ESAIM: M2AN 56 (2022) 529–564
https://doi.org/10.1051/m2an/2022010

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

THE ADAPTIVE BIASING FORCE ALGORITHM WITH NON-CONSERVATIVE FORCES AND RELATED TOPICS

Tony Lelièvre[1], Lise Maurin[2,3,*] and Pierre Monmarché[2,3]

Abstract. We propose a study of the Adaptive Biasing Force method's robustness under generic (possibly non-conservative) forces. We first ensure the flat histogram property is satisfied in all cases. We then introduce a fixed point problem yielding the existence of a stationary state for both the Adaptive Biasing Force and Projected Adapted Biasing Force algorithms, relying on generic bounds on the invariant probability measures of homogeneous diffusions. Using classical entropy techniques, we prove the exponential convergence of both biasing force and law as time goes to infinity, for both the Adaptive Biasing Force and the Projected Adaptive Biasing Force methods.

Mathematics Subject Classification. 35B40, 60J60.

Received March 4, 2021. Accepted January 21, 2022.

1. Introduction

After presenting in Sections 1.1, 1.2 and 1.3 the motivation and well-known results on the Adaptive Biasing Force (ABF) method applied to the overdamped Langevin dynamics with conservative forces, we present in Section 1.4 the dynamics we are interested in, namely the ABF method applied to the overdamped Langevin dynamics with non-conservative forces.

1.1. Setting

Let us work within the so-called *canonical ensemble* (or *NVT ensemble*), where a system of N particles is contained in a fixed volume $\mathcal{V}$, and is in contact with a thermostat of constant temperature T. Denote by $q=(q_1,\dots,q_N)\in\mathcal{D}$ the positions, $p=(p_1,\dots,p_N)\in\mathbb{R}^{dN}$ the momenta, and $(m_1,\dots,m_N)\in\mathbb{R}^N$ the masses of the particles, where $\mathcal{D}$ is the configuration space and $d\in\{1,2,3\}$ is the space dimension. Usually, $\mathcal{D}$ is an open subset of $\mathbb{R}^{dN}$ (or $\mathbb{T}^{dN}$, where the dN-dimensional torus is viewed as the cube $[0,1]^{dN}$ with opposite sides identified, in other words, $\mathbb{T}^{dN}=\mathbb{R}^{dN}/\mathbb{Z}^{dN}$). Interactions between particles are taken into account *via* a

Keywords and phrases. Adaptive Bisaing Force, nonlinear Fokker-Planck equation, long-time behaviour, free energy, entropy techniques.

[1] Université Paris-Est, CERMICS (ENPC), Inria, 77455 Marne-la-Vallée, France.

[2] Sorbonne Université, LJLL, 4 place Jussieu, 75005 Paris, France.

[3] Sorbonne Université, LCT, 4 place Jussieu, 75005 Paris, France.

*Corresponding author: lise.maurin@sorbonne-universite.fr

potential function $V : \mathcal{D} \to \mathbb{R}$, so that the system's total energy is given by the following Hamiltonian:

$$H(q,p) = V(q) + \frac{1}{2} p^\top M^{-1} p,$$

with $M = \mathrm{diag}(m_1 I_d, \dots, m_N I_d)$ being the mass matrix. Since this Hamiltonian is *separable*, the positions and the momenta are independent variables in the canonical ensemble, namely under the probability distribution $Z^{-1} e^{-\beta H(q,p)} \, \mathrm{d}q\mathrm{d}p$ where $\beta = 1/(k_B T)$, with k_B being the Boltzmann constant, and $Z = \int_{\mathcal{D}\times\mathbb{R}^{dN}} e^{-\beta H(q,p)} \, \mathrm{d}q\mathrm{d}p$ is the normalization constant, or *partition function*. The momenta p being distributed according to a Gaussian measure, the main issue resides in sampling the positions q, which are distributed according to the *Boltzmann-Gibbs measure*:

$$\mu(\mathrm{d}q) = Z_\mu^{-1} e^{-\beta V(q)} \mathrm{d}q, \qquad Z_\mu = \int_{\mathcal{D}} e^{-\beta V(q)} \, \mathrm{d}q.$$

Thermodynamic properties are obtained by averaging functions of the microstate q which are called *observables*. Given an observable ψ, one would like to compute the following thermodynamic quantity:

$$\mathbb{E}_\mu[\psi] = \int_{\mathcal{D}} \psi \, \mathrm{d}\mu.$$

One of the simplest dynamics to sample the Boltzmann-Gibbs measure is the *overdamped Langevin dynamics*:

$$\mathrm{d}Q_t = -\nabla V(Q_t) \, \mathrm{d}t + \sqrt{2\beta^{-1}} \mathrm{d}W_t, \tag{1.1}$$

where $(W_t)_{t\geq 0}$ is a dN-dimensional standard Brownian motion, and $-\nabla V : \mathcal{D} \to \mathbb{R}^{dN}$ is the *interaction force*. Notice that here, the interaction force is *conservative*, namely it is the gradient of a function (here, minus the gradient of the potential energy V). Under reasonable assumptions on the potential V (see [19] for more details), the process $(Q_t)_{t\geq 0}$ is ergodic with respect to μ. In other words, for any observable $\psi \in \mathcal{C}_0^\infty(\mathcal{D})$, the average over a trajectory of the process converges to the canonical average:

$$\lim_{\tau\to+\infty} \frac{1}{\tau} \int_0^\tau \psi(Q_t) \, \mathrm{d}t = \mathbb{E}_\mu[\psi]. \tag{1.2}$$

