first objective, what has been said in Section 2.4.1 for the usual ABF algorithm is still true for the TABF variant as far as we consider the continuous-time processes in physical time (again, see Figure 4, which is obtained with the TABF algorithm). However,

in contrast with the usual case where the free energy is recorded on a grid, since the greedy tensor approximation is numerically demanding, the numerical cost of the TABF algorithm is increased with respect to a non-biased standard MCMC simulation. In the context of the numerical experiments conducted in Section 6, which have a purely illustrative purpose, the TABF algorithm does not yield any gain with respect to a non-biased simulation in term only of sampling at a given computational cost, for two reasons:

- As discussed in Section 5, to be competitive with unbiased algorithms, a more efficient implementation of the TABF algorithm would be necessary.
- In the models used in the numerical experiments (e.g. 100 particles with pairwise interactions), the numerical cost for computing the forces ∇V is low, which implies that the construction of the tensor approximation becomes the bottleneck. This is not representative of the real-life test cases where evaluating the force is much more expensive.

As a consequence, the current implementation of the algorithm which has been used in Section 6 does not allow for a real comparison in term of sampling efficiency of the TABF algorithm with a standard MCMC method. This would require the interfacing of a high-performance TABF module within a general molecular dynamics software, in order to test the algorithm on applications in computational chemistry where the computation of the forces at each time step is the overwhelming limiting factor, so that the extra cost of the tensor algorithm does not impair the overall complexity. This exceeds the scope of the present work.

Concerning the second objective, namely the estimation of the free energy, the numerical experiments of Section 6 show that a simple implementation of Algorithm 3 is able to estimate the free energy of 5-dimensional highly-correlated reaction coordinates (see in particular Figure 7). Let us emphasize that, to our knowledge, there are no other implementable provably convergent algorithm to perform such a task due to the curse of dimensionality.

3 Proof of the long-time convergence

In the whole Section 3 we consider the ABF process $(Q_t, Z_t, A_t)_{t\geqslant 0}$ obtaind through Algorithm 1 (with $N_{up} = +\infty$), and Assumption 1 holds.

Lemma 6. *For all $r \in \mathbb{N}^*$ and all multi-index $\alpha \in \mathbb{N}^r$, there exists a constant $C_\alpha > 0$ such that, for all $t \geqslant 0$, $\|\partial^\alpha A_t\|_\infty \leqslant C_\alpha$.*

Proof. Since $\mathbb{R}_+ \ni t \mapsto A_t$ is piecewise constant, we may assume that $t = t_k = kT_{up}$ for some $k \in \mathbb{N}$ without loss of generality. Using the notation of Section 2.2, A_t is then the minimizer over H of $\widetilde{\mathcal{J}}_{\nu_t}$ defined in (16). Recall that for all $f \in H$,

$$\widetilde{\mathcal{J}}_{\nu_t}(f) \;=\; \int_{\mathbb{T}^d} |F_{\nu_t}(z) - \nabla f(z)|^2 \theta_{\nu_t}(z) \mathrm{d}z.$$

Remark that θ_{ν_t} is bounded from below uniformly in t and z by $(\lambda + \min K)/(1+\lambda) > 0$, and similarly all the derivatives in z of θ_{ν_t} and of F_{ν_t} are bounded in $L^\infty(\mathbb{T}^d)$ by constants which depend on K and V but not on t. The Euler-Lagrange equation associated to the minimization of $\widetilde{\mathcal{J}}_{\nu_t}$ reads

$$\nabla \cdot (\theta_{\nu_t} \nabla A_t) \;=\; \nabla \cdot (\theta_{\nu_t} F_{\nu_t}) \; . \tag{20}$$

By elliptic regularity (cf. [2]), A_t is thus $\mathcal{C}^\infty$ and, differentiating (20), multiplying it by derivatives of ∇A_t and integrating, we classically get by induction that

$$\int_{\mathbb{T}^d} |\partial^\alpha \nabla A_t|^2 \theta_{\nu_t} \;\leqslant\; C_\alpha$$

where $\alpha \in \mathbb{N}^r$ is any multi-index for any $r \in \mathbb{N}_*$, for some constant $C_\alpha > 0$ which does not depend on t. Conclusion follows from Sobolev embeddings. $\square$

Theorem 2 will be a direct corollary of:

Proposition 7. *Almost surely,* $\nu_t \overset{weak}{\underset{t\to\infty}{\longrightarrow}} \mu$.

The proof of Proposition 7 is postponed to the end of this section. Let us prove that indeed, given the latter, Theorem 2 holds:

Proof of Theorem 2. By the arguments of the previous proof, for all $t \geq 0$, the function $\mathbb{T}^d \ni z \mapsto \theta_{\nu_t}(z)$ is bounded and Lipschitz with constants which are uniform in t. Hence, for any $\varepsilon > 0$, we can find $N_\varepsilon \in \mathbb{N}^*$ and a finite set of points $z_1, \cdots, z_{N_\varepsilon} \in \mathbb{T}^d$ such that for all $z \in \mathbb{T}^d$, there exists $i_z \in [\![1, N_\varepsilon]\!]$ such that, for all $t > 0$, $|\theta_{\nu_t}(z_{i_z}) - \theta_{\nu_t}(z)| \leqslant \varepsilon$. The same holds for θ_μ. On the other hand, according to Proposition 7, almost surely,

$$\sup_{i\in[\![1,N_\varepsilon]\!]} |\theta_{\nu_t}(z_i) - \theta_\mu(z_i)| \underset{t\to\infty}{\longrightarrow} 0,$$

so that $\|\theta_{\nu_t} - \theta_\mu\|_\infty$ goes to zero as $t \to \infty$. Similar arguments enable us to obtain the same results for all the derivatives of θ_{ν_t} and for F_{ν_t} and all its derivatives. Note that A_* is the minimizer of

$$\widetilde{\mathcal{J}}_\mu(f) \;=\; \int_{\mathbb{T}^d} |F_\mu(z) - \nabla f(z)|^2 \theta_\mu(z) \mathrm{d}z.$$

Let $t = t_n$ for some $n \in \mathbb{N}^*$. The associated Euler-Lagrange equations associated with the two minimization problems on A_t and A_* lead to

$$\nabla \cdot (\theta_\mu \nabla (A_t - A_*)) \;=\; \nabla \cdot (\theta_{\nu_t} F_{\nu_t} - \theta_\mu F_\mu - (\theta_{\nu_t} - \theta_\mu) \nabla A_t). \tag{21}$$

Multiplying this equality by $A_t - A_*$, integrating and using the uniform control on ∇A_t established in Lemma 6 (and the lower bound on θ_μ), we get that

$$\int_{\mathbb{T}^d} |\nabla (A_t(z) - A_*(z))|^2 \mathrm{d}z \underset{t\to\infty}{\longrightarrow} 0.$$

More generally, differentiating (21), multiplying it by derivatives of $A_t - A_*$, integrating and using the uniform controls of the derivatives of A_t, we obtain by induction that

$$\int_{\mathbb{T}^d} |\nabla \partial^\alpha (A_t - A_*)(z)|^2 \mathrm{d}z \underset{t\to\infty}{\longrightarrow} 0$$

for all multi-index $\alpha \in \mathbb{N}^d$. The first statement of Theorem 2 then follows from Sobolev embeddings.

Finally, inequality (12) stems from the fact that $\mathcal{J}_\mu(A_*) \leqslant \mathcal{J}_\mu(A)$. More precisely, using that

$$\int_{\mathbb{T}^d} \nabla A(y) \mu(q,y) \mathrm{d}q \;=\; \int_{\mathbb{T}^d} \nabla_y V(q,y) \mu(q,y) \mathrm{d}q,$$

we get that for all $f \in H$, $\mathcal{J}_\mu(f) = \widehat{\mathcal{J}}_\mu(f) + \int |\nabla_y V|^2 \mathrm{d}\mu - \int |\nabla A|^2 \mathrm{d}\mu$ where

$$\widehat{\mathcal{J}}_\mu(f) \;=\; \int_{\mathbb{T}^p\times\mathbb{T}^d\times\mathbb{T}^d} |\nabla A(y) - \nabla f(z)|^2 K(y,z) \mathrm{d}z \mathrm{d}\mu(q,y) + \lambda \int_{\mathbb{T}^d} |\nabla f(z)|^2 \mathrm{d}z.$$

In other words, $\mathcal{J}_\mu$ and $\widehat{\mathcal{J}}_\mu$ only differ by an additive constant, so that A_* is the minimizer of $\widehat{\mathcal{J}}_\mu$ over H. Then

$$\begin{aligned}
&\int_{\mathbb{T}^p\times\mathbb{T}^d\times\mathbb{T}^d} |\nabla A(z) - \nabla A_*(z)|^2 K(y,z)\mathrm{d}z\mathrm{d}\mu(q,y) \\
&\quad\leqslant\ 2\widehat{\mathcal{J}}_\mu(A_*) + 2\int_{\mathbb{T}^p\times\mathbb{T}^d\times\mathbb{T}^d} |\nabla A(y) - \nabla A(z)|^2 K(y,z)\mathrm{d}z\mathrm{d}\mu(q,y) \\
&\quad\leqslant\ 2\widehat{\mathcal{J}}_\mu(A) + 2\int_{\mathbb{T}^p\times\mathbb{T}^d\times\mathbb{T}^d} |\nabla A(y) - \nabla A(z)|^2 K(y,z)\mathrm{d}z\mathrm{d}\mu(q,y) \\
&\quad\leqslant\ 2\lambda\int_{\mathbb{T}^d} |\nabla A(z)|^2\mathrm{d}z + 4\int_{\mathbb{T}^p\times\mathbb{T}^d\times\mathbb{T}^d} |\nabla A(y) - \nabla A(z)|^2 K(y,z)\mathrm{d}z\mathrm{d}\mu(q,y) \\
&\quad\leqslant\ 2\lambda\int_{\mathbb{T}^d} |\nabla A(z)|^2\mathrm{d}z + 4\|\nabla^2 A\|_\infty^2 \sup_{y\in\mathbb{T}^d}\int_{\mathbb{T}^d} |y-z|^2 K(y,z)\mathrm{d}z
\end{aligned}$$

□

The rest of the section is dedicated to the proof of Proposition 7 and Theorem 3. We start with a presentation of the the so-called ordinary differential equation (ODE) method of [9], which introduces some general ideas of the proof of Proposition 7 (although, as we will see, we are in a very simple case so that we won't really use the fully general method).

3.1 Time change and the ODE method

Following an idea of [6], we introduce the (random) time change:

$$\tau(t) \ :=\ \int_0^t e^{-\beta A_s(Z_s)}\mathrm{d}s,$$

so that

$$\nu_t \ =\ \frac{1}{\tau(t)}\int_0^t \delta_{Q_s,Z_s}\tau'(s)\mathrm{d}s \ =\ \frac{1}{\tau(t)}\int_0^{\tau(t)} \delta_{Q_{\tau^{-1}(s)},Z_{\tau^{-1}(s)}}\mathrm{d}s\,.$$

In other words, considering the time-changed process $\overline{X}_t := \left(Q_{\tau^{-1}(t)}, Z_{\tau^{-1}(t)}\right)$ and its occupation measure

$$\bar{\nu}_t \ =\ \frac{1}{t}\int_0^t \delta_{\overline{X}_s}\mathrm{d}s\,, \tag{22}$$

then $\nu_t = \bar{\nu}_{\tau(t)}$. Since, at a fixed time $t \geqslant 0$, A_t is smooth and with Lebesgue integral zero, there always exists $z \in \mathbb{T}^d$ such that $A_t(z) = 0$, so that

$$\|A_t\|_\infty \ \leqslant\ \sqrt{d}/2\|\nabla A_t\|_\infty \tag{23}$$

where we used that $\sqrt{d}/2$ is the diameter of $\mathbb{T}^d$. Together with Lemma 6, this implies that in particular, $\tau(t)$ goes to infinity with t.

Denoting $S_t(x) := A_{\tau^{-1}(t)}(z)$ for all $x = (q,z) \in \mathbb{T}^p\times\mathbb{T}^d$, the inhomogeneous Markov process $\overline{X}$ solves the SDE

$$\mathrm{d}\overline{X}_t \ =\ -e^{\beta S_t\left(\overline{X}_t\right)}\nabla\left(V - S_t\right)\left(\overline{X}_t\right) + \sqrt{2\beta^{-1}e^{\beta S_t\left(\overline{X}_t\right)}}\mathrm{d}\overline{B}_t\,, \tag{24}$$

where $(\overline{B}_t)_{t\geqslant 0}$ is a standard Brownian motion on $\mathbb{T}^D$, obtained from $(B_t)_{t\geqslant 0}$ through rescaling. We denote by $(L_t)_{t\geqslant 0}$ its infinitesimal generator, defined by: for all $\varphi \in \mathcal{C}^2(\mathbb{T}^p\times\mathbb{T}^d) = \mathcal{C}^2(\mathbb{T}^D)$ and all $x \in \mathbb{T}^p\times\mathbb{T}^d$,

$$L_t\varphi(x) \ =\ \lim_{h\to 0}\frac{\mathbb{E}\left(\varphi(\overline{X}_{t+h}) \mid \overline{X}_t = x\right) - \varphi(x)}{h}$$

whenever the limit exists. Here,

$$L_t\varphi(x) \;=\; \left(-\nabla\left(V-S_t\right)(x)\cdot\nabla\varphi(x)+\frac{1}{\beta}\Delta\varphi(x)\right)e^{\beta S_t(x)}.$$

We denote by $(P_s^{(t)})_{s\geqslant 0}$ the Markov semi-group generated by L_t for a fixed t. Formally, $P_s^{(t)}=e^{sL_t}$. For all $t\geqslant 0$ the unique invariant measure of $(P_s^{(t)})_{s\geqslant 0}$ is μ (see [6, Proposition 3.1]), which is a natural consequence of the fact we consider a process interacting with its *unbiased* occupation measure. From Lemma 6 and the bound (23), we consider $C_0>0$ such that $S_t\in\mathcal{B}_{C_0}$ for all $t\geqslant 0$ where

$$\mathcal{B}_{C_0} \;:=\; \left\{S\in\mathcal{C}^\infty\left(\mathbb{T}^D\right),\ \int_{\mathbb{T}^D}S(x)\mathrm{d}x=0,\ \|S\|_{\mathcal{C}^2(\mathbb{T}^D)}\leqslant C_0\right\}.$$

The principle of the ODE method is the following: for large values of the time t, the evolution of $\bar{\nu}_t$ is slow (because of the t^{-1} factor in (22)). Hence, for $1\ll s\ll t$, in principle, it holds that $S_u\simeq S_t$ for $u\in[t,t+s]$, so that

$$\bar{\nu}_{t+s} \;=\; \frac{t}{t+s}\bar{\nu}_t+\frac{s}{s+t}\left(\frac{1}{s}\int_t^{t+s}\delta_{\overline{X}_u}\mathrm{d}u\right) \;\simeq\; \frac{t}{t+s}\bar{\nu}_t+\frac{s}{s+t}\mu. \tag{25}$$

In other words, the evolution of $\bar{\nu}_t$ approximately follows the deterministic flow

$$\partial_t m_t \;=\; \frac{1}{t}\left(\mu-m_t\right),$$

which converges to μ, so that $\bar{\nu}_t$ (hence ν_t) should also converge to μ.

In general cases of self-interacting processes, as those studied in [9], the asymptotic deterministic flow may be more complicated (see in particular [8] for the ABF algorithm with the *non-reweighted* occupation measure). Here, as mentioned above, we are in a very simple case since for all $t\geqslant 0$, μ is the invariant measure of the time-homogenous Markov process with generator L_t, so that the flow is simply a relaxation toward this equilibrium. For this reason, in order to make rigorous the previous heuristic, instead of applying the technical arguments of [9], we may use a shortcut that yields a simpler proof and more explicit estimates (allowing in particular to tackle the question of the asymptotic variance, which may be much more intricate in other cases), similarly to [7] for the ABP algorithm.