1.2. Metastability, reaction coordinate and free-energy profiles

Computing thermodynamic averages can be troublesome, as microscopic and macroscopic timescales can violently differ. Typical microscopic phenomena occur on timescales of the order of 10^{-15}s, while macroscopic ones can take up to 1 h [18]. Furthermore, N needs to be sufficiently large so that the targeted macroscopic phenomena can emerge from the collective, microscopic behaviour of the system.

Such timescales differences are linked to the system's *metastability*: low-energy regions of the configuration space are separated by either high-energy or high-entropy barriers. These regions are called metastable: the process (1.1) remains trapped in a metastable region and occasionally jumps to another one after a long period of time. From a probabilistic point of view, metastability is linked to the *multimodality* of the measure μ: likely regions are separated by low probability regions. The exploration of the state space by the process and the convergence of the trajectorial averages (1.2) can thus take a considerably long time.

One way of avoiding metastability is to capture some slow components of the dynamics $(Q_t)_{t\geq 0}$. To do so, we consider *transition coordinates* (also called reaction coordinates or collective variables), namely mappings $\xi : \mathcal{D} \to \mathcal{M}$, where $\mathcal{M}$ is a manifold of dimension $m \ll dN$. Transition coordinates are designed to provide a *coarse-grained information* on the system's state (for example, the dihedral angle of a molecule, in which case $\mathcal{M} = \mathbb{T}$, or the signed distance to a hypersurface of $\mathcal{D}$, in which case $\mathcal{M} = \mathbb{R}$). In other words, $\xi(q) \in \mathcal{M}$ is the macroscopic state of a microscopic state $q \in \mathcal{D}$. Designing a good reaction coordinate is a difficult problem, that will not be discussed further in the present work (see [12] for a recent review on the question of automatic learning of transition coordinates).

Decomposing

$$\mathcal{D} = \bigsqcup_{z\in\mathcal{M}} \Sigma_z = \bigsqcup_{z\in\mathcal{M}} \{q \in \mathcal{D} | \xi(q) = z\},$$

and denoting by σ_{Σ_z} the measure on Σ_z induced by the Lebesgue measure on $\mathcal{D}$, one can define the measure $\delta_{\xi(q)-z}(\mathrm{d}q)$ by

$$\delta_{\xi(q)-z}(\mathrm{d}q) = \frac{1}{\sqrt{\det G(q)}} \sigma_{\Sigma_z}(\mathrm{d}q),$$

where $G = (\nabla\xi)^\top \nabla\xi$, in other words,

$$G_{i,j} = \nabla\xi_i \cdot \nabla\xi_j, \quad \text{for all } (i,j) \in [\![1,m]\!]^2.$$

The free energy associated to ξ is then expressed as follows: for every $z \in \mathcal{M}$,

$$A(z) = -\frac{1}{\beta}\ln(Z_{\Sigma_z}), \quad Z_{\Sigma_z} = \int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q)-z}(\mathrm{d}q), \tag{1.3}$$

assuming V and ξ are such that $Z_{\Sigma_z} < +\infty$. As can be seen using the co-area formula [19], this definition ensures that the image of μ by ξ is given by

$$\xi * \mu\,(\mathrm{d}z) = \frac{e^{-\beta A(z)}\mathrm{d}z}{\displaystyle\int_{\mathcal{M}} e^{-\beta A(u)}\mathrm{d}u}. \tag{1.4}$$

1.3. The Adaptive Biasing Force method

Introducing a reaction coordinate allows us to construct a less metastable dynamics, the idea being to substitute the potential V in (1.1) for a *biased potential* $V - A\circ\xi$. The new equilibrium measure is then

$$\mu_A(\mathrm{d}q) = Z_{\mu_A}^{-1} e^{-\beta(V-A\circ\xi)(q)}\,\mathrm{d}q, \tag{1.5}$$

where $Z_{\mu_A} = \int_{\mathcal{D}} e^{-\beta(V-A\circ\xi)(q)}\mathrm{d}q$. Given the expression (1.4), the image of μ_A by ξ is the uniform measure: $\xi * \mu_A = \lambda(\mathcal{M})^{-1}\mathbb{1}_{\mathcal{M}}$, with $\lambda(\mathcal{M})$ being the Lebesgue measure of $\mathcal{M}$ (which is here assumed to be compact). Since, contrary to the initial probability measure $\xi * \mu$, the uniform measure is no longer multimodal, we expect a faster sampling of the phase space, provided ξ is well chosen so that μ_A is less multimodal than μ.