We will need some quantitative estimates. Indeed, note that, for the approximation (25) to hold, the speed of convergence of $P^{(t)}$ toward μ should be uniform in t, and the time evolution of A_t should be controlled in some sense. As we will see below, these are direct consequences of the estimates of Lemma 6.

3.2 Preliminary estimates

For a fixed $S\in\mathcal{C}^\infty(\mathbb{T}^D)$, consider L_S defined for $\varphi\in\mathcal{C}^\infty(\mathbb{T}^D)$ by

$$L_S\varphi(x) \;=\; \left(-\nabla\left(V-S\right)(x)\cdot\nabla\varphi(x)+\frac{1}{\beta}\Delta\varphi(x)\right)e^{\beta S(x)},$$

which is the infinitesimal generator of the SDE

$$\mathrm{d}X_t^S \;=\; -e^{\beta S(X_t^S)}\nabla\left(V-S\right)(X_t^S)\mathrm{d}t+\sqrt{2\beta^{-1}e^{\beta S(X_t^S)}}\mathrm{d}B_t\,.$$

Denote $\left(P_t^S\right)_{t\geqslant 0}$ the associated (homogeneous) semi-group and Γ_S the associated carré-du-champs operator, defined for $\varphi, \psi \in \mathcal{C}^\infty(\mathbb{T}^D)$ and all $x \in \mathbb{T}^D$ by

$$\Gamma_S(\varphi,\psi)(x) \;:=\; \frac{1}{2}\left(L_S(\varphi\psi) - \varphi L_S\psi - \psi L_S\varphi\right)(x) \;=\; \beta^{-1}e^{\beta S(x)}\nabla\varphi(x)\cdot\nabla\psi(x)\,,$$

and $\Gamma_S(\varphi) := \Gamma_S(\varphi,\varphi)$. By classical elliptic regularity arguments, if $\varphi \in \mathcal{C}^\infty(\mathbb{T}^D)$ then $P_t^S\varphi \in \mathcal{C}^\infty(\mathbb{T}^D)$, in particular $\mathcal{C}^\infty(\mathbb{T}^D)$ is a core for L_S, see [3, Section 1.13]. More precisely each derivative of $P_t^S\varphi$ is uniformly bounded over all finite time interval, which ensures the validity of the computations in the proofs of the next lemmas. Integrating twice by parts, it can be easily seen that for all $\varphi, \psi \in \mathcal{C}^\infty(\mathbb{T}^D)$,

$$\int_{\mathbb{T}^D} \varphi(x) L_S\psi(x)\mu(\mathrm{d}x) \;=\; \int_{\mathbb{T}^D} \psi(x) L_S\varphi(x)\mu(\mathrm{d}x)\,,$$

in other words L_S is a self-adjoint operator on $L^2(\mu)$.

Lemma 8. *Let us assume that $D \geq 3$. Then, there exists $C_1 > 0$ such that for all $S \in \mathcal{B}_{C_0}$, (μ, L_S) satisfies a Poincaré inequality and a Sobolev inequality both with constant C_1, in the sense that for all $\varphi \in \mathcal{C}^\infty(\mathbb{T}^D)$,*

$$\begin{aligned}
\|\varphi\|^2_{L^2(\mu)} &\leqslant C_1 \int_{\mathbb{T}^D} \Gamma_S(\varphi)\, d\mu \\
\|\varphi\|^2_{L^p(\mu)} &\leqslant C_1 \left(\|\varphi\|^2_{L^2(\mu)} + \int_{\mathbb{T}^D} \Gamma_S(\varphi)\, d\mu\right),
\end{aligned}$$

where $p = \frac{2D}{D-2}$.

Proof. For $S = 0$, the first inequality is the classical Poincaré inequality, which holds here since the density of μ with respect to the Lebesgue measure is bounded above and below away from zero, see [3, Proposition 5.1.6]. As a consequence, there exists $c > 0$ such that for all $S \in \mathcal{B}_{C_0}$ and $\varphi \in \mathcal{C}^\infty(\mathbb{T}^D)$,

$$\|\varphi\|^2_{L^2(\mu)} \;\leqslant\; c\int_{\mathbb{T}^D} |\nabla\varphi|^2 \mathrm{d}\mu \;\leqslant\; ce^{\beta C_0}\int_{\mathbb{T}^D} \Gamma_S(\varphi)\mathrm{d}\mu\,.$$

Similarly, from the Sobolev inequality satisfied by the Lebesgue measure on $\mathbb{T}^D$ [3, Section 6],

$$\begin{aligned}
\|\varphi\|^2_{L^p(\mu)} &\leqslant \|\mu\|^{2/p}_\infty \|\varphi\|^2_{L^p(\mathbb{T}^D)} \\
&\leqslant C\|\mu\|^{2/p}_\infty \left(\|\varphi\|^2_{L^2(\mathbb{T}^D)} + \|\nabla\varphi\|^2_{L^2(\mathbb{T}^D)}\right) \\
&\leqslant C\|\mu\|^{2/p}_\infty \|\mu^{-1}\|^2_\infty \left(\|\varphi\|^2_{L^2(\mu)} + \|\nabla\varphi\|^2_{L^2(\mu)}\right) \\
&\leqslant Ce^{\beta C_0}\|\mu\|^{2/p}_\infty \|\mu^{-1}\|^2_\infty \left(\|\varphi\|^2_{L^2(\mu)} + \int_{\mathbb{T}^D} \Gamma_S(\varphi)\mathrm{d}\mu\right).
\end{aligned}$$

□

These inequalities, in turn, yield the following estimates:

Lemma 9. *There exist $C_2 > 0$ such that, for all $S \in \mathcal{B}_{C_0}$, $t \geqslant 0$ and $\varphi \in \mathcal{C}^\infty(\mathbb{T}^D)$,*

$$\begin{aligned}
\|P_t^S\Pi\varphi\|_{L^2(\mu)} &\leqslant e^{-t/C_2}\|\Pi\varphi\|_{L^2(\mu)} \\
\|P_t^S\varphi\|_\infty &\leqslant \frac{C_2}{\min(1, t^{d/2})}\|\varphi\|_{L^2(\mu)} \\
\|\nabla P_t^S\varphi\|_\infty &\leqslant \frac{C_2}{\min(1,\sqrt{t})}\|\varphi\|_\infty,
\end{aligned}$$

with $\Pi\varphi := \varphi - \int_{\mathbb{T}^D}\varphi d\mu$.

Proof. The first estimate is a usual consequence of the Poincaré inequality, see [3, Proposition 5.1.3]. The second one, namely the ultracontractivity of the semi-group, is a consequence of the Sobolev inequality (see [3, Theorem 6.3.1]). The last one can be established thanks to the Bakry-Emery calculus (see [3, Section 1.16] for an introduction), by showing that L_S satisfies a curvature estimate, as we now detail. We would like to compare $|\nabla P_t\varphi|^2$ and $P_t(\varphi^2)$. A seminal idea of the Bakry-Emery calculus is that quantities of the form $\Theta(P_t\varphi)$ and $P_t\Theta(\varphi)$, where Θ is some operator can be linked through the interpolation $P_{t-s}\Theta(P_s\varphi)$, $s\in[0,t]$, so that $\Theta(P_t\varphi)-P_t\Theta(\varphi)=\int_0^t\partial_s\left(P_{t-s}\Theta(P_s\varphi)\right)\mathrm{d}s$. When differentiating with respect to s, we obtain quantities of the form $-2P_{t-s}\Gamma_\Theta(P_s\varphi)$ for some operator Γ_Θ, which is of a form similar to the interpolation (Θ being replaced by Γ_Θ).

More precisely, when $\Theta(\varphi)=\varphi^2$, then Γ_Θ is the usual carré-du-champ operator, and when $\Theta(\varphi)=|\nabla\varphi|^2$ we end up with

$$\Gamma_{\nabla,S}(\varphi) \;=\; \frac{1}{2}L_S\left(|\nabla\varphi|^2\right)-\nabla\varphi\cdot\nabla L_S\varphi\,,$$

for $\varphi\in\mathcal{C}^\infty(\mathbb{T}^D)$. Writing $[\varphi,\psi]=\varphi\psi-\psi\varphi$, we compute

$$\begin{aligned}
\Gamma_{\nabla,S}(\varphi) &= \sum_{i=1}^D\left(\Gamma_S(\partial_{x_i}\varphi)+\partial_{x_i}\varphi[\partial_{x_i},L_S]\varphi\right)\\
&\geqslant \sum_{i=1}^D\Big[\beta^{-1}e^{-\beta\|S\|_\infty}|\nabla\partial_{x_i}\varphi|^2-\beta^{-1}e^{\beta\|S\|_\infty}|\nabla\partial_{x_i}(V-S)||\nabla\varphi||\partial_{x_i}\varphi|\\
&\qquad -\beta|\partial_{x_i}S|e^{\beta\|S\|_\infty}|\partial_{x_i}\varphi||\nabla\left(V-S\right)\cdot\nabla\varphi+\frac{1}{\beta}\Delta\varphi|\Big]\\
&\geqslant -c|\nabla\varphi|^2
\end{aligned}$$

for some $c>0$ which is uniform over $S\in\mathcal{B}_{C_0}$. Now, following [42, Lemma 4], we want to consider the interpolation between $\alpha(t)|\nabla P_t\varphi|^2+(P_t\varphi)^2$ and $\alpha(0)P_t|\nabla\varphi|^2+P_t(\varphi^2)$ for some α with $\alpha(0)=0<\alpha(t)$. For fixed $\varphi\in\mathcal{C}^\infty(\mathbb{T}^D)$, $x\in\mathbb{T}^D$ and $t\geqslant 0$, we set for all $s\in[0,t]$

$$\Psi(s) \;=\; \alpha(s)P^S_{t-s}|\nabla P^S_s\varphi|^2(x)+e^{\beta C_0}P^S_{t-s}\left(P^S_s\varphi\right)^2(x)$$

with $\alpha(s)=(1-\exp(-2cs))/c$, so that

$$\begin{aligned}
\partial_s\Psi(s) &= P^S_{t-s}\left(-2\alpha(s)\Gamma_{\nabla,S}+\alpha'(s)|\nabla\cdot|^2-2e^{\beta C_0}\Gamma_S\right)\left(P^S_s\varphi\right)(x)\\
&\leqslant \left(2\alpha(s)c+\alpha'(s)-2\right)P^S_{t-s}|\nabla P^S_s\varphi|^2(x) \;=\; 0\,.
\end{aligned}$$

In particular,

$$\alpha(t)|\nabla P^S_t\varphi|^2(x) \;\leqslant\; \Psi(t) \;\leqslant\; \Psi(0) \;=\; e^{\beta C_0}P^S_t\varphi^2(x) \;\leqslant\; e^{\beta C_0}\|\varphi\|^2_\infty$$

which yields the desired estimate. □

Lemma 10. *There exists $C_3>0$ such that for all $S\in\mathcal{B}_{C_0}$, the operator R_S defined for all $\varphi\in\mathcal{C}^\infty(\mathbb{T}^D)$ by*

$$R_S\varphi \;=\; -\int_0^\infty P^S_t\Pi\varphi dt$$

satisfies $L_SR_S=R_SL_S=\Pi$ and, for all $\varphi\in\mathcal{C}^\infty(\mathbb{T}^D)$,

$$\|R_S\varphi\|_\infty+\|\nabla R_S\varphi\|_\infty\;+\|\Delta R_S\varphi\|_\infty \;\leqslant\; C_3\|\varphi\|_\infty. \tag{26}$$

Proof. We follow the proof of [9, Section 5.2 and Lemma 5.1]. First, from Lemma 9 (and using the fact that $\|P_t^S\varphi\|_\infty \leqslant \|\varphi\|_\infty$ for all $t \geqslant 0$),

$$\begin{aligned}\int_0^\infty \|P_t^S\Pi\varphi\|_\infty dt &\leqslant \int_0^1 \|\Pi\varphi\|_\infty dt + \int_1^\infty \|P_t^S\Pi\varphi\|_\infty dt \\ &\leqslant 2\|\varphi\|_\infty + C_2\int_1^\infty \|P_{t-1}^S\Pi\varphi\|_{L^2(\mu)} dt \\ &\leqslant 2\|\varphi\|_\infty + C_2\int_1^\infty e^{-(t-1)/C_2}\|\Pi\varphi\|_{L^2(\mu)} dt \\ &\leqslant 2(1+C_2^2)\|\varphi\|_\infty\,,\end{aligned}$$

and similarly, using the fact that $\|\nabla P_t^S\Pi\varphi\|_\infty = \|\nabla P_1 P_{t-1}^S\Pi\varphi\|_\infty \leq C_2\|P_{t-1}^S\Pi\phi\|_\infty$ for $t \geqslant 1$,

$$\begin{aligned}\int_0^\infty \|\nabla P_t^S\Pi\varphi\|_\infty dt &\leqslant \int_0^1 \frac{C_2}{\sqrt{t}}\|\Pi\varphi\|_\infty dt + C_2\int_1^\infty \|P_{t-1}^S\Pi\varphi\|_\infty dt \\ &\leqslant C_2\left(4 + 2(1+C_2^2)\right)\|\varphi\|_\infty\,.\end{aligned}$$

In particular $R_S\varphi$ and $\nabla R_S\varphi$ are well defined in $L^\infty(\mathbb{T}^D)$ for $\varphi \in \mathcal{C}^\infty(\mathbb{T}^D)$. Moreover, using the fact that, from Lemma 9, $\|P_t^S\Pi\varphi\|_\infty \leqslant C_2 e^{-(t-1)/C_2}\|\Pi\varphi\|_{L^2(\mu)} \to 0$ as $t \to +\infty$,

$$\begin{aligned}L_S R_S\varphi &= -\int_0^\infty L_S P_t^S\Pi\varphi dt \\ &= -\int_0^\infty \partial_t\left(P_t^S\Pi\varphi\right) dt \;=\; \Pi\varphi\,.\end{aligned}$$

The case of $R_S L_S$ is similar: since μ is invariant for L_S, $L_S\Pi = L_S = \Pi L_S$, and thus $P_t^S\Pi L_S = P_t^S L_S\Pi = \partial_t(P_t^S\Pi\varphi)$ for all $t \geqslant 0$.