Although this change of potential can accelerate the phase space sampling, the free-energy A is *a priori* unknown. The main idea to get round this issue will be to approximate on the fly A, or ∇A, its derivative with respect to the reaction coordinate. To do so, we will consider the *Adaptive Biasing Force (ABF) algorithm* [7,13]:

$$\begin{cases} \mathrm{d}Q_t = \left(-\nabla V(Q_t) + B_t\left(\xi(Q_t)\right)\nabla\xi(Q_t)\right)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}W_t \\ B_t(z) = \mathbb{E}[F(Q_t)\,|\,\xi(Q_t) = z] \qquad \forall z \in \mathcal{M}, \end{cases} \tag{1.6}$$

where $-\nabla V$ is the conservative interaction force, and F is the so-called *local mean force*, which is the vector with components $(F_i)_{i\in[\![1,m]\!]}$ given by:

$$F_i = \sum_{j=1}^m G_{i,j}^{-1}\nabla\xi_j\cdot\nabla V - \beta^{-1}\mathrm{div}\left(\sum_{j=1}^m G_{i,j}^{-1}\nabla\xi_j\right),$$

where $G_{i,j}^{-1}$ denotes the (i,j)-component of the inverse of the matrix G defined above.

This process is motivated by the fact that the aforementioned free energy satisfies:

$$\nabla A(z) = \mathbb{E}[F(Q)|\xi(Q) = z], \quad \forall z \in \mathcal{M} \qquad \text{if } X \sim \mu_A\,,$$

so that μ_A is a fixed point of the Fokker-Planck equation associated to the process. In other words, if $Q_0 \sim \mu_A$, then $Q_t \sim \mu_A$ for all $t \geqslant 0$ and $(Q_t)_{t\geqslant 0}$ is exactly the diffusion (1.1) with the biased potential $V - A \circ \xi$.

Starting from another initial distribution, using entropy estimates and functionnal inequalities, it has been proven in [19], under mild assumptions, that this fixed point is in fact an attractor of the dynamics, in the sense that B_t converges to ∇A in the long-time limit, and the law of Q_t converges to μ_A.

Remark 1.1.

- ▷ In some cases $\mathcal{M}$ is not bounded, for example when ξ is a distance. If so, an additional confining potential $W \circ \xi$ is needed in the drift [19].
- ▷ As discussed in [19], the algorithm (1.6) can be modified in order to obtain a diffusive behaviour for the law of $\xi(Q_t)$. Additional terms depending on ξ are added to obtain the following variant:

$$\begin{cases} \mathrm{d}Q_t = \left(-\nabla V + B_t \circ \xi - \nabla W \circ \xi + \beta^{-1} \nabla \ln(|\nabla \xi|^{-2})\right) |\nabla \xi|^{-2}(Q_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}|\nabla\xi|^{-1}(Q_t)\mathrm{d}W_t \\ B_t(z) = \mathbb{E}[F(Q_t)\,|\,\xi(Q_t) = z], \quad \forall z \in \mathcal{M}. \end{cases}$$

In this case the longtime convergence of B_t towards ∇A is stronger than in the case of (1.6), in that it requires less hypothesis.

We might also consider a variant of the ABF method, namely the *Projected Adaptive Biasing Force* (PABF) algorithm, introduced in [1]:

$$\begin{cases} \mathrm{d}Q_t = \left(-\nabla V(Q_t) + B_t\left(\xi(Q_t)\right)\nabla\xi(Q_t)\right)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}W_t \\ \quad B_t = \mathsf{P}_{L^2(\lambda)}\left(G_t\right) \\ G_t(z) = \mathbb{E}[F(Q_t)\,|\,\xi(Q_t) = z] \qquad \forall z \in \mathcal{M}, \end{cases}$$

where $\mathsf{P}_{L^2(\lambda)}(f)$ stands for the Helmholtz projection with respect to the Lebesgue measure λ of a vector field f on an open bounded set $\mathcal{M} \subset \mathbb{R}^{dN}$ with Lipschitz boundary $\partial\mathcal{M}$ [2]. In other words, it is the gradient of the minimizer on $\{g \in H^1(\mathcal{M}),\ \int_{\mathcal{M}} g\mathrm{d}x = 0\}$ of

$$g \mapsto \int_{\mathcal{M}} |f(x) - \nabla g(x)|^2 \mathrm{d}x\,.$$

More generally, if ν is a continuous positive measure on $\mathcal{M}$, the Helmholtz projection with respect to ν is the minimizer on $\{g \in H^1(\mathcal{M}),\ \int_{\mathcal{M}} g\mathrm{d}x = 0\}$ of $g \mapsto \int_{\mathcal{M}} |f(x) - \nabla g(x)|^2 \nu(\mathrm{d}x)$.

1.4. The non-conservative case

From now on, we only consider periodic boundary conditions and reaction coordinates that are Euclidean coordinates of the system, namely $\mathcal{D} = \mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ for some $n \in \mathbb{N}^*$, $\mathcal{M} = \mathbb{T}^m$ for $m \in \mathbb{N}^*$ such that $m \leqslant n$ and $\xi(x,y) = x$, where we decomposed the position as $q = (x,y) \in \mathcal{D}$, with $x \in \mathbb{T}^m$ and $y \in \mathbb{T}^{n-m}$. Note that the considered process $(Q_t)_{t\geqslant 0}$ can now be read as $(X_t, Y_t)_{t\geqslant 0}$. This latter restriction may seem quite narrow: nevertheless, it is the generic case used for alchemical reactions [16]. Besides, more general reaction coordinates can be reduced to this setting by adding extended variables [11]. Here, such restriction is made only for the sake of clarity: most arguments could be extended (at the price of heavier computations) to the general case $\xi(q) \in \mathcal{M}$.

We are interested in the case where the force in (1.1) is not necessarily conservative, namely is not the gradient of some potential energy V. There are several motivations for this approach, one of them being that the numerical computation of conservative forces $-\nabla V$ sometimes relies on approximations which make the force *a priori* not conservative, in particular in the context of *ab initio* molecular dynamics, see *e.g.*, [6, 24, 26]. In this case, one is interested in knowing if, by controlling the error made on the force $-\nabla V$, one can deduce an estimation of the error made on the system's free energy. The robustness of a diffusion's invariant measure with respect to the perturbation of its drift is a classical problem (see *e.g.*, Sect. 4.3), but note that in the ABF case, the adaptive procedure makes the question more subtle. Moreover, the convergence of the ABF method in such a context cannot be deduced from the aforementionned convergence analysis. We consequently consider the ABF algorithm in the case where $-\nabla V$ is replaced by a general force field $\mathcal{F} \in \mathcal{C}^1(\mathcal{D}, \mathbb{R}^n)$ that we rewrite as $\mathcal{F}(x,y) = (\mathcal{F}_1(x,y), \mathcal{F}_2(x,y)) \in \mathbb{R}^m \times \mathbb{R}^{n-m}$. The local mean force is simply $F = -\mathcal{F}_1$, and the corresponding process is thus, for all $t \geq 0$:

$$\begin{cases} \mathrm{d}X_t = \mathcal{F}_1(X_t, Y_t)\mathrm{d}t + B_t(X_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}W_t^1 \\ \mathrm{d}Y_t \;= \mathcal{F}_2(X_t, Y_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}W_t^2 \end{cases} \tag{1.7}$$

where $W = (W^1, W^2)$ is a standard Brownian motion on $\mathbb{T}^m \times \mathbb{T}^{n-m}$, and, given the average mean force

$$G_t(x) = -\mathbb{E}[\,\mathcal{F}_1(X_t, Y_t) \,|\, X_t = x\,], \quad \forall t \geq 0, \forall x \in \mathbb{T}^m,$$

one has for all $t \geq 0$ and $x \in \mathbb{T}^m$, in the case of the ABF method,

$$B_t(x) = G_t(x),$$

or, in the case of the PABF method,

$$B_t(x) = \mathsf{P}_{L^2(\lambda)}(G_t)(x) := \nabla H_t(x).$$

In either case, denoting by π_t the law of $Z_t = (X_t, Y_t)$ and $\pi_t^\xi(x) = \int_{\mathbb{T}^{n-m}} \pi_t(x,y)\mathrm{d}y$ the density of $X_t = \xi(Z_t)$, then

$$G_t(x) = \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x,y)\frac{\pi_t(x,y)}{\pi_t^\xi(x)}\mathrm{d}y,$$

so that π_t is a weak solution of the Fokker-Planck equation associated to (1.7), that is

$$\begin{cases} \partial_t \pi_t = \beta^{-1}\Delta\pi_t - \nabla\cdot(\mathcal{F}\,\pi_t) - \nabla_x\cdot(B_t\,\pi_t) \\ B_t = \begin{cases} G_t & \text{in the ABF case} \\ \nabla H_t \;=\; \mathsf{P}_{L^2(\lambda)}(G_t) & \text{in the PABF case} \end{cases} \\ G_t(x) = \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x,y)\frac{\pi_t(x,y)}{\pi_t^\xi(x)}\mathrm{d}y \qquad \forall x \in \mathbb{T}^m. \end{cases} \tag{1.8}$$

For a given initial condition π_0, the existence of the process and the proof that it admits a density with respect to the Lebesgue measure, being a strong solution of (1.8), can be established by fixed point arguments or by the convergence of an interacting particles system [15]. We will not address this question here. As a consequence, we would like to emphasize that our arguments will be partially formal, in the sense that we work under the assumption that a density π_t that solves (1.8) exists and is sufficiently regular so that the algebraic computations in the proofs are valid.

Let us emphasize that the bias B_t in (1.8) (*i.e.*, either G_t or $\nabla H_t = \mathsf{P}_{L^2(\lambda)}(G_t)$) depends on π_t, which makes (1.8) a non-linear PDE.

Remark 1.2. In the conservative case, where $\mathcal{F} = -\nabla V$, and $\mu \propto e^{-\beta V}$, up to an additive constant, the free energy A is characterized by either one of these properties:

1. $\xi * \mu \propto e^{-\beta A}$ (distribution of the reaction coordinate at equilibrium).
2. $\nabla A = \mathbb{E}[\nabla_1 V(Z)|\xi(Z) = \cdot\,]$ with $Z \sim \mu$ (average local mean force at equilibrium).
3. $\nabla A = \mathbb{E}[\nabla_1 V(Z)|\xi(Z) = \cdot\,]$ with $Z \sim \mu_A$ (fixed point of the ABF algorithm).