As a consequence,

$$|\Delta R_S\varphi| \;\leqslant\; e^{\beta C_0}|e^{\beta S}\Delta R_S\varphi| \;\leqslant\; e^{\beta C_0}\left(\|e^{\beta S}\nabla(V-S)\cdot\nabla R_S\varphi\|_\infty + \|\Pi\varphi\|_\infty\right) \;\leqslant\; C_3\|\varphi\|_\infty$$

for some $C_3 > 0$ uniform over $S \in \mathcal{B}_{C_0}$, which yields the desired result. □

Lemma 11. *There exist $C_4 > 0$ such that for all $S_1, S_2 \in \mathcal{B}_{C_0}$ and $\varphi \in \mathcal{C}^\infty(\mathbb{T}^D)$,*

$$\|R_{S_1}\varphi - R_{S_2}\varphi\|_\infty + \|\nabla R_{S_1}\varphi - \nabla R_{S_2}\varphi\|_\infty \;\leqslant\; C_4\|\nabla S_1 - \nabla S_2\|_\infty\|\varphi\|_\infty.$$

Proof. From $R_S L_S = \Pi$,

$$\left(R_{S_1} - R_{S_2}\right)L_{S_1} + R_{S_2}\left(L_{S_1} - L_{S_2}\right) \;=\; 0.$$

Multiplying this equality by R_{S_1} on the right, and using that $R_S\Pi = R_S$, we get for all $\varphi \in \mathcal{C}^\infty(\mathbb{T}^D)$,

$$\left(R_{S_1} - R_{S_2}\right)\varphi \;=\; R_{S_2}\left(L_{S_2} - L_{S_1}\right)R_{S_1}\varphi\,.$$

Thus, from (26),

$$\begin{aligned}&\|\left(R_{S_1} - R_{S_2}\right)\varphi\|_\infty + \|\nabla R_{S_1}\varphi - \nabla R_{S_2}\varphi\|_\infty \\ &\quad\leqslant C_3\|\left(L_{S_2} - L_{S_1}\right)R_{S_1}\varphi\|_\infty \\ &\quad\leqslant C_3\|e^{\beta S_2}\nabla(S_1 - S_2)\cdot\nabla R_{S_1}\varphi\|_\infty + C_3\|(1 - e^{\beta(S_2 - S_1)})L_{S_1}R_{S_1}\varphi\|_\infty \\ &\quad\leqslant C_3 e^{\beta C_0}\|\nabla(S_1 - S_2)\|_\infty\|\nabla R_{S_1}\varphi\|_\infty + C_3 e^{2\beta C_0}\|S_2 - S_1\|_\infty\|L_{S_1}R_{S_1}\varphi\|_\infty\,.\end{aligned}$$

Conclusion follows from (26). Indeed, notice that $S_1 - S_2$ is a continuous function with integral zero, so that there exist $x \in \mathbb{T}^D$ such that $(S_1 - S_2)(x) = 0$, and then for all $y \in \mathbb{T}^D$

$$|(S_1 - S_2)(y)| \leqslant |x - y| \|\nabla(S_1 - S_2)\|_\infty$$

so that $\|S_1 - S_2\|_\infty \leqslant \sqrt{D}\|\nabla S_1 - \nabla S_2\|_\infty$.

□

Lemma 12. *There exists $C_5 > 0$ such that for all $k \geqslant 1$ and $\varphi \in \mathcal{C}^\infty(\mathbb{T}^D)$*

$$\|R_{A_{t_k}}\varphi - R_{A_{t_{k-1}}}\varphi\|_\infty \;\leqslant\; \frac{C_5}{k}\|\varphi\|_\infty.$$

Proof. From Lemma 6, $A_t \in \mathcal{B}_{C_0}$ for all $t \geqslant 0$, so that Lemma 11 applies. It remains to obtain a bound on $\|\nabla A_{t_k} - \nabla A_{t_{k-1}}\|_\infty$. In this proof, to simplify the notation, we write $\theta_k = \theta_{\nu_{t_k}}$ and $F_k = F_{\nu_{t_k}}$. Denoting by

$$m := \frac{\int_{t_k}^{t_{k+1}} \delta_{(Q_s,Z_s)} e^{\beta A_s(Z_s)} \mathrm{d}s}{\int_{t_k}^{t_{k+1}} e^{\beta A_s(Z_s)} \mathrm{d}s} \quad \text{and} \quad p := \frac{\int_{t_k}^{t_{k+1}} e^{\beta A_s(Z_s)} \mathrm{d}s}{\int_0^{t_{k+1}} e^{\beta A_s(Z_s)} \mathrm{d}s},$$

it holds that

$$\nu_{t_{k+1}} = (1-p)\nu_{t_k} + pm.$$

In particular, for some $c, c' > 0$, for all $z \in \mathbb{T}^d$ and $k \in \mathbb{N}$,

$$|\theta_{k+1}(z) - \theta_k(z)| \;= \frac{p}{1+\lambda}\left|\int_{(q,y)\in\mathbb{T}^p\times\mathbb{T}^d} K(z,y)(\mathrm{d}m(q,y) - \mathrm{d}\nu_{t_k}(q,y))\right| \;\leqslant\; cp \;\leqslant\; \frac{c'}{k},$$

where we used that $A_s \in \mathcal{B}_{C_0}$ for all $s \in [0, t_{k+1}]$. The same argument also works for the derivatives of θ_{ν_t}, for F_{ν_t} and its derivatives, so that for any multi-index $\alpha \in \mathbb{N}^d$, there exists a constant C_α such that for all $k \geqslant 1$,

$$\|\partial^\alpha F_{k+1} - \partial^\alpha F_k\|_\infty + \|\partial^\alpha \theta_{k+1} - \partial^\alpha \theta_k\|_\infty \;\leqslant\; \frac{C_\alpha}{k}.$$

Now, from the Euler equations satisfied by A_{t_k} and $A_{t_{k+1}}$, we get

$$\nabla \cdot \left(\theta_k \nabla\left(A_{t_k} - A_{t_{k+1}}\right)\right) \;=\; \nabla \cdot \left(\nabla A_{t_{k+1}}\left(\theta_{k+1} - \theta_k\right)\right) - \nabla \cdot \left(\theta_{k+1}F_{k+1} - \theta_k F_k\right). \tag{27}$$

Multiplying this equation by $A_{t_k} - A_{t_{k+1}}$, integrating and using Lemma 6 and the lower bound on θ_k, we get

$$\int_{\mathbb{T}^d} |\nabla\left(A_{t_k} - A_{t_{k+1}}\right)(z)|^2 \mathrm{d}z \;\leqslant\; \frac{c}{k^2}$$

for some $c > 0$. Next, differentiating (27), multiplying it by derivatives of $A_{t_k} - A_{t_{k+1}}$, integrating and using by induction the previous estimates, we obtain in fact that

$$\int_{\mathbb{T}^d} |\nabla\partial^\alpha\left(A_{t_k} - A_{t_{k+1}}\right)(z)|^2 \mathrm{d}z \;\leqslant\; \frac{c_\alpha}{k^2}$$

for some $c_\alpha > 0$ for all $\alpha \in \mathbb{N}^d$, and Sobolev embeddings then yield the conclusion. □

3.3 Proof of the main results

In this section we denote $\nu(\varphi) = \int_{\mathbb{T}^D} \varphi \mathrm{d}\nu$ the expectation of an observable φ with respect to a probability measure ν.

Proof of Proposition 7. In the following, we use the same notation C for various constants. For all $t \geqslant 0$ and $\varphi \in \mathcal{C}^\infty\left(\mathbb{T}^D\right)$,

$$\begin{aligned}\nu_t(\varphi) - \mu(\varphi) &= \frac{1}{\tau(t)} \int_0^t e^{-\beta A_s(Z_s)} \Pi\varphi(X_s) \mathrm{d}s \\ &= \frac{1}{\tau(t)} \int_0^t e^{-\beta A_s(Z_s)} L_{A_s} R_{A_s} \varphi(X_s) \mathrm{d}s \,.\end{aligned}$$

To alleviate notations, write $\tilde{\varphi}_s = R_{A_s}\varphi$. For $k \in \mathbb{N}$ and $t \in [t_k, t_{k+1})$, from (5) by Itô's formula,

$$\tilde{\varphi}_t(X_t) - \tilde{\varphi}_{t_k}(X_{t_k}) = \int_{t_k}^t e^{-\beta A_s(Z_s)} L_{A_s} \tilde{\varphi}_s(X_s) \mathrm{d}s + \sqrt{\frac{2}{\beta}} \int_{t_k}^t \nabla \tilde{\varphi}_s(X_s) \mathrm{d}B_s \,,$$

so that

$$\begin{aligned}\tau(t)\left(\nu_t(\varphi) - \mu(\varphi)\right) &= \int_0^t e^{-\beta A_s(Z_s)} L_{A_s} \tilde{\varphi}_s(X_s) \mathrm{d}s \;=\; \tilde{\varphi}_t(X_t) - \tilde{\varphi}_0(X_0) \\ &\quad + \sum_{0<t_k\leqslant t} \left(\tilde{\varphi}_{t_{k-1}}(X_{t_k}) - \tilde{\varphi}_{t_k}(X_{t_k})\right) - \sqrt{\frac{2}{\beta}} \int_0^t \nabla \tilde{\varphi}_s(X_s) \mathrm{d}B_s \,.\end{aligned} \tag{28}$$

Recall that, from Lemma 6, there exists $C > 0$ such that almost surely $\tau(t) \geqslant t/C$ for all $t \geqslant 0$. Together with Lemma 12, we get that there exists $C > 0$ such that, almost surely, for all $t > 0$ and $\varphi \in \mathcal{C}^\infty\left(\mathbb{T}^D\right)$,

$$\frac{1}{\tau(t)} \left(|\tilde{\varphi}_t(X_t) - \tilde{\varphi}_0(X_0)| + \sum_{0<t_k<t} |\tilde{\varphi}_{t_k}(X_{t_k}) - \tilde{\varphi}_{t_{k-1}}(X_{t_k})| \right) \leqslant \frac{C \ln(1+t)}{t} \|\varphi\|_\infty \,. \tag{29}$$

Moreover, applying Itô's isometry,

$$\begin{aligned}\mathbb{E}\left(\left| \int_0^t \nabla \tilde{\varphi}_s(X_s) \mathrm{d}B_s \right|^2 \right) &= \mathbb{E}\left(\int_0^t |\nabla \tilde{\varphi}_s(X_s)|^2 \, \mathrm{d}s \right) \\ &\leqslant t C_3^2 \|\varphi\|_\infty^2 \,,\end{aligned}$$

where we used (26). As a consequence, there exists $C > 0$ such that

$$\mathbb{E}\left(|\nu_t(\varphi) - \mu(\varphi)|^2 \right) \leqslant \frac{C}{t} \|\varphi\|_\infty^2 \tag{30}$$

for all $t > 0$ and $\varphi \in \mathcal{C}^\infty\left(\mathbb{T}^D\right)$. As in the proof of [7, Lemma 5.1], this implies the almost sure weak convergence of ν_t to μ as follows. Indeed, for all $r > 0$, the Borel-Cantelli Lemma yields the almost sure convergence of $\nu_{\exp(nr)}(\varphi)$ toward $\mu(\varphi)$ as $n \to +\infty$, $n \in \mathbb{N}$, so that

$$\mathbb{P}\left(\forall k \in \mathbb{N}_*, \; \nu_{\exp(n/k)}(\varphi) \underset{n\to+\infty}{\longrightarrow} \mu(\varphi) \right) = 1 \,. \tag{31}$$

Moreover, using the almost sure bounds $t/C \leqslant \tau(t) \leqslant Ct$ and $\|\exp(-\beta A_s)\|_\infty \leqslant C$ for some $C > 0$, we get that there exists $C, C' > 0$ such that for all $t \geqslant s > 0$,

$$\begin{aligned} |\nu_t(\varphi) - \nu_s(\varphi)| &\leqslant C\left(\left|\frac{1}{\tau(t)} - \frac{1}{\tau(s)}\right| s + \frac{|t-s|}{\tau(t)}\right)\|\varphi\|_\infty \\ &\leqslant C\left(\frac{|\tau(t)-\tau(s)|s}{\tau(t)\tau(s)} + \frac{|t-s|}{\tau(t)}\right)\|\varphi\|_\infty \\ &\leqslant C'\frac{|t-s|}{t}\|\varphi\|_\infty\,. \end{aligned}$$

As a consequence, for all $\varphi \in \mathcal{C}^\infty\left(\mathbb{T}^D\right)$, $t \mapsto \nu_{\exp(t)}(\varphi)$ is almost surely $C'\|\varphi\|_\infty$-Lipschitz, and in particular

$$\mathbb{P}\left(\forall k \in \mathbb{N}_*, \forall t \geqslant 0,\ |\nu_{\exp(t)}(\varphi) - \nu_{\exp(\lfloor tk\rfloor/k)}(\varphi)| \leqslant \frac{C'\|\varphi\|_\infty}{k}\right) \;=\; 1\,. \tag{32}$$

The almost sure convergence of $\nu_t(\varphi)$ to $\mu(\varphi)$ for a given φ then follows from

$$\begin{aligned} \left\{\nu_t(\varphi) \underset{t\to+\infty}{\longrightarrow} \mu(\varphi)\right\} &= \bigcap_{k\in\mathbb{N}_*}\left\{\limsup_{t\to\infty}|\nu_{\exp(t)} - \mu(\varphi)| \leqslant \frac{C'\|\varphi\|_\infty}{k}\right\} \\ &\supset \bigcap_{k\in\mathbb{N}_*}\left(\left\{\sup_{t\geqslant 0}|\nu_{\exp(t)}(\varphi) - \nu_{\exp(\lfloor tk\rfloor/k)}(\varphi)| \leqslant \frac{C'\|\varphi\|_\infty}{k}\right\} \cap \left\{\nu_{\exp(\lfloor tk\rfloor/k)} \underset{t\to+\infty}{\longrightarrow} \mu(\varphi)\right\}\right)\,, \end{aligned}$$

the last event having probability 1 from (31) and (32). Considering a sequence $(\varphi_k)_{k\in\mathbb{N}}$ of $\mathcal{C}^\infty\left(\mathbb{T}^D\right)$ functions that is dense in $\mathcal{C}\left(\mathbb{T}^D\right)$, we get that

$$\mathbb{P}\left(\nu_t(\varphi_k) \underset{t\to+\infty}{\longrightarrow} \mu(\varphi_k)\ \forall k \in \mathbb{N}\right) \;=\; 1\,,$$

so that almost surely ν_t converges weakly to μ as $t \to +\infty$. This concludes the proof of Proposition 7, hence of Theorem 2. □

Proof of Theorem 3. The first claim of the theorem has already been established in the proof of Proposition 7, see (30). Fix $\varphi \in \mathcal{C}^\infty\left(\mathbb{T}^D\right)$. We have seen in the proof of Proposition 7 (see (28) and (29)) that, denoting $\tilde{\varphi}_s = R_{A_s}\varphi$,

$$\nu_t(\varphi) - \mu(\varphi) \;=\; \varepsilon_t - \frac{1}{\tau(t)}\sqrt{\frac{2}{\beta}}\int_0^t \nabla\tilde{\varphi}_s(X_s)\mathrm{d}B_s$$

for some ε_t such that almost surely $|\varepsilon_t| \leqslant C\ln(1+t)/t$ for all $t > 0$ for some $C > 0$. The martingale part having zero expectation, the bias is bounded as

$$|\mathbb{E}(\nu_t(\varphi) - \mu(\varphi))|^2 = |\mathbb{E}(\varepsilon_t)|^2 \leqslant \frac{C^2\ln^2(1+t)}{t^2} = \underset{t\to+\infty}{o}\left(\frac{1}{t}\right)\,.$$

In other words, the asymptotic mean-square error is only due to the asymptotic variance, which is itself only due to the martingale part of $\nu_t(\varphi) - \mu(\varphi)$.