In the non-conservative case, there is no reason for these various definitions to coincide. Besides, $x \mapsto \mathbb{E}[-\mathcal{F}_1(Z)|\xi(Z) = x\,]$ is *a priori* not a gradient. Denoting by $\mu_{\mathcal{F}}$ the invariant measure of the non-biased, out-of-equilibrium dynamics $\partial_t\pi_t = \beta^{-1}\Delta\pi_t - \nabla\cdot(\mathcal{F}\pi_t)$, we are then led to consider the (in general different) functions H_1, H_2 and H_3 given, up to an additive constant, by

1. $\xi * \mu_{\mathcal{F}} \propto e^{-\beta H_1}$.
2. $\nabla H_2 = \mathsf{P}_{L^2(\lambda)}\left(\mathbb{E}[-\mathcal{F}_1(Z)|\xi(Z) = \cdot\,]\right)$ with $Z \sim \mu_{\mathcal{F}}$.
3. $\nabla H_3 = \mathsf{P}_{L^2(\lambda)}\left(\mathbb{E}[-\mathcal{F}_1(Z)|\xi(Z) = \cdot\,]\right)$ with $Z \sim \pi_\infty^{\mathcal{F}}$ an equilibrium of the (P)ABF algorithm.

In other words, in the non-conservative case, an equilibrium of an adaptive algorithm yields an alternative generalization of the notion of free energy that does not coincide in general with the log-density of the law of the reaction coordinates at (unbiased) equilibrium, and whose gradient is not in general the average local mean force at (unbiased) equilibrium.

Outline of this paper. Section 2 introduces several preliminary notions, before stating the main results. Section 3 focuses on the law π_t^ξ of the process $(X_t)_{t\geq 0} = (\xi(Z_t))_{t\geq 0}$. More precisely, we show that π_t^ξ satisfies a particular Fokker-Planck equation, which differs depending on the method considered, and that π_t^ξ converges in the long-time limit towards the Lebesgue measure λ. Section 4 then states several results on the invariant measure of a generic diffusion, in order to adress the issue of the existence of both stationary measure and stationary biais to equation (1.8), and later handles the robustness of the conservative equilibrium to non-conservative perturbations. Eventually, Section 5 is devoted to the long-time convergence of both the ABF and PABF methods, in the conservative case, with a force $\mathcal{F} = -\nabla V$ (generalizing in particular results from [1]), and in the non-conservative case, with a generic force $\mathcal{F}$.

2. Main results

2.1. Relative entropy and preliminary inequalities

Let us first introduce several tools that will be used in the following. For μ, ν two probability measures on the same space, we will denote by $\mu \ll \nu$ the absolute continuity of μ with respect to ν. Now consider the relative entropy of μ with respect to ν:

$$\mathcal{H}(\mu|\nu) = \begin{cases} \displaystyle\int \ln\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) \mathrm{d}\mu \text{ if } \mu \ll \nu, \\ +\infty \qquad\qquad\qquad \text{otherwise.} \end{cases}$$

Recall the Csiszàr-Kullback inequality:

$$\|\mu - \nu\|_{TV} \leq \sqrt{2\mathcal{H}(\mu|\nu)}\,, \tag{2.1}$$

where $\|\cdot\|_{TV}$ stands for the total variation norm. In particular, while the relative entropy is not a distance (it lacks the symmetry property), its convergence towards zero implies the convergence in total variance norm of μ towards ν.

Similarly, let us define the Fisher information: for $\mu \ll \nu$,

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

The probability measure ν is said to satisfy a Logarithmic Sobolev Inequality $LSI(\rho)$ of constant $\rho > 0$ if:

$$\forall \mu \ll \nu, \qquad \mathcal{H}(\mu|\nu) \leq \frac{1}{2\rho} I(\mu|\nu).$$

From [25], if ν satisfies a log-Sobolev inequality with constant $\rho > 0$, then it also satisfies the so-called Talagrand inequality $\mathcal{T}(\rho)$ with constant $\rho > 0$:

$$\forall \mu \ll \nu, \qquad W_2^2(\mu,\nu) \leq \frac{2}{\rho}\mathcal{H}(\mu|\nu), \tag{2.2}$$

where $W_2(\mu,\nu)$ is the Wasserstein distance with quadratic cost between the probability measures μ and ν. More precisely, if μ and ν are defined on a general Riemannian manifold Ω:

$$W_2^2(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\Omega\times\Omega} \omega(x,y)^2 \,\mathrm{d}\pi(x,y),$$

where ω is the geodesic distance on Ω, and $\Pi(\mu,\nu)$ is the set of coupling probability measures, i.e probability measures on $\Omega \times \Omega$ whose marginals are μ and ν respectively.

In the following, we will slightly abuse notations and denote $I(\mu|\nu)$, $\mathcal{H}(\mu|\nu)$ or $W_2(\mu|\nu)$ both in the case where μ and ν are probability measures, or probability density functions.

2.2. Precise statements of the results

In all this section, π_t satisfies (1.8). First of all, let us consider the equation satisfied by the density π_t^ξ in the general case where $\mathcal{F}$ is either conservative or non-conservative.

Lemma 2.1. *The density π_t^ξ satisfies the following Fokker-Planck equation:*

$$\partial_t \pi_t^\xi \;=\; \Delta \pi_t^\xi - \nabla \cdot \Big((B_t - G_t)\pi_t^\xi \Big) . \tag{2.3}$$

Proof. Take a test function $\varphi \in \mathcal{C}^\infty(\mathbb{T}^m)$. Then, using an integration by parts,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{T}^m} \varphi \pi_t^\xi &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{T}^n} \varphi(x)\pi_t(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{\mathbb{T}^n} \left(\Delta_x \varphi(x) + (\mathcal{F}_1(x,y) + B_t(x))\nabla_x \varphi(x)\right)\pi_t(\mathrm{d}x,\mathrm{d}y) \\
&= \int_{\mathbb{T}^m} \left(\Delta_x\varphi(x)\pi_t^\xi(x) + \left(\int_{\mathbb{T}^{n-m}} \mathcal{F}_1(x,y)\frac{\pi_t(x,y)}{\pi_t^\xi(x)}\mathrm{d}y + B_t(x)\pi_t^\xi(x)\right)\nabla_x\varphi(x)\right)\mathrm{d}x \\
&= \int_{\mathbb{T}^m} \left(\Delta_x\varphi + (B_t - G_t)\nabla_x\varphi\right)\pi_t^\xi .
\end{aligned}$$

□

Remark that in Proposition 2 from [1], the Helmoltz projection is done in $L^2(\pi_t^\xi)$, so that $\nabla\cdot((B_t - G_t)\pi_t^\xi) = 0$ and one ends up with the heat equation. Here, we get the heat equation in the ABF case ($B_t = G_t$) and, in the PABF case ($B_t = \mathsf{P}_{L^2(\lambda)}(G_t)$), an additional time-dependent divergence-free drift.

Remark 2.2. Since the density π_t^ξ, as well as constants, satisfies the Fokker-Planck equation (2.3) which preserves positivity, provided there exists $m_0^\xi > 0$ such that $\pi_0^\xi \geq m_0^\xi$, one can show that $\pi_t^\xi \geq m_0^\xi$ for all $t \geq 0$ on the torus $\mathbb{T}^m$. Note that if π_0^ξ was to be zero at some points or not sufficiently smooth, the conditional mean G_0 given in (1.8) might not be well defined.

In view of Remark 2.2, from now on, assume the following:

Assumption 2.3. *The initial condition π_0 admits a smooth density with respect to the Lebesgue measure, such that π_0^ξ is positive.*

As a consequence, the conditional means G_t are well defined for all $t \geq 0$, along with the entropy $\mathcal{H}(\pi_0 \,|\, \lambda)$, which is ensured to be finite. Furthermore, π_0^ξ belongs to $L^2(\mathbb{T}^m)$.

Both the ABF and PABF algorithms are designed in order to ensure that all the values of the transition coordinate have been visited. In other words, the density of $\xi(X_t, Y_t)$ should converge to a flat histogram, namely the Lebesgue measure λ. In the conservative case, this is known to hold in both the ABF case [20] and the PABF case [1]. We now extend the flat histogram property to the general –possibly non-conservative– case.

Proposition 2.4. *For both the ABF and PABF algorithm, under Assumption 2.3, π_t^ξ converges towards the Lebesgue measure as $t \to \infty$. More precisely, for all $t \geq 0$:*

$$\mathcal{H}(\pi_t^\xi|\lambda) \;\leqslant\; e^{-8\beta^{-1}\pi^2 t}\mathcal{H}(\pi_0^\xi|\lambda) .$$

Furthermore, the entropic convergence of the density can be strengthened to an L^∞ one, that will prove useful in the rest of the study:

Proposition 2.5. *For both the ABF and PABF algorithm, under Assumption 2.3, there exists $C > 0$ such that for all initial distribution $\pi_0^\xi \in L^2(\mathbb{T}^m)$, for all $t \geq 1$:*

$$\|\pi_t^\xi - 1\|_\infty \ \leqslant \ Ce^{-4\beta^{-1}\pi^2 t}\|\pi_0^\xi - 1\|_2 \,.$$

As detailed in [1, 19], in the conservative case $\mathcal{F} = -\nabla V$, $\pi_\infty = \mu_A$ is a stationary state of (1.8). In the non-conservative case, the existence of such a stationary state may be unclear, and this issue will be treated in Theorem 2.7 below, which will pe proved in Section 4.2. For now, let us consider the following assumption:

Assumption 2.6. *The interaction force $\mathcal{F}$ is in $\mathcal{C}^1(\mathbb{T}^n, \mathbb{R}^n)$, and we denote by $M > 0$ a constant such that for all $y \in \mathbb{T}^{n-m}$, $x \mapsto \mathcal{F}_1(x, y)$ is M-Lipschitz.*