Remark that, as a corollary of Theorem 2, $\|A_t - A_*\|_\infty \to 0$ almost surely. Together with the uniform bounds of Lemma 6 and the weak convergence of ν_t to μ, we get that

$$\begin{aligned}
\frac{t}{\tau(t)} \;=\; \frac{t-\sqrt{t}}{\tau(t)} + \underset{t\to+\infty}{o}(1) \quad &= \quad \frac{1}{\tau(t)}\int_{\sqrt{t}}^t e^{-\beta A_*(Z_s)} e^{\beta A_*(Z_s)}\mathrm{d}s + \underset{t\to+\infty}{o}(1) \\
&= \quad \frac{1}{\tau(t)}\int_{\sqrt{t}}^t e^{-\beta A_s(Z_s)} e^{\beta A_*(Z_s)}\mathrm{d}s + \underset{t\to+\infty}{o}(1) \\
&= \quad \frac{1}{\tau(t)}\int_0^t e^{-\beta A_s(Z_s)} e^{\beta A_*(Z_s)}\mathrm{d}s + \underset{t\to+\infty}{o}(1) \\
&\underset{t\to+\infty}{\longrightarrow} \quad \mu\left(e^{\beta A_*\circ\xi}\right) \;=:\; \kappa
\end{aligned}$$

almost surely. As a consequence,

$$\begin{aligned}
t\mathbb{E}\left(|\nu_t(\varphi)-\mu(\varphi)|^2\right) &= \frac{2t}{\beta}\mathbb{E}\left(\frac{1}{\tau^2(t)}\left|\int_0^t \nabla\tilde{\varphi}_s(X_s)\mathrm{d}B_s\right|^2\right) + \underset{t\to+\infty}{o}(1) \\
&= \frac{2\kappa^2}{\beta t}\mathbb{E}\left(\left|\int_0^t \nabla\tilde{\varphi}_s(X_s)\mathrm{d}B_s\right|^2\right) + \underset{t\to+\infty}{o}(1) \\
&= \frac{2\kappa^2}{\beta t}\int_0^t \mathbb{E}\left(|\nabla\tilde{\varphi}_s(X_s)|^2\right)\mathrm{d}s + \underset{t\to+\infty}{o}(1)\,.
\end{aligned}$$

From Lemma 11, $\|\nabla\tilde{\varphi}_s - \nabla R_{A_*}\varphi\|_\infty \leqslant C_4\|\nabla A_s - \nabla A_*\|_\infty\|\varphi\|_\infty$, which together with Theorem 2 and (26) yields

$$\begin{aligned}
t\mathbb{E}\left(|\nu_t(\varphi)-\mu(\varphi)|^2\right) \quad &= \quad \frac{2\kappa^2}{\beta t}\int_0^t \mathbb{E}\left(|\nabla R_{A_*}\varphi(X_s)|^2\right)\mathrm{d}s + \underset{t\to+\infty}{o}(1) \\
&= \quad \frac{2\kappa^2}{\beta t}\mathbb{E}\left(\int_0^t |\nabla R_{A_*}\varphi(X_s)|^2\, e^{-\beta(A_s(Z_s)-A_*(Z_s))}\mathrm{d}s\right) + \underset{t\to+\infty}{o}(1) \\
&= \quad \frac{2\kappa}{\beta}\mathbb{E}\left(\nu_t\left(|\nabla R_{A_*}\varphi|^2\, e^{\beta A_*\circ\xi}\right)\right) + \underset{t\to+\infty}{o}(1) \\
&\underset{t\to+\infty}{\longrightarrow} \quad \frac{2\kappa}{\beta}\mu\left(|\nabla R_{A_*}\varphi|^2\, e^{\beta A_*\circ\xi}\right)\,.
\end{aligned}$$

In other words, the asymptotic variance is

$$\frac{2}{\beta}\mu\left(e^{\beta A_*\circ\xi}\,|\nabla\psi|^2\right)\mu\left(e^{\beta A_*\circ\xi}\right)$$

where $\psi = R_{A_*}\varphi$ solves $L_{A_*}\psi = \Pi\varphi$, which reads

$$\left(\frac{1}{\beta}\Delta - \nabla\left(V - A_*\circ\xi\right)\nabla\right)\psi \;=\; e^{-\beta A_*\circ\xi}\Pi\varphi\,.$$

It remains to see that we get the same formula for the asymptotic variance of $\sqrt{t}\,(\nu_t^*(\varphi)-\mu(\varphi))$. The computations are similar, so we only sketch the main points. Denoting $\tau_*(t) = \int_0^t \exp(-\beta A_*(Z_s^*))\mathrm{d}s$, as in Proposition 7,

$$\begin{aligned}
\tau_*(t)\,(\nu_t^*(\varphi)-\mu(\varphi)) &= \int_0^t e^{-\beta A_*(Z_s^*)}\Pi\varphi(X_s^*)\mathrm{d}s \\
&= \int_0^t e^{-\beta A_*(Z_s^*)} L_{A_*}R_{A_*}\varphi(X_s^*)\mathrm{d}s \\
&= R_{A_*}\varphi(X_t^*) - R_{A_*}\varphi(X_0^*) - \sqrt{\frac{2}{\beta}}\int_0^t \nabla R_{A_*}\varphi(X_s)\mathrm{d}B_s\,.
\end{aligned}$$

Again, from the almost sure bound $\tau_*(t) \geqslant t/C$ for some $C > 0$, Itô's isometry and the bounds on $R_{A_*}\varphi$ given by (26), we get that there exists $C > 0$ such that for all $t > 0$ and $\varphi \in \mathcal{C}(\mathbb{T}^D)$,

$$\mathbb{E}\left(\left|\nu_t^*(\varphi) - \mu(\varphi)\right|^2\right) \leqslant \frac{C}{t}\|\varphi\|_\infty^2 .$$

The proof that this implies the almost sure weak convergence of ν_t^* to μ is similar to the end of the proof of Proposition 7. As a corollary,

$$\frac{t}{\tau_*(t)} = \nu_t^*\left(e^{\beta A_* \circ \xi}\right) \underset{t\to+\infty}{\longrightarrow} \kappa .$$

Finally,

$$\begin{aligned}
t\mathbb{E}\left(|\nu_t(\varphi) - \mu(\varphi)|^2\right) &= \frac{2t}{\beta}\mathbb{E}\left(\frac{1}{\tau_*^2(t)}\left|\int_0^t \nabla R_{A_*}\varphi(X_s^*)\mathrm{d}B_s\right|^2\right) + \underset{t\to+\infty}{o}(1) \\
&= \frac{2\kappa^2}{\beta t}\mathbb{E}\left(\left|\int_0^t \nabla R_{A_*}\varphi(X_s^*)\mathrm{d}B_s\right|^2\right) + \underset{t\to+\infty}{o}(1) \\
&= \frac{2\kappa^2}{\beta t}\int_0^t \mathbb{E}\left(|\nabla R_{A_*}\varphi(X_s^*)|^2\right)\mathrm{d}s + \underset{t\to+\infty}{o}(1) \\
&= \frac{2\kappa^2}{\beta t}\mathbb{E}\left(\tau_*(t)\nu_t^*\left(|\nabla R_{A_*}\varphi|^2 e^{\beta A_*\circ\xi}\right)\right)\mathrm{d}s + \underset{t\to+\infty}{o}(1) \\
\underset{t\to+\infty}{\longrightarrow} &\ \frac{2\kappa}{\beta}\mu\left(|\nabla R_{A_*}\varphi|^2 e^{\beta A_*\circ\xi}\right) .
\end{aligned}$$

□

4 Consistency of the tensor approximation

This section is devoted to the proof of Propositions 1 and 4 and Theorem 5. In all this section, Assumption 1 holds and we write $\theta = \theta_\nu$, $F = F_\nu$, $\mathcal{J} = \mathcal{J}_\nu$ for some fixed $\nu \in \mathcal{P}(\mathbb{T}^p \times \mathbb{T}^d)$. Recall that $F, \theta \in \mathcal{C}^\infty(\mathbb{T}^d)$, that θ is a positive probability density on $\mathbb{T}^d$, and that the minimizers of $\mathcal{J}$ in H are exactly the minimizers of $\widetilde{\mathcal{J}}$ in H, where for all $f \in H$,

$$\widetilde{\mathcal{J}}(f) = \int_{\mathbb{T}^d} |F(z) - \nabla f(z)|^2 \theta(z)\mathrm{d}z ,$$

the link between $\mathcal{J}$ and $\widetilde{\mathcal{J}}$ begin given by (13).

Since θ is bounded from above and below by positive constants, the weighted spaces $L^2(\mathbb{T}^d;\theta)$ and $H^1(\mathbb{T}^d;\theta)$ are equal to the flat spaces $L^2(\mathbb{T}^d;\mathrm{d}z)$ and $H^1(\mathbb{T}^d;\mathrm{d}z)$. We endow H (whose definition is given in (6), with the norm

$$\|f\| = \sqrt{(1+\lambda)\int_{\mathbb{T}^d} |\nabla f(z)|^2 \theta(z)\mathrm{d}z},$$

which is indeed a norm, equivalent to the usual H^1 norm from the Poincaré-Wirtinger inequality: there exists $C > 0$ such that for all $f \in H$,

$$\begin{aligned}
\int_{\mathbb{T}^d} f^2(z)\theta(z)\mathrm{d}z &\leqslant \|\theta\|_\infty \int_{\mathbb{T}^d} f^2(z)\mathrm{d}z \\
&\leqslant C\|\theta\|_\infty \int_{\mathbb{T}^d} |\nabla f(z)|^2 \mathrm{d}z \\
&\leqslant C\|\theta\|_\infty \|\theta^{-1}\|_\infty \int_{\mathbb{T}^d} |\nabla f(z)|^2 \theta(z)\mathrm{d}z.
\end{aligned}$$

The scalar product associated with $\|\cdot\|$ is denoted by $\langle\cdot\rangle$. The choice of such a norm is motivated by the fact that, denoting by $\mathcal{J}'$ the differential of $\mathcal{J}$, then for all $f, g \in H$,

$$\begin{aligned}\mathcal{J}(f) &= \mathcal{J}(0) + \mathcal{J}'(0)\cdot f + \|f\|^2 \\ \mathcal{J}'(f)\cdot g &= \mathcal{J}'(0)\cdot g + 2\langle f, g\rangle\,.\end{aligned}$$

Proposition 1 is then a direct consequence of the strict convexity of $\mathcal{J}$. The unique minimizer f_* of $\mathcal{J}$ over H being a minimizer of $\widetilde{\mathcal{J}}$, it satisfies $\widetilde{\mathcal{J}}'(f_*) = 0$, which reads

$$\forall g \in H\,, \qquad \int F(z)\cdot\nabla g(z)\theta(z)\mathrm{d}z \;=\; \int \nabla f_*(z)\cdot\nabla g(z)\theta(z)\mathrm{d}z$$

Moreover, using that $\mathcal{J}'(f_*) = 0$, we get that $\mathcal{J}'(0)\cdot g = -2\langle f_*, g\rangle$ for all $g \in H$ and then $\mathcal{J}(0) = \mathcal{J}(f_*) - \mathcal{J}'(0)\cdot f_* - \|f_*\|^2 = \mathcal{J}(f_*) + \|f_*\|^2$. As a consequence, for all $g \in H$,

$$\mathcal{J}(g) \;=\; \mathcal{J}(0) + \mathcal{J}'(0)\cdot g + \|g\|^2 \;=\; \mathcal{J}(f_*) + \|f_* - g\|^2\,. \tag{33}$$

Proof of Proposition 4. Let $f \in H$ and $i \in [\![1, d]\!]$. If $\mathcal{J}(f) = \inf\{\mathcal{J}(f+g),\ g \in \Sigma_i\}$ then the result is correct since $0 \in \Sigma_i$ is a minimizer over Σ_i. Suppose now that 0 is not a minimizer, i.e. that $\mathcal{J}(f) > \inf\{\mathcal{J}(f+g),\ g \in \Sigma_i\}$ and consider a minimizing sequence $(g^{(l)})_{l\in\mathbb{N}}$ in Σ_i such that $\mathcal{J}\left(f + g^{(l)}\right)$ converges to $\inf\{\mathcal{J}(f+g),\ g \in \Sigma_i\}$ as l goes to infinity. For l large enough, $\mathcal{J}\left(f + g^{(l)}\right) < \mathcal{J}(f)$ so that $g^{(l)} \neq 0$, and thus up to an extraction we suppose that $g^{(l)} \neq 0$ for all $l \in \mathbb{N}$. Moreover the sequence is bounded in H^1 and thus, up to the extraction of a subsequence, we suppose that it weakly converges in H^1 to some $g^* \in H$. The function $H \ni g \mapsto \mathcal{J}(f+g)$ being convex on H, it is weakly lower semi-continuous, so that

$$\mathcal{J}(g_*) \;\leqslant\; \inf_{g\in\Sigma_i} \mathcal{J}(f+g)\,.$$

For all $l \in \mathbb{N}$, there exist $r_1^{(l)}, \cdots, r_d^{(l)} \in H^1(\mathbb{T})$ such that $g^{(l)} = \bigotimes_{j=1}^d r_j^{(l)}$. Since $g^{(l)} \neq 0$, we can normalize the $r_j^{(l)}$'s so that $\|r_j^{(l)}\|_{L^2(\mathbb{T})} = 1$ for all $j \neq i$ and $l \in \mathbb{N}$. As a consequence, up to the extraction of a subsequence, for all $1 \leq j \neq i \leq d$, there exists $r_j^* \in L^2(\mathbb{T})$ such that the sequence $(r_j^{(l)})_{l\in\mathbb{N}}$ weakly converges to r_j^* in $L^2(\mathbb{T})$. Now, since the sequence $g^{(l)}$ is bounded in H and

$$\|\nabla g^{(l)}\|^2_{L^2(\mathbb{T}^d)} \;=\; \sum_{j=1}^d \|\partial_{z_j} r_j^{(l)}\|^2_{L^2(\mathbb{T})} \prod_{h\neq j} \|r_h^{(l)}\|^2_{L^2(\mathbb{T})},$$

we get that the sequence $(r_i^{(l)})_{l\in\mathbb{N}}$ is bounded in $H^1(\mathbb{T})$ (since $\int_{\mathbb{T}} r_i^{(l)} = 0$ for all $l \in \mathbb{N}$). Thus, up to the extraction of another subsequence, the sequence $(r_i^{(l)})_{l\in\mathbb{N}}$ weakly converges in $H^1(\mathbb{T})$ to some $r_i^* \in H^1(\mathbb{T})$ such that $\int_{\mathbb{T}} r_i^* = 0$. From [32, Lemma 2], $(g^{(l)})_{l\in\mathbb{N}}$ converges in the distributional sense to $\bigotimes_{j=1}^d r_j^*$. Thus, $g^* = \bigotimes_{j=1}^d r_j^*$ and since $g^* \neq 0$, this implies that for all $1 \leq j \leq d$, $r_j^* \neq 0$. Finally, since

$$\|\nabla g^*\|^2_{L^2(\mathbb{T}^d)} = \sum_{j=1}^d \|\partial_{z_j} r_j^*\|^2_{L^2(\mathbb{T})} \prod_{h\neq j} \|r_h^*\|^2_{L^2(\mathbb{T})}$$

is a finite quantity, this implies that for all $1 \leq j \neq i \leq d$,

$$\|\partial_{z_j} r_j^*\|_{L^2(\mathbb{T})} \leq \frac{\|\nabla g^*\|^2_{L^2(\mathbb{T}^d)}}{\prod_{h\neq j}\|r_h^*\|^2_{L^2(\mathbb{T})}} < +\infty,$$

and thus $r_j^* \in H^1(\mathbb{T})$. This implies that $g_* \in \Sigma_i$ and yields the desired result. □

Remark 13. *The problem would be ill-posed if we were to try and minimize $\mathcal{J}(f+g-\int_{\mathbb{T}^d} g)$ over all $g \in \Sigma := \{r_1 \otimes \cdots \otimes r_d,\ r_j \in H^1(\mathbb{T})$ for all $1 \leq j \leq d\}$. This is the reason why we introduced the condition that one of the r_j's has zero mean. Indeed, consider the situation where $d=2$, $f=0$ and $F(z) = (a'(z_1), b'(z_2))$ for some smooth functions $a, b : \mathbb{T} \to \mathbb{R}$ with zero mean. Then, the minimum of $\mathcal{J}$ over H is 0, and only attained at $f^*(z_1,z_2) = a(z_1)+b(z_2)$, which is not of the form $g - \int_{\mathbb{T}^d} g$ for some $g \in \Sigma$. Nevertheless, the sequence $(g^{(l)})_{l\in\mathbb{N}^*}$ defined by: for all $l \in \mathbb{N}^*$, $g^{(l)} = r_1^{(l)} r_2^{(l)}$ with*

$$r_1^{(l)}(z_1) = 1 + \frac{a(z_1)}{l}, \qquad r_2^{(l)}(z_2) = l + b(z_2)$$

is a minimizing sequence. Indeed, the sequence $\left(g^{(l)} - \int_{\mathbb{T}^d} g^{(l)}\right)_{l\in\mathbb{N}^*}$ weakly converges to f^*. In other words, the set $\left\{g - \int_{\mathbb{T}^d} g,\ g \in \Sigma\right\}$ is not weakly closed in H.

Proposition 4 proves that all the iterations of Algorithm 2 are well-defined. In the following, we consider a sequence $(f_n)_{n\in\mathbb{N}}$ given by the latter and $g_n = f_{n+1} - f_n$ for $n \in \mathbb{N}$. The general idea of the proof of Theorem 5 is that, if the sequence $(f_n)_{n\in\mathbb{N}}$ converges to some f_∞ in H, it holds that $\mathcal{J}'(f_\infty)\cdot g = 0$ for all $g \in \cup_{i=1}^d \Sigma_i$, and a density argument enables to conclude. Nevertheless, remark that Σ_i is not a vector space and that its elements all have null integral, so that one should be careful. In the following, we essentially adapt the arguments of [12].

Lemma 14. *For all $g \in \cup_{i=1}^d \Sigma_i$ and $n \in \mathbb{N}$,*

$$|\mathcal{J}'(f_n)\cdot g| \ \leqslant\ 6\|g\| \sum_{j=0}^{d-1} \|g_{n+j}\|.$$

Proof. Let $i \in [\![1,d]\!]$ be such that $g \in \Sigma_i$. Let $n_i \in [\![n, n+d-1]\!]$ be such that $g_{n_i} \in \Sigma_i$. We bound, first,

$$|\mathcal{J}'(f_n)\cdot g| \ \leqslant\ |\mathcal{J}'(f_{n_i})\cdot g| + 2\left|\langle g, f_{n_i} - f_n\rangle\right|.$$

The second term of the right hand side is bounded by $2\|g\|\sum_{j=n}^{n_i-1}\|g_j\|$. To deal with the first one, note that, even though g_{n_i} is a minimizer of $\mathcal{J}(f_{n_i}+\cdot)$ over Σ_i, it is not necessarily true that $\mathcal{J}'(f_{n_i}+g_{n_i})\cdot g = 0$, since Σ_i is not a vector space. We follow the proof of [12, Proposition 3.3]. By convexity of

$$t \in \mathbb{R} \ \mapsto\ \psi(t) \ :=\ \mathcal{J}\left(f_{n_i} + g + t(g_{n_i} - g)\right),$$

and since $t=0$ minimizes $\psi(t)$, we get

$$\psi'(0) \ \leqslant\ \psi(1) - \psi(0) \ \leqslant\ 0,$$

which reads

$$\mathcal{J}'\left(f_{n_i}+g\right)\cdot g \ \geqslant\ \mathcal{J}'\left(f_{n_i}+g\right)\cdot g_{n_i}.$$

Hence,

$$\begin{aligned} -\mathcal{J}'(f_{n_i})\cdot g &= -\mathcal{J}'(f_{n_i}+g)\cdot g + 2\|g\|^2 \\ &\leqslant -\mathcal{J}'(f_{n_i}+g)\cdot g_{n_i} + 2\|g\|^2 \\ &\leqslant -\mathcal{J}'(f_{n_i}+g_{n_i})\cdot g_{n_i} + 2\langle g_{n_i}, g_{n_i} - g\rangle + 2\|g\|^2. \end{aligned}$$

Now, 1 being a minimizer over $\mathbb{R}$ of $t \mapsto \mathcal{J}(f_{n_i} + t g_{n_i})$, $\mathcal{J}'(f_{n_i} + g_{n_i}) \cdot g_{n_i} = 0$, so that

$$-\mathcal{J}'(f_{n_i}) \cdot g \;\leqslant\; 2\left(\|g_{n_i}\|^2 + \|g\|\|g_{n_i}\| + \|g\|^2\right).$$

When applied to $\widetilde{g} = \pm\|g_{n_i}\| g/\|g\|$, this inequality yields

$$|\mathcal{J}'(f_{n_i}) \cdot g| \;\leqslant\; 6\|g\|\|g_{n_i}\|,$$

which concludes the proof. □

Proposition 15. *Let f_* be the unique minimizer of $\mathcal{J}$ over H. Then*

$$\|f_n - f_*\| \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. As in the previous proof,

$$0 \;=\; \mathcal{J}'(f_n + g_n) \cdot g_n \;=\; 2(1+\lambda)\int \nabla g_n \cdot (\nabla f_{n+1} - F)\theta = 0$$

for all $n \in \mathbb{N}$, so that,

$$\mathcal{J}(f_n) - \mathcal{J}(f_{n+1}) \;=\; \mathcal{J}(f_{n+1} - g_n) - \mathcal{J}(f_{n+1}) \;=\; \|g_n\|^2. \tag{34}$$

In particular, since $(\mathcal{J}(f_n))_{n\geqslant 0}$ is a decreasing sequence bounded from below,

$$\sum_{n\geqslant 0} \|g_n\|^2 < \infty. \tag{35}$$

Together with Lemma 14 and the fact $\mathcal{J}'(f_*) = 0$, this implies that for all $g \in \mathrm{Span}\left(\cup_{i=1}^d \Sigma_i\right)$,

$$2\langle f_* - f_n, g\rangle \;=\; \mathcal{J}'(f_*) \cdot g - \mathcal{J}'(f_n) \cdot g \underset{n\to\infty}{\longrightarrow} 0.$$

Now, for all $r_1, \cdots, r_d \in \mathcal{C}^\infty(\mathbb{T})$, denoting by $h := \bigotimes r_i \in \mathcal{C}^\infty(\mathbb{T}^d)$ (note that we do not have necessarily that $\int_{\mathbb{T}^d} h = 0$), we can write

$$h - \int_{\mathbb{T}^d} h \;=\; \sum_{j=1}^d \left(r_j - \int_{\mathbb{T}} r_j\right)\left(\prod_{l<j}\int_{\mathbb{T}} r_l\right)\prod_{l>j} r_l\,,$$

which proves that $h - \int_{\mathbb{T}^d} h \in \mathrm{Span}\left(\cup_{i=1}^d \Sigma_i\right)$. As a consequence,

$$(1+\lambda)\int_{\mathbb{T}^d} \theta\, \nabla(f_* - f_n) \cdot \nabla h \;=\; \langle f_* - f_n, h - \int_{\mathbb{T}^d} h\rangle \underset{n\to\infty}{\longrightarrow} 0\,,$$

As a consequence, the limit f_∞ of any convergent (in the weak sense in H) subsequence of $(f_n)_{n\in\mathbb{N}}$ necessarily satisfies that

$$\int_{\mathbb{T}^d} \theta\, \nabla(f_* - f_\infty) \cdot \nabla h = 0,$$

for any tensor product function $h = \bigotimes r_i$, with $r_i \in \mathcal{C}^\infty(\mathbb{T})$ for all $1 \leq i \leq d$. By [12, Lemma 2.1], this implies that $f_\infty = f_*$. On the other hand, since $(\|f_n\|)_{n\in\mathbb{N}}$ is bounded, all its subsequences admits weak convergent subsequences, and the fact they all have the same limit f_* proves that the whole sequence $(f_n)_{n\in\mathbb{N}}$ weakly converges in H to f_*. In particular

$$\langle f_* - f_n, f_*\rangle \underset{n\to\infty}{\longrightarrow} 0. \tag{36}$$

Thus, it only remains to prove that $(\langle f_* - f_n, f_n\rangle)_{n\in\mathbb{N}}$ also converges to zero as n goes to infinity to obtain the *strong* convergence of the sequence $(f_n)_{n\in\mathbb{N}}$ to f_* in H. From Lemma 14,

$$\begin{aligned} 2|\langle f_* - f_n, f_n\rangle| &= |\mathcal{J}'(f_n)\cdot f_n| \\ &\leqslant \sum_{k=0}^{n} |\mathcal{J}'(f_n)\cdot g_k| \\ &\leqslant 6\left(\sum_{k=0}^{n-1}\|g_k\|\right)\sum_{j=n}^{n+d-1}\|g_j\| \\ &\leqslant 6\sqrt{nda_n\sum_{k=0}^{\infty}\|g_k\|^2}. \end{aligned}$$

with $a_n = \sum_{j=n}^{n+d-1}\|g_j\|^2$. Using (35), since $\sum_{n\in\mathbb{N}} a_n \leq d\sum_{n\in\mathbb{N}}\|g_n\|^2 < \infty$, there exists an extracted subsequence $(n_k)_{k\geqslant 1}$ such that $(n_k a_{n_k})_{k\geq 1}$ converges to 0 as k goes to infinity. As a consequence, $(\langle f_* - f_{n_k}, f_{n_k}\rangle)_{k\geq 1}$ goes to zero as $k\to\infty$, and thus so does $(\|f_{n_k} - f_*\|)_{k\geq 1}$ by (36). Finally, from (33), the sequence $(\|f_n - f_*\|)_{n\in\mathbb{N}}$ is non-increasing, so that the whole sequence goes to zero. Hence the result. □

5 Discussion and variations

For the sake of clarity, the TABF algorithm defined in Section 2.3 has been kept relatively simple, and there is obviously room for many variations or fine-tuning. We list here a few of them.

5.1 Extended ABF

Consider general reaction coordinates $\xi : \mathbb{T}^D \to \mathcal{M}$ where $\mathcal{M}$ is a submanifold of $\mathbb{R}^d$ or $\mathbb{T}^d$. In the Extended ABF (EABF) algorithm introduced in [38] (see also [39, 21]), the state space is extended to $\mathbb{T}^D\times\mathcal{M}$ with the addition of auxiliary variables (or fictitious particles) $z\in\mathcal{M}$, and the potential V on $\mathbb{T}^D$ is extended to a potential $\tilde{V}$ on $\mathbb{T}^D\times\mathcal{M}$ as

$$\tilde{V}(q,z) = V(x) + \frac{1}{2\sigma^2}\left(\mathrm{dist}_{\mathcal{M}}\left(\xi(q),z\right)\right)^2, \quad \forall (q,z)\in\mathbb{T}^D\times\mathcal{M},$$

for some small parameter $\sigma > 0$, where $\mathrm{dist}_{\mathcal{M}}$ stands for the distance on $\mathcal{M}$. The reaction coordinates on the extended space are then defined by $\tilde{\xi}(q,z) = z$, which means the framework considered in the present paper is general for the EABF algorithm. If (Q,Z) is distributed according to $\mu_{\tilde{V},\beta}$, the law of Z is obtained from the law of $\xi(Q)$ through a Gaussian convolution of variance σ^2/β. There are several practical advantages to EABF:

- In the potential $\tilde{V}$, in the case where $\mathrm{dist}_{\mathcal{M}}$ is an Euclidean distance, the z_i's for $i\in[\![1,d]\!]$ appear in separate terms of a sum, they are not directly coupled. As a consequence, in the EABF case, $\mu_{A,\beta}$ should be, in some sense, closer to the product of its marginal (namely, at equilibrium, the Z_i's should be closer to be independent) than in the non-extended ABF case. In [21], this was a crucial point since the density $\mu_{A,\beta}$ was approximated by a tensor product. But even in our case where the approximation as a sum of tensor product is made at the level of A, we can expect this form of $\tilde{V}$ to improve the approximation.

- After convolution, the so-called mean-force $\nabla_z A$ is smoother than the initial mean force in the non-extended ABF case. Since it varies less, its estimation is expected to be easier.

That being said, the tensorized ABF introduced above can also be straightforwardly extended to a general ABF framework, without extended coordinates.

5.2 Non-periodic reaction coordinates

In general, $\mathcal{M}$ may be different from $\mathbb{T}^d$. If it has boundaries, for instance if $\mathcal{M} = [0,1]^d$, the definition of the algorithm is the same except that the diffusion (5) is reflected at the boundaries of $\mathcal{M}$. The proof of well-posedness and convergence of the tensor algorithm, i.e. Theorem 5, is unchanged. The proof of the long-time convergence of the idealized algorithm, i.e. Theorem 2, is similar up to technical considerations in particular to take into account boundary conditions in Section 3.2.

Moreover, $\mathcal{M}$ may not be compact, for instance $\mathcal{M} = \mathbb{R}^d$, with U satisfying suitable growth conditions at infinity. Since the Lebesgue measure has not a finite mass, a confining biasing potential has to be added to the adaptive biasing potential, see [35, Section 1.2], in which case the law of $\xi(X_t)$ is not supposed to converge to a uniform law (flat histogram) but to a target unimodal law on $\mathbb{R}^d$. Extending our long-time convergence results to this case would lead to additional theoretical difficulties and is out of the scope of the present work.

5.3 Real implementation

The algorithm really implemented for the numerical experiments in Section 6 differs from the theoretical Algorithm 3 in the following points:

1. Time and space are discretized. The SDE (5) is replaced by an Euler-Maruyama scheme with some time step δt and the time integral in (8) is replaced by a discrete sum with a time step Δt (not necessarily small; it can be of the order of the decorrelation length of the process $(Q_t, Z_t)_{t\geqslant 0}$). The one-dimensional functions in the tensor terms are restricted to be continuous piecewise linear, determined by their value on a discrete grid with $q \in \mathbb{N}_*$ points, so that solving (19) amount to solve a $q \times q$ linear system. In particular, the discrete space interpolation plays a role similar to the regularization kernel K which is no more necessary, hence is discarded.

2. In fact, it is not necessary to re-weight the occupation distribution, namely (8) can be replaced by $\nu_t = 1/t \int_0^t \delta_{(Q_s,Z_s)} \mathrm{d}s$. In that case, ν_t is expected to converge to $\mu_{V-\tilde{A}_*,\beta}$ for some $\tilde{A}_*$ instead of $\mu_{V,\beta}$ but the conditional law of Q given $Z = z$ is the same for these two laws. Since the free energy only depends on these conditional laws, A_t is still expected to converge to $\tilde{A}_*$, that should be close to the true free energy in a sense similar to (12). This is clear in the mean-field limit of the algorithm, where no regularization is needed so that $\tilde{A}_* = A$ (see [35]). It is more difficult to establish for the self-interacting ABF process, but in parallel of the present paper it has been done in [8]. We tried numerically both cases, and the results were similar. The results presented in Section 6 are obtained with the full (non reweighted) occupation distribution.

3. Instead of a single particle, in practice, several replicas of the process (5) are simulated in parallel. Denoting N the number of replicas and $(Q_t^i, Z_t^i)_{t\geqslant 0}$ the i^{th} replica, $i \in [\![1, N]\!]$,

the total empirical distribution of the system is

$$\tilde{\nu}_{N,t} = \frac{1}{N\lfloor t/\Delta t\rfloor}\sum_{i=1}^{N}\sum_{k=1}^{\lfloor t/\Delta t\rfloor}\delta_{(Q^i_{k\Delta t},Z^i_{k\Delta t})}\,. \tag{37}$$

The replicas all use the same bias A_t obtained from this empirical distribution by minimizing $\mathcal{J}_{\tilde{\nu}_{N,t}}$ at times $t = t_k = kT_{up}$.