Theorem 2.7. *For the ABF (resp. PABF) algorithm, under Assumption 2.6, there exists a couple of stationary measure and bias $\left(\pi_\infty^{\mathcal{F}}, B_\infty^{\mathcal{F}}\right)$ (resp. $\left(\pi_\infty^{\mathcal{F}}, \nabla H_\infty^{\mathcal{F}}\right)$) to (1.8), such that $\pi_\infty^{\mathcal{F}} \in \mathcal{C}^0(\mathbb{T}^n)$ is stricly positive. As a consequence,*

(i) $\pi_\infty^{\mathcal{F}}$ satisfies a log-Sobolev inequality for some constant $R > 0$,
(ii) the conditional density $y \mapsto \pi_{\infty,x}^{\mathcal{F}}(y) := \pi_\infty^{\mathcal{F}}(x,y)/\pi_\infty^{\mathcal{F},\xi}(x)$ satisfies a log-Sobolev inequality for some constant $\rho > 0$, for all $x \in \mathbb{T}^m$.

Remark 2.8. Note that there is no reason whatsoever for $\pi_\infty^{\mathcal{F}}$ to be the same in both the ABF and PABF case. Nevertheless, as shown in Proposition 2.4, $\pi_\infty^\xi \equiv 1$ in all cases.

Remark 2.9. An important remark is that, at small temperatures (*i.e.*, $\beta \gg 1$), the optimal log-Sobolev constant of a probability measure with density proportional to $\exp(\beta W)$ for some W, roughly scales like $\exp(\beta d_W)$ where d_W is the so-called critical depth of W [21] (the critical depth is the highest energy barrier to overcome in order to reach a global minimum of W). If the transition coordinate is well-chosen, the metastability in the orthogonal space should be small, meaning that for all $x \in \mathbb{T}^m$ the critical depth of $W(x, \cdot)$ should be small with respect to the critical depth of W. As a consequence, as a function of β, ρ is expected to be much larger than the log-Sobolev constant of $\mu \propto e^{-\beta V}$, which is the convergence rate to equilibrium of the original (unbiased) dynamics (1.1).

The following result deals with the robustness of the conservative equilibrium to non-conservative perturbations, and will be proved in Section 4.3.

Proposition 2.10. *For the PABF algorithm, under Assumption 2.3 and Assumption 2.6, for all $V \in \mathcal{C}^2(\mathbb{T}^n)$ and $p \geq 1$, there exists $K_V > 0$ and $K_p > 0$ such that the following holds. Denote by A the free energy associated to V (see Eq. (1.3) for the definition of A). For all $\mathcal{F} \in \mathcal{C}^1(\mathbb{T}^n)$ satisfying $\|\mathcal{F} + \nabla V\|_\infty \leq 1$, for all equilibrium measure $\pi_\infty^{\mathcal{F}}$ of (1.8), considering the corresponding bias $\nabla H_\infty^{\mathcal{F}}$, one has*

$$\|\nabla A - \nabla H_\infty^{\mathcal{F}}\|_{L^p(\mathbb{T}^m)} \ \leq \ K_V K_p \|\mathcal{F} + \nabla V\|_\infty \,,$$

and, for all $\psi \in L^\infty(\mathbb{T}^n)$, considering

$$\hat{I}_\psi \ := \ \frac{\int_{\mathbb{T}^n} \psi(x,y) e^{-\beta H_\infty^{\mathcal{F}}(x)} \pi_\infty^{\mathcal{F}}(x,y) \mathrm{d}x\mathrm{d}y}{\int_{\mathbb{T}^n} e^{-\beta H_\infty^{\mathcal{F}}(x)} \pi_\infty^{\mathcal{F}}(x,y) \mathrm{d}x\mathrm{d}y} \,,$$

one has

$$\left| \int_{\mathbb{T}^n} \psi \mathrm{d}\mu - \hat{I}_\psi \right| \ \leq \ K_V \, \|\psi\|_\infty \, \|\mathcal{F} + \nabla V\|_\infty \,.$$

The first point of Proposition 2.10 states that, if the error on the forces ∇V is small, then the bias of the free energy estimation is small. The second point states that similarly, the bias on the computations of averages with respect to μ is small if the error on the forces is small. Indeed, in practice, in order to compute averages with respect to the initial target law μ from the biased trajectory, two strategies are available: either standard importance sampling re-weighting, or estimation of the conditional expectations given $\xi(X,Y)=x$ and then average with respect to $\exp(-H_{\mathcal{F}}(x))$. In both cases, if $-\nabla V$ is replaced by $\mathcal{F}$ due to some numerical errors and the process converges in large time towards an equilibrium $\pi_\infty^{\mathcal{F}}$, then a quantity of the form $\int_{\mathbb{T}^n}\psi\mathrm{d}\mu$ is approximated by an estimator that converges in large time towards the quantity $\hat{I}_\psi$ defined in Proposition 2.10.