4. It is possible to use the tensor approximation only as a correction of the classical ABF, or more precisely of the Generalized ABF (GABF) algorithm proposed in [51] where the bias is just a sum of one-dimensional functions. Namely, for all time $t > 0$, for $j \in [\![1, d]\!]$, let

$$\begin{aligned}\alpha_{j,t}(z_j) &= \int_{\mathbb{T}^p\times\mathbb{T}^d}\partial_{y_k}V(q,y)K(y_j,z_j)\mathrm{d}\nu_t(q,y)\\ \beta_{j,t}(z_j) &= \int_{\mathbb{T}^p\times\mathbb{T}^d}K(y_j,z_j)\mathrm{d}\nu_t(q,y)\,.\end{aligned}$$

These functions can be recorded on d one-dimensional grids and are easily updated on the fly. Denoting $\gamma_{t,j}(z_j) = \mathbb{1}_{\beta_{j,t}(z_j)>s}\alpha_{j,t}(z_j)/\beta_{j,t}(z_j)$ for some burn-in time $s > 0$, let $A_{t,j}(z_j) = \int_0^{z_j}\gamma_{t,j}(z)\mathrm{d}z$ if the j^{th} reaction coordinate z_j lies in $\mathbb{R}$ and $A_{t,j}(z_j) = \int_0^{z_j}\gamma_{t,j}(u)\mathrm{d}u - z_j\int_0^1\gamma_{t,j}(u)\mathrm{d}u$ if z_j lies in $\mathbb{T}$ (so that, in both cases, $\partial_{z_j}A_{t,j}$ is the Helmoltz projection in $L^2(\mathrm{d}z_j)$ of $\gamma_{t,j}$). Then, at time t, in the dynamics (5), use the bias $\nabla_z A_t$ with $A_t(z) = \sum_{j=1}^d A_{t_k,j}(z_j) + f_m(z)$ where f_m is a tensor approximation obtained through Algorithm 2 of the minimizer of $H \ni f \mapsto \mathcal{J}_{\nu_{t_k}}(f - \sum_{j=1}^d A_{t_k,j})$, where $t_k = \sup\{t_{k'} < t,\ k' \in \mathbb{N}\}$ is the last update time.

5.4 Some limitations and perspectives

A practical limitation observed in the algorithm is the following. Recall that d is too large to keep in memory the empirical measure on a grid by simply recording how many times each d-dimensional cell has been visited by the process, as in the classical ABF algorithm. Instead, the sequence $(Z_{k\Delta t}, \nabla_z V(Q_{k\Delta t}, Z_{k\Delta t}))_{k\in\mathbb{N}}$ is kept in memory for some $\Delta t > 0$, and thus computing an expectation with respect to ν_t has a numerical cost proportional to t. Such integrals are computed when solving the one-dimension equations (19), which have to be solved repeatedly at each addition of a tensor term to the bias. As t grows, the update of the bias gets numerically more expensive. We list here some possible directions to address this question. The analysis of these variations is beyond the reach of the present work.

1. At the beginning of Algorithm 2, a clustering or quantization algorithm (see [47]) can be used to reduce the memory $(Z_{k\Delta t}, \nabla_z V(Q_{k\Delta t}, Z_{k\Delta t}))_{k\in[\![1,t_n/\Delta t]\!]}$ to fewer points.

2. Another way to deal with this problem would be to use a fixed small size for the memory. For instance, at an update time t_k, the empirical measure used to define $\mathcal{J}_{\bar{\nu}_{t_k}}$ could be

$$\bar{\nu}_{t_k} = \left(\int_{t_{k-l}}^{t_k}e^{-\beta A_s(Z_s)}\mathrm{d}s\right)^{-1}\int_{t_{k-l}}^{t_k}\delta_{(Q_s,Z_s)}e^{-\beta A_s(Z_s)}\mathrm{d}s$$

for some small $l \in \mathbb{N}_*$, say $l = 1$. In that case, in order to expect a long-time convergence of the bias, following classical stochastic algorithms, we would define the new bias as

$A_{t_k} = A_{t_{k-1}} + \gamma_k f_k$ where f_k is (a tensor approximation of) a minimizer over H of $H \ni f \mapsto \mathcal{J}_{\nu_{t_k}}(A_{t_k} + f)$ and $(\gamma_k)_{k\in\mathbb{N}}$ is a positive sequence with $\gamma_k \to 0$ and $\sum_{l=1}^{k} \gamma_l \to \infty$ as $k \to \infty$.

3. A third way to deal with the memory management as time increases could be to use a stochastic gradient descent when solving the one-dimensional partial differential equation (19). In other words, when optimizing r_i for some $1 \leq i \leq d$, instead of computing averages over all steps $l \in [\![1, t_k/\Delta t]\!]$, only use an approximation of ν_{t_k} by picking a random (and comparatively small) set of steps among $[\![1, t_k/\Delta t]\!]$. Then only an estimation of the gradient of $H^1(\mathbb{T}) \ni r_i \mapsto \mathcal{J}_{t_k}(f + \bigotimes_{j=1}^{d} r_j)$ is computed, which is exactly the settings of the stochastic gradient descent.

4. Finally, since integrals with respect to the occupation measure are computed at each step of the Alternating Least Square (ALS) method, the cost can be reduced by using more sophisticated tensor approximation algorithms [46, 29, 25].

A second possible limitation is the following. Note that, as the number of reaction coordinates increases, we can expect that, at some point, the biasing scheme becomes unefficient. Indeed, by flattening the energy landscape, we replace the initial sampling problem (that was mainly restricted to low-energy regions, which form a low-dimensional manifold) by the sampling of the uniform measure on some hypercube, which is not so easy. In some sense, following the definitions of [34], at some point, energy barriers are replaced by entropic ones (which means that, in the exploration of the space, what takes time is not crossing high energy areas but visiting all areas in a relatively high dimensional space). Moreover, the variance of the estimator (3) increases due to the exponential weights. As a consequence, as d increases, a partial biasing with $V_{bias,t} = \theta A_t \circ \xi$ for some $\theta \in (0, 1)$ may be more appropriate than the full biasing (i.e. $\theta = 1$). At the biased equilibrium, if $A_t = A$ is the true free energy, the marginal law of the reactions coordinates is thus $\mu_{(1-\theta)A,\beta}$, i.e. the temperature is increased. Then the choice of θ such that this measure satisfies a Poincaré inequality with minimal constant (which means the corresponding overdamped Langevin process mixes the fastest) may not be $\theta = 1$, see e.g. [20].

5.5 Other possible simple variations

1. From the biased trajectory $(Q_t, Z_t)_{t\geqslant 0}$ provided by the TABF algorithm, in order to compute expectations with respect to the target Gibbs measure $\mu = \mu_{V,\beta}$, an alternative to the reweighting step (3) is the following. Remark that only the Z variable is biased, so that for all $z \in \mathbb{T}^d$, the conditional expectations $\int_{\mathbb{T}^p} f(q, z)\mu(q, z)\mathrm{d}q / \int_{\mathbb{T}^p} \mu(q, z)\mathrm{d}q$ can be estimated without re-weighting. On the other hand, the marginal law of Z is estimated by $\exp(-\beta A_{T_{tot}})/\int_{\mathbb{T}^d} \exp(-\beta A_{T_{tot}}(z))\mathrm{d}z$.

2. The bias update period T_{up} and the number m of tensor products added at each update in Algorithm 3, instead of having fixed values, could be adaptively chosen. For instance, the bias could be updated when the histogram of the reaction coordinates have reached some stability, and m could be the lowest integer $n \in \mathbb{N}$ such that $\mathcal{J}_\nu(A_{t_k} + f_{n-d}) - \mathcal{J}_\nu(A_{t_k} + f_n) \leqslant \varepsilon$ for some threshold $\varepsilon > 0$.

3. A time-dependent weight in the definition (8) of ν_t (or, in practice, in (37) for $\tilde{\nu}_{N,t}$) can be added in such a way that old samples have less influence than new ones since they are more biased toward the initial distribution.

4. The regularization kernel K and parameter λ may depend on time. Indeed, as time goes, the size of the sample increases. Since the problem of minimizing $\mathcal{J}_{\nu_t}$ is in practice solved on a finite dimension space, for a time large enough the regularization is actually not necessary anymore and the minimization problem with $K(z,y) = \delta_z(y)$ and $\lambda = 0$ is well-posed.

5. For $k \in \mathbb{N}_+$, denote by $\mathcal{J}_k^\lambda$ the function given by (11) for some $\lambda > 0$ with $\nu = \nu_{t_k}$. Rather than setting A_{t_k} to be the minimizer of $\mathcal{J}_k^\lambda$, we can set it to be $A_{t_{k-1}} + f$ where f is the minimizer of

$$H \ni f \;\; \mapsto \;\; \mathcal{J}_k^0\left(A_{t_{k-1}} + f\right) + \lambda \int_{\mathbb{T}^d} |\nabla_z f(z)|^2 \mathrm{d}z\,, \tag{38}$$

the difference being that $A_{t_{k-1}}$ does not appear in the last regularization term any more. Note that, when $\lambda = 0$, there is no difference. When $\lambda > 0$, the theoretical results of Section 4, i.e. the well-posedness of the tensor approximation, can be straightforwardly adapted. The long-time behaviour study of Section 3 should be similar, although a bit more troublesome since A_{t_k} would not depend only on the empirical distribution η_{t_k} but also on the previous bias $A_{t_{k-1}}$. On the other hand, remark that 0 is a minimizer of (38) if and only if A_{t_k} is a minimizer of $\mathcal{J}_k^0$. As a consequence, the long-time limit of A_t should be the minimizer of $\mathcal{J}_\mu$ with $\lambda = 0$ which, in view of (12), advocates for this alternative form of cost function.

6. Since the bias is stored in memory in a tensor form, it is possible to use at some times a compression algorithm (see [18]) to reduce the number of tensor terms, if needed.

6 Numerical experiments

Let us fix some details and parameters that will hold for the different examples below. In this section, the modifications discussed in Section 5.3 are enforced.

The one-dimensional functions $r_{n,j}$ are stored for all $n \in \mathbb{N}$ and $1 \leq j \leq d$ on a discrete grid with $M = 30$ points, so that the minimization of functions of the form $\bigotimes_{j=1}^d r_j \mapsto \mathcal{J}_{\nu_t}(f + \bigotimes_{j=1}^d r_j)$ is restricted to tensor products of one-dimensional continuous piecewise linear functions on this grid, and the Euler-Lagrange equations (19) are replaced by $M \times M$ linear systems. This discretization replaces the regularization by a kernel K, which is no more necessary. The process (5) is discretized with a time-step $\delta t = 25.10^{-5}$, while the time integral in the empirical measure ν_t defined in (8) is discretized with a time-step $\Delta t = 20\delta t$. In other words, the reaction coordinates and the associated local mean forces are recorded in memory only every 20 steps of the Euler scheme. Moreover, N independent replicas of the processes are run in parallel and the occupation measure used to defined the bias is $\tilde{\nu}_{N,t}$ given by (37). The update times t_k of the bias are fixed at $t_k = kT$, with T a multiple of Δt and the number of tensor terms g_n added at each update time is fixed with value m.

6.1 A low dimensional example

We start to test the method on a toy model, with $N = 30$ replicas, a bias update period of $T = 100\Delta t$, a regularization parameter $\lambda = 10^{-5}$, and $m = 8$ tensor products added at each update. The reaction coordinates are Euclidean coordinates, more precisely $\xi(x) = (x_1, x_2)$, so that we don't introduce any additional extended coordinate.

The dimensions are $D = 3$, $d = 2$, particles start at $(0, 0, 0)$ and

$$\begin{aligned} V(x_1, x_2, x_3) &= -\sin(3x_1)\sin(x_2)\cos(x_3 - 1) + \cos(3x_2 + 2)(0.5 + \cos(x_3 - 2)) \\ &\quad + 2\sin(2x_1 + 0.5)\cos(x_3) - 5\cos(x_1)\cos(x_2)\cos(x_3 + 1)\,. \end{aligned}$$

This potential has the following properties: it is not a tensor product and yields a metastable process but, since $V(x_1, x_2, x_3) = \psi(x_1, x_2)\cos(x_3 + \varphi(x_1, x_2))$ for some functions ψ and φ, there is no metastability in the orthogonal space for fixed x_1, x_2.

The results are given in Figures 1 and 2 (for $\beta = 1$) and 3 and 4 (for $\beta = 5$). In both cases, the theoretical free energy is successfully computed and the histograms of the reaction coordinates is eventually flat. This is a bit slower with the inverse temperature $\beta = 5$, since the initial metastability is very strong. As can be seen in Figure 4, at that temperature and in the same times, a non-biased process is stuck in its initial well.

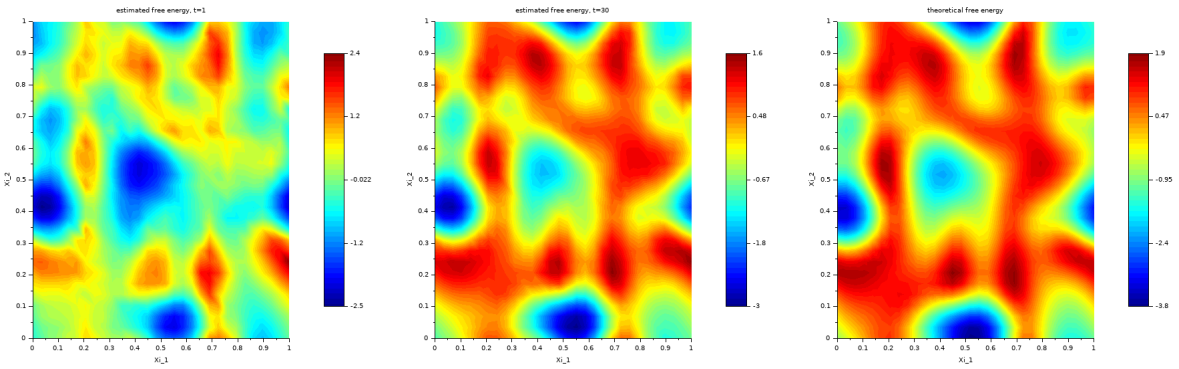

Figure 1: For $\beta = 1$, left and middle: estimated free energy respectively at $t = 1$ and $t = 30$. Right: theoretical free energy.

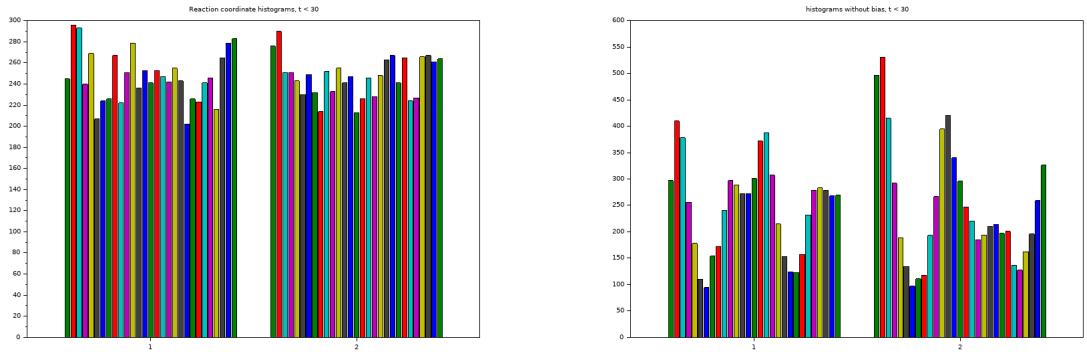

Figure 2: For $\beta = 1$, cumulated histograms of the reaction coordinates at $t = 30$ respectively for the TABF algorithm (left) and a non-biased process (right).

6.2 Polymer ring in solvent

We now consider a system inspired from [1]. The system is constituted of two types of particles, solvent particles, and polymer particles. The polymer particles interact through a potential

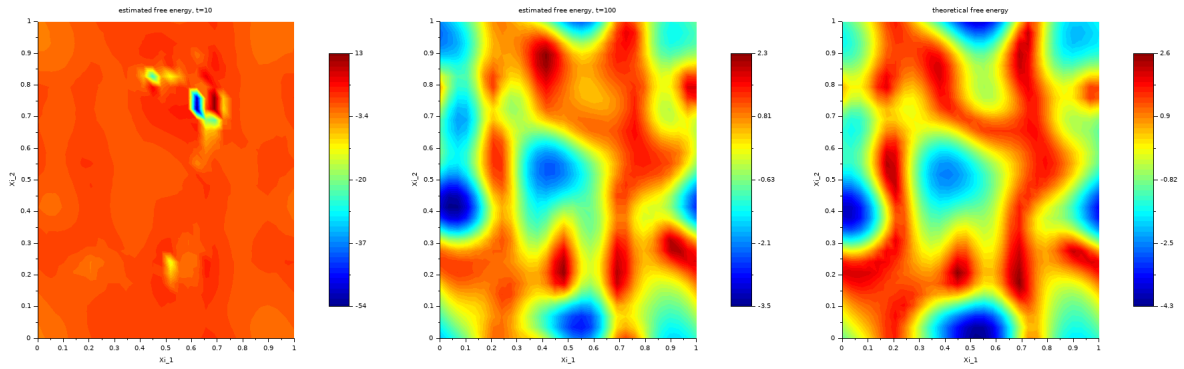

Figure 3: For $\beta = 5$, left and middle: estimated free energy respectively at $t = 10$ and $t = 100$. Right: theoretical free energy.

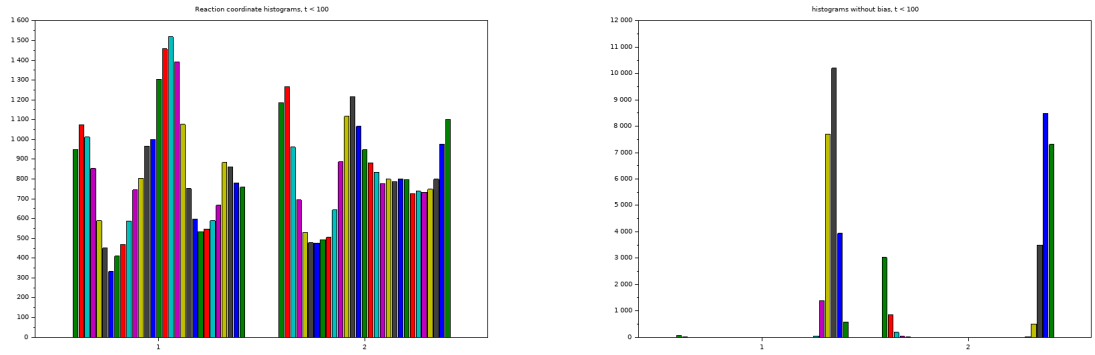

Figure 4: For $\beta = 5$, cumulated histograms of the reaction coordinates at $t = 100$ respectively for the TABF algorithm (left) and a non-biased process (right).

made precise below to form a ring. The reaction coordinates are the bond lengths between consecutive polymer particles. This gives a large dimensional problem, for which the total dimension and the number of reaction coordinates are easily prescribed, and moreover where the reaction coordinates should exhibit some correlations (if it wasn't the case, then the TABF algorithm would not give different results than GABF [51]).

In a two-dimensional periodic box, we consider $D/2 = 100$ particles among which d (labeled from 1 to d) form a polymer and the others are solvent particles. The length of the box is $L = \sqrt{D/2}$, to ensure a concentration independent from D. Each pair of particles that involves at least one solvent particle interacts through the purely repulsive WCA pair potential, which is the Lennard-Jones potential truncated at its minimum, namely

$$V_{WCA}(r) \quad = \quad \varepsilon \mathbb{1}_{r \leqslant r_0} \left(1 + \left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right)$$

where r denotes the distance between the two particles, $\varepsilon = 1$, $\sigma = 0.5$ and $r_0 = 2^{1/6}\sigma$. Each pair of consecutive particles in the polymer ring (where the d^{th} and first particles are considered to be consecutive, closing the loop) interacts through a double well potential

$$V_{DW}(r) \quad = \quad h \left(1 - \frac{(2r - 2r_1 - \omega)^2}{\omega^2} \right)^2 ,$$

where $r_1 = r_0$, $\omega = 1$ and $h = 3$. The minimum of V_{DW} is attained at $r = r_1$ (compact state) and $r = r_1 + w$ (stretched state). Finally, each triplet of consecutive particles in the polymer also interacts through the angle θ they form with the potential

$$V_A(\theta) \;=\; \frac{1}{2}\left(\cos(\theta) - \cos(\theta_d)\right)^2$$

with an equilibrium angle $\theta_d = \pi(1 - 2/d)$ that ensures that the total angular potential is minimized when the polymer particles form a regular polygon.

There are d reaction coordinates, which are the distances between two consecutive polymer particles. Following Section 5.1, the interaction between an extended reaction coordinate z and the corresponding distance r in the system is given via the extended potential

$$V_E(z, r) \;=\; \frac{1}{2\delta}\left(z - \frac{r - r_1}{w}\right)^2$$

for $\delta = 0.01$. The scaling ensures that the minimum of $V_E(z, r) + V_{DW}(r)$ is attained at $z = (r - r_1)/w \in \{0, 1\}$. Moreover, in line with Section 5.2, the extended variable is confined in $[\xi_{min}, \xi_{max}]^d$ by orthogonal reflection at the boundary, with $\xi_{min} = -0.2$ and $\xi_{max} = 1.2$.

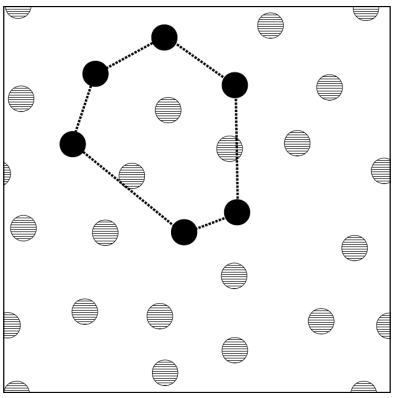

Figure 5: The slow motions of the system are the transitions of each bond between two consecutive particles of the polymer from its compact state to its stretched state.

The total energy of the (extended) system is thus, for $(q, z) \in (L\mathbb{T})^D \times [\xi_{min}, \xi_{max}]^d$,

$$\begin{aligned} V(q, z) \;=\; & \sum_{i=d+1}^{D/2} \sum_{j<i} V_{WCA}\left(|q_i - q_j|\right) + \sum_{i=1}^{d} V_E\left(z_i, |\tilde{q}_{i+1} - q_i|\right) \\ & + \sum_{i=1}^{d} V_{DW}\left(|q_i - \tilde{q}_{i+1}|\right) + \sum_{i=1}^{d-1} V_A\left(\arccos\left(\frac{\tilde{q}_{i+1} - q_i}{|\tilde{q}_{i+1} - q_i|} \cdot \frac{\tilde{q}_{i+2} - \tilde{q}_{i-1}}{|\tilde{q}_{i+2} - \tilde{q}_{i+1}|}\right)\right), \end{aligned}$$

where $\tilde{q}_i = q_i$ for all $i \in [\![1, d]\!]$ and $\tilde{q}_{d+j} = q_j$ for $j = 1, 2$. Initially, the polymer is in a compact state, i.e. the distances between two consecutive of its particles are at distance r_1, the angles are θ_d and all the extended variables $(z_i)_{i \in [\![1,d]\!]}$ are at 0. For this model, we use the variant described in point 4 of Section 5.3 namely, following the GABF algorithm, we keep in memory one dimensional free energies on a grid and we use the tensor approximation as a correction of this initial guess. There are $N = 50$ replicas, the update period, regularization parameter, and inverse temperature are respectively $T = 10^4 \Delta t$, $\lambda = 0.05$ and $\beta = 1$, and at each update, $m = 4d$ tensor products are added.

The free energy is expected to be close to a sum of one-dimensional double well potentials, with minima attained at points close to 0 and 1. Nevertheless the angular force should favor configurations where the consecutive distances in the polymer are close. This fact cannot be grasped by the GABF algorithm alone, for which reaction coordinates are treated independently one from the others.

The results are presented in Figure 6 for $d = 3$ and Figures 7 and 8 for $d = 5$. In Figure 6, we see that indeed the one-dimensional free energies recovered by the GABF algorithm have two wells approximately at 0 and 1, and that the non-independent part of the free energy has the following effect: when $z_3 = 0$, the well $(0, 0)$ is favored, when $z_3 = 1$ the same goes for $(1, 1)$, and when z_3 is intermediate the landscape is flatter and the two wells $\{z_1 = z_2 = x\}$ with $x \in \{0, 1\}$ are favored with respect to the wells $(0, 1)$ and $(1, 0)$. The result is similar in Figure 7, even though the quality of the estimation is lower for $z_3 = z_4 = z_5 = 0.5$, which is to be expected as this lies in a very low probability area (since 0.5 is the saddle point of the two well potential). This shows that the TABF algorithm is able to recover non-trivial dependency structures between reaction coordinates.

For $d = 5$, the free energy is eventually approximated with 140 tensor terms (7 updates, adding 20 terms each), which means $5 \times 140 = 700$ one-dimensional functions have been stored, each represented by $M = 30$ numbers. This is orders of magnitude below the cost $30^5 = 2,43 \times 10^7$ required to store a 5-dimensional function on a grid of the same precision. Moreover, since the total number of time steps of the simulation is of order 10^6, most of the points of the 5-dimensional grid have never been visited during the whole simulation, so we would'nt have any estimation of the free energy with a classical ABF algorithm.

We investigate further the quality of the tensor approximation by computing the evolution of the square error on the free energy as a function of the number of tensor terms. To this aim, first, using the same data as in Figure 7, in Figure 9 we plot $\|A_m - A_{m_f}\|_2^2$ with $m_f = 140$ where A_m is the approximation obtained with the m first tensor terms constructed by the algorithm (the GABF one-dimensional free energies are not taken into account). We see a fast decay at $m \in \{0, 20, 60\}$, which corresponds to different update times (recall 20 terms are added at each update), and thus to different empirical distributions and different GABF estimations. This makes this figure difficult to interpret. For this reason, second, in order to focus only on the tensor approximation step without the adaptive bias process or the GABF one-dimensional free energies, we consider $N = 50$ independent long standard (non-biased) trajectories with 5.10^6 time steps followed by a unique tensor approximation of the free energy with Algorithm 2. Denoting again by A_m the approximation obtained with m tensor terms, we plot $\|A_m - A_{m_f}\|_2^2$ with $m_f = 200$. The result, displayed in Figure 10, shows a fast convergence. More quantitatively, we observe that the relative error $\|A_m - A_{m_f}\|_2^2 / \|A_{m_f}\|_2^2$ is below 20% with $m = 19$, below 10% with $m = 53$ and below 1% with $m = 123$.

Acknowledgements

This work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement number 614492 and under the European Union's Horizon 2020 Research and Innovation Programme, ERC Grant Agreement number 810367, project EMC2. It was also supported by the ANR JCJC project COMODO (ANR-19-CE46-0002) and the ANR Project EFI (ANR-17-CE40-0030) of the French National Research Agency. The authors would like to thank the associate editor and the referees for their work.

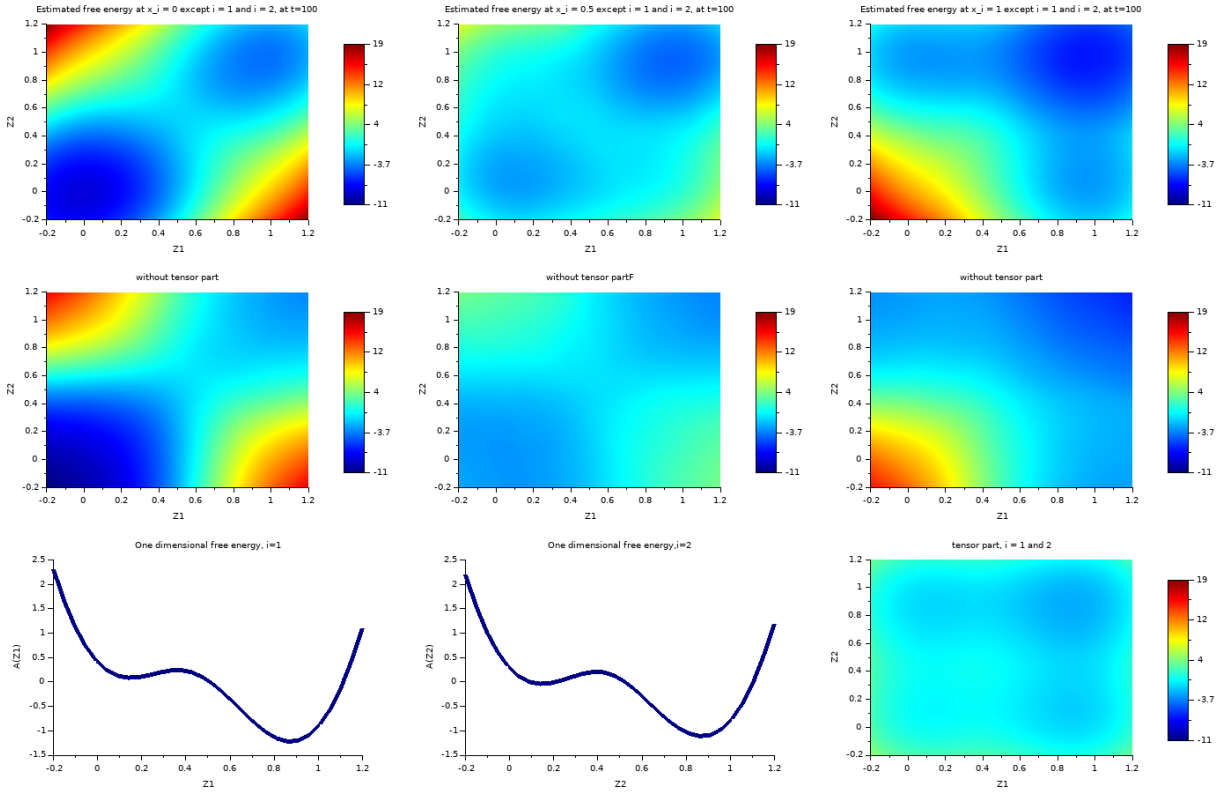

Figure 6: For $d = 3$, up: estimated free energy as a function of (z_1, z_2) when z_3 is, respectively, 0 (left) 0.5 (middle) and 1 (right). Middle: idem but without the independent, one-dimensional parts given by the GABF algorithm. Bottom: one-dimensional potential given by the GABF algorithm for z_1 (left) z_2 (middle) and their sum (right).

References

[1] H. Alrachid and T. Lelièvre. Long-time convergence of an adaptive biasing force method: Variance reduction by Helmholtz projection. *SMAI Journal of Computational Mathematics*, 1:55–82, 2015.

[2] L. Ambrosio, A. Carlotto, and A. Massaccesi. *Lectures on elliptic partial differential equations*, volume 18 of *Appunti. Scuola Normale Superiore di Pisa (Nuova Serie)*. Edizioni della Normale, Pisa, 2018.

[3] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[4] A. Barone, M. G. Carlino, A. Gizzi, S. Perotto, and A. Veneziani. Efficient estimation of cardiac conductivities: A proper generalized decomposition approach. *Journal of Computational Physics*, 423:109810, 2020.

[5] P.A. Bash, Singh U.C., R. Langridgeand, and P.A. Kollman. Free energy calculations by computer simulation. *Science*, 236(4801):564–568, 1987.

[6] 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.

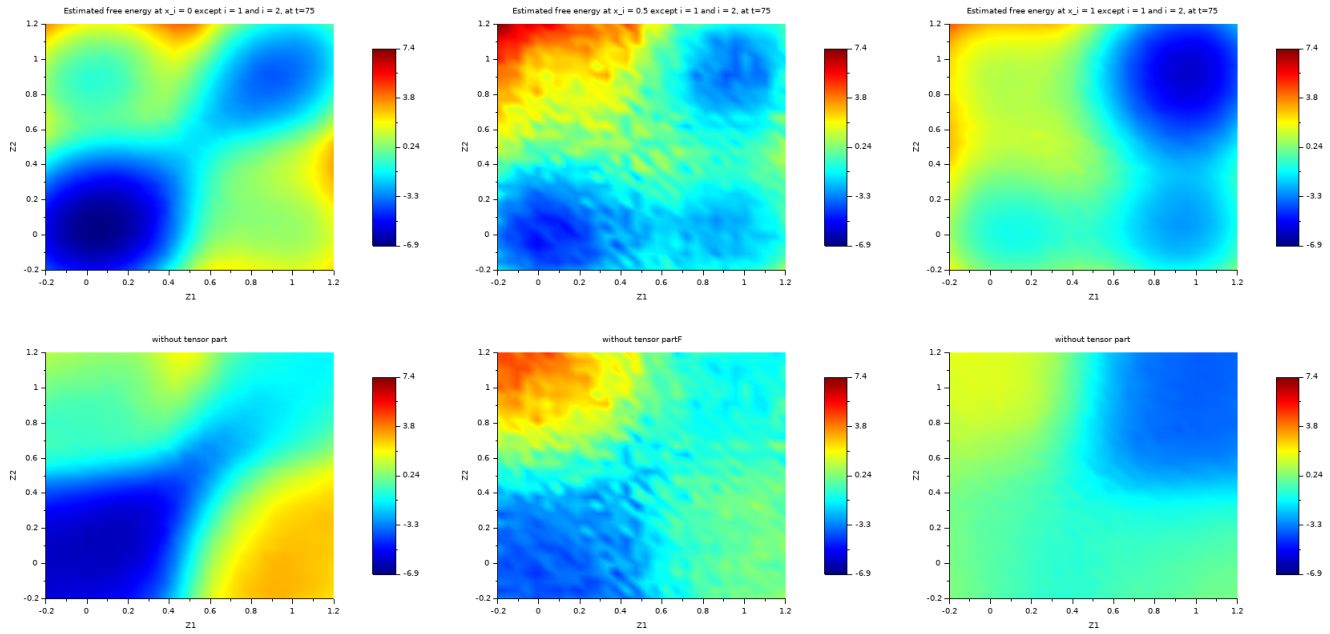

Figure 7: For $d = 5$, up: estimated free energy as a function of (z_1, z_2) when z_3, z_4, z_5 are, respectively, $(0, 0, 0)$ (left) $(0.5, 0.5, 0.5)$ (middle) and $(1, 1, 1)$ (right). Bottom: idem but without the independent, one-dimensional parts given by the GABF algorithm.

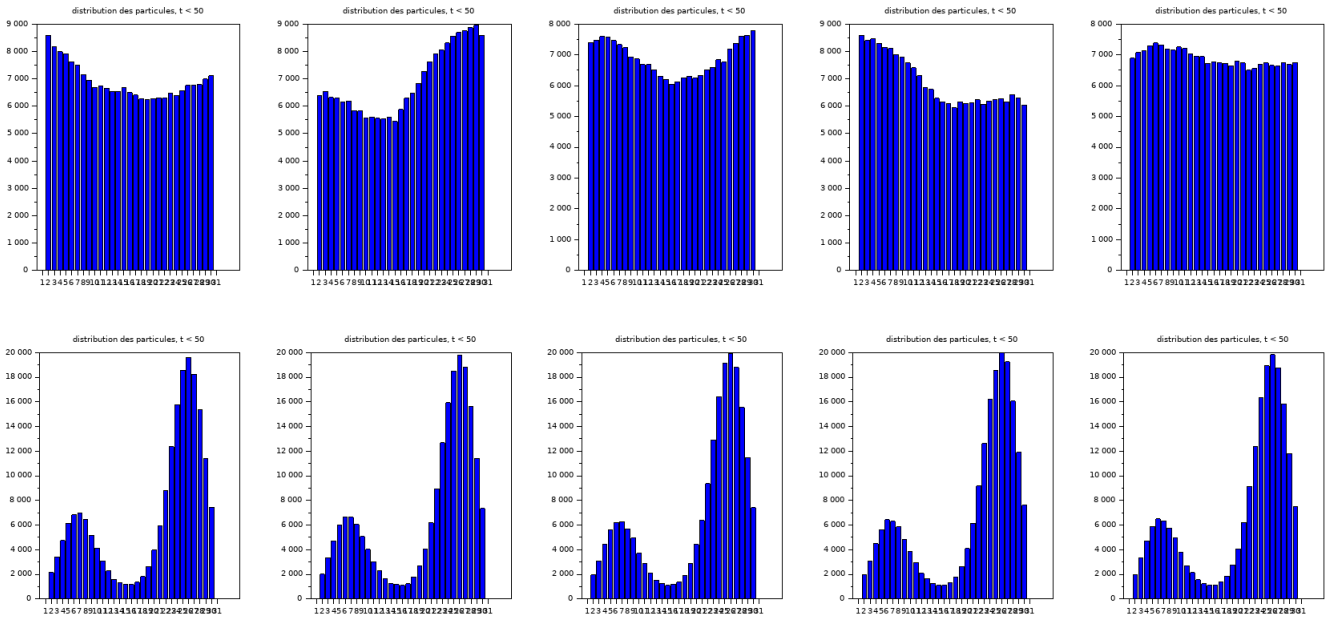

Figure 8: For $d = 5$, cumulative one-dimensional histograms of the five reaction coordinates at $t = 50$ for the TABF algorithm (up) and for a non biased process (bottom).

[7] M. Benaïm and C.-E. Bréhier. Convergence analysis of Adaptive Biasing Potential methods for diffusion processes. *Communications in Mathematical Sciences*, 2019.

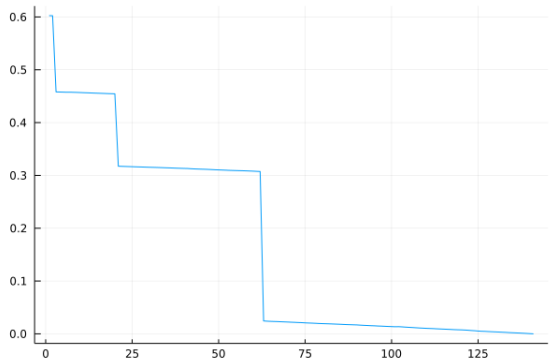

Figure 9: Square error on the free energy as a function of the number of tensor terms (same simulation as Figure 7).

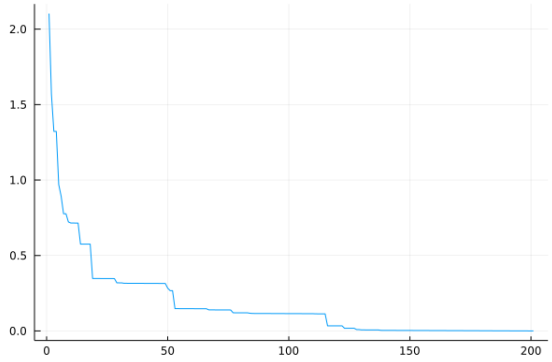

Figure 10: Square error on the free energy as a function of the number of tensor terms (long unbiased trajectories followed by Algorithm 2).

[8] M. Benaïm, C.-E. Bréhier, and P. Monmarché. Analysis of an Adaptive Biasing Force method based on self-interacting dynamics. *Electronic Journal of Probability*, 25(none):1 – 28, 2020.

[9] M. Benaïm, M. Ledoux, and O. Raimond. Self-interacting diffusions. *Probab. Theory Related Fields*, 122(1):1–41, 2002.

[10] A. Bittracher, R. Banisch, and C. Schütte. Data-driven computation of molecular reaction coordinates. *The Journal of Chemical Physics*, 149(15):154103, 2018.

[11] S. Brandt, F. Sittel, M. Ernst, and G. Stock. Machine learning of biomolecular reaction coordinates. *The Journal of Physical Chemistry Letters*, 9(9):2144–2150, 2018.

[12] E. Cancès, V. Ehrlacher, and T. Lelièvre. Convergence of a greedy algorithm for high-dimensional convex nonlinear problems. *Math. Models Methods Appl. Sci.*, 21(12):2433–2467, 2011.

[13] F. Chinesta, A. Ammar, A. Leygue, and R. Keunings. An overview of the proper generalized decomposition with applications in computational rheology. *Journal of Non-Newtonian Fluid Mechanics*, 166(11):578–592, 2011.

[14] F. Chinesta, R. Keunings, and A. Leygue. *The proper generalized decomposition for advanced numerical simulations: a primer*. Springer Science & Business Media, 2013.

[15] F. Chinesta, P. Ladeveze, and E. Cueto. A short review on model order reduction based on proper generalized decomposition. *Archives of Computational Methods in Engineering*, 18(4):395, 2011.

[16] 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. *The Journal of Physical Chemistry B*, 119(3):1129–1151, 2015. PMID: 25247823.

[17] E. Darve and A. Pohorille. Calculating free energies using average force. *The Journal of Chemical Physics*, 2001.

[18] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4):1253–1278, 2000.

[19] M. Espig, W. Hackbusch, and A. Khachatryan. On the Convergence of Alternating Least Squares Optimisation in Tensor Format Representations. *arXiv e-prints*, page arXiv:1506.00062, May 2015.

[20] G. Fort, B. Jourdain, T. Lelièvre, and G. Stoltz. Convergence and efficiency of adaptive importance sampling techniques with partial biasing. *J Stat Phys*, 171:220–268, 2018.

[21] H. Fu, X. Shao, C. Chipot, and W. Cai. Extended adaptive biasing force algorithm. an on-the-fly implementation for accurate free-energy calculations. *J. Chem. Theory Comput.*, 12(8):3506–3513, 2016.

[22] P. Gkeka, G. Stoltz, A. B. Farimani, Z. Belkacemi, M. Ceriotti, J. Chodera, A. R. Dinner, A. Ferguson, J.-B. Maillet, H. Minoux, C. Peter, F. Pietrucci, A. Silveira, A. Tkatchenko, Z. Trstanova, R. Wiewiora, and T. Lelièvre. Machine learning force fields and coarse-grained variables in molecular dynamics: application to materials and biological systems. *To appear in Journal of Chemical Theory and Computation*, 2020.

[23] L. Grasedyck, D. Kressner, and C. Tobler. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36(1):53–78, 2013.

[24] Z. Guo, C. L. Brooks, and X. Kong. Efficient and flexible algorithm for free energy calculations using the λ-dynamics approach. *The Journal of Physical Chemistry B*, 102(11):2032–2036, 1998.

[25] W. Hackbusch. *Tensor spaces and numerical tensor calculus*, volume 42. Springer, 2012.

[26] J. Hénin and C. Chipot. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *The Journal of Chemical Physics*, 121:2904–2914, 2004.

[27] R. A. Holley, S. Kusuoka, and D. W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2):333–347, 1989.

[28] W. L. Jorgensen and C. Ravimohan. Monte carlo simulation of differences in free energies of hydration. *The Journal of Chemical Physics*, 83(6):3050–3054, 1985.

[29] B.N. Khoromskij. *Tensor Numerical Methods in Scientific Computing*. De Gruyter, 2018.

[30] S. Klus and C. Schütte. Towards tensor-based methods for the numerical approximation of the Perron–Frobenius and Koopman operator. *Journal of Computational Dynamics*, 3:139, 2016.

[31] K. Konakli and B. Sudret. Global sensitivity analysis using low-rank tensor approximations. *Reliability Engineering & System Safety*, 156:64–83, 2016.

[32] C. Le Bris, T. Lelièvre, and Y. Maday. Results and questions on a nonlinear approximation approach for solving high-dimensional partial differential equations. *Constr. Approx.*, 30(3):621–651, 2009.

[33] T. Lelièvre. A general two-scale criteria for logarithmic sobolev inequalities. *Journal of Functional Analysis*, 256(7):2211 – 2221, 2009.

[34] T. Lelièvre. Two mathematical tools to analyze metastable stochastic processes. In *Numerical mathematics and advanced applications 2011*, pages 791–810. Springer, Heidelberg, 2013.

[35] T. Lelièvre, M. Rousset, and G. Stoltz. Long-time convergence of an adaptive biasing force method. *Nonlinearity*, 21(6):1155–1181, 2008.

[36] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations: A mathematical perspective*. Imperial College Press, 2010.

[37] T. Lelièvre and G. Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, May 2016.

[38] T. Lelièvre, M. Rousset, and G. Stoltz. Computation of free energy profiles with parallel adaptive dynamics. *The Journal of chemical physics*, 126(13):134111, April 2007.

[39] A. Lesage, T. Lelièvre, G. Stoltz, and J. Hénin. Smoothed biasing forces yield unbiased free energies with the extended-system adaptive biasing force method. *The Journal of Physical Chemistry B*, 121(15):3676–3685, 2017. PMID: 27959559.

[40] C. Lu, X. Li, D. Wu, L. Zheng, and W. Yang. Predictive sampling of rare conformational events in aqueous solution: Designing a generalized orthogonal space tempering method. *Journal of Chemical Theory and Computation*, 12(1):41–52, 2016.

[41] L. Lu and G. A. Voth. The multiscale coarse-graining method. vii. free energy decomposition of coarse-grained effective potentials. *The Journal of Chemical Physics*, 134(22):224107, 2011.

[42] P. Monmarché. Generalized Γ calculus and application to interacting particles on a graph. *Potential Anal.*, 50(3):439–466, 2019.

[43] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink. The martini coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008. PMID: 26621095.

[44] S. Niroomandi, I. Alfaro, D. González, E. Cueto, and F. Chinesta. Model order reduction in hyperelasticity: a proper generalized decomposition approach. *International Journal for Numerical Methods in Engineering*, 96(3):129–149, 2013.

[45] F. Nüske, R. Schneider, F. Vitalini, and F. Noé. Variational tensor approach for approximating the rare-event kinetics of macromolecular systems. *The Journal of Chemical Physics*, 144(5):054105, 2016.

[46] I. Oseledets. *SIAM J. Sci. Comput.*, 33:2295–2317, 2011.

[47] G. Pagès. Introduction to vector quantization and its applications for numerics. In *CEMRACS 2013—modelling and simulation of complex systems: stochastic and deterministic approaches*, volume 48 of *ESAIM Proc. Surveys*, pages 29–79. EDP Sci., Les Ulis, 2015.

[48] S. Piana and A. Laio. A bias-exchange approach to protein folding. *The Journal of Physical Chemistry B*, 111(17):4553–4559, 2007. PMID: 17419610.

[49] E. Pruliere, F. Chinesta, and A. Ammar. On the deterministic solution of multidimensional parametric models using the proper generalized decomposition. *Mathematics and Computers in Simulation*, 81(4):791–810, 2010.

[50] A. Uschmajew. Local convergence of the alternating least squares algorithm for canonical tensor approximation. *SIAM Journal on Matrix Analysis and Applications*, 33(2):639–652, 2012.

[51] T. Zhao, H. Fu, T. Lelièvre, X. Shao, C. Chipot, and W. Cai. The extended generalized adaptive biasing force algorithm for multidimensional free-energy calculations. *Journal of Chemical Theory and Computation*, 13(4):1566–1576, 2017.