Finally, we turn to the long-time convergence of the density π_t on the whole space. The first theorem concerns the classical, conservative case, whereas the second concerns the general case, where the force $\mathcal{F}$ can be non conservative. These will respectively be proved in Sections 5.2 and 5.3.

Theorem 2.11. *Let us consider (π_t, B_t) solution of (1.8) for either the ABF or PABF algorithm, under Assumption 2.3 and Assumption 2.6. Let us suppose moreover that $\mathcal{F}=-\nabla V$, with $V\in\mathcal{C}^2(\mathbb{T}^n)$. Then, there exists $K>0$ such that, for all $\varepsilon>0$ and for all $t\geq 0$:*

$$\mathcal{H}\left(\pi_t|\mu_A\right)\leq K\left(1+\frac{1}{\varepsilon^2}\right)e^{-(\Lambda-\varepsilon)t},$$

with μ_A being given by (1.5), $\Lambda=\left(8\pi^2\wedge 2\rho\right)\beta^{-1}$ in the ABF case, $\Lambda=\left(4\pi^2\wedge 2\rho\right)\beta^{-1}$ in the PABF case, and ρ is the log-Sobolev constant of the conditional density $y\mapsto\mu_{A,x}(y):=\mu_A(x,y)/\mu_A^\xi(x)$. Furthermore, (1.8) consequently admits a unique stationary state: using the notations of Theorem 2.7, $\left(\pi_\infty^{-\nabla V}, B_\infty^{-\nabla V}\right)=(\mu_A,\nabla A)$.

This extends Theorem 1 from [19], which is restricted to the ABF algorithm with $m=1$. Besides, for the PABF algorithm, Theorem 1 [1] is a similar convergence result but for a variant of the algorithm where the classical Helmholtz projection in $L^2(\lambda)$ is replaced by the Helmholtz projection in the weighted space $L^2(\pi_t^\xi)$. This variant is motivated in [1] by some cancellations in the computations of the proofs. Nevertheless, as already noted in [1], the classical Helmholtz projection is used in practice. Theorem 2.11 in the PABF case is thus a new result which fills a gap between the existing theoretical convergence results and the practical algorithm.

Remark 2.12. For $t>0$, applying Theorem 2.11 with $\varepsilon=1/t$ yields

$$\mathcal{H}\left(\pi_t|\mu_A\right)\leq Ke^1(1+t^2)e^{-\Lambda t}.$$

The next results address the general –possibly non-conservative– case, and as such are new.

Theorem 2.13. *Let us consider (π_t, B_t) solution of (1.8) for either the ABF or PABF algorithm, under Assumption 2.3 and Assumption 2.6. Let $\pi_\infty^{\mathcal{F}}$, R,ρ be a stationary measure for (1.8) and the two associated constants, as introduced in Theorem 2.7. Suppose moreover that $M\beta<2\rho$, where M is the constant introduced in Assumption 2.6. Then there exists $K\geq 0$ such that, for all $t\geq 0$:*

$$\mathcal{H}\left(\pi_t|\pi_\infty^{\mathcal{F}}\right)\leq Ke^{-\Lambda t},$$

with $\Lambda=2R(1-\frac{M\beta}{2\rho})\beta^{-1}$. As a consequence, the dynamics (1.8) admits a unique stationary state.

Remark 2.14. The condition $M\beta<2\rho$ is a technical restriction in our proof, we do not think that it is necessary for the result to hold. Notice that ρ depends on β in a non-trivial way. At high temperature, the conditional densities $\pi_{\infty,x}^{\mathcal{F}}$ are bounded above and below by constants which are uniform in $\beta\leqslant 1$ (it is easily seen that they converge to the uniform density as $\beta\to 0$) which, by classical perturbation results for log-Sobolev inequalities, implies that ρ is bounded uniformly in $\beta\leqslant 1$, and thus the condition is satisfied for β small enough. To fix some ideas in the low temperature case, consier the conservative case where $\mathcal{F}=-\nabla V$. If $y\mapsto V(x,y)$ is strongly convex uniformly in x, by the Bakry-Emery curvature criterion (see *e.g.*, [3]), $\rho \geqslant \beta \inf \mathrm{Hess_y}(V)$. The assumption is thus implied by $M < 2 \inf \mathrm{Hess_y}(V)$, which simply means that the force should be "flatter" in the direction of $\xi(x,y) = x$ than in the orthogonal direction y. This is somehow the ideal situation for the use of ABF, namely a fast contraction orthogonally to the reaction coordinate.

Eventually, one has the following result, which will be proved in Section 5.4.

Corollary 2.15. *Under the settings of either Theorem 2.11 or 2.13, there exists a unique stationary state* $\left(\pi_\infty^{\mathcal{F}}, B_\infty^{\mathcal{F}}\right)$ *for the dynamics* (1.8)*. Furthermore, there exists* $K \geq 0$ *such that for all* $t \geq 0$*,*

$$\int_{\mathbb{T}^m} |B_t - B_\infty^{\mathcal{F}}|^2 \mathrm{d}x \leq K e^{-\Lambda t},$$

where Λ *is given by either Theorem 2.11 (where* $\mathcal{F} = -\nabla V$*) or 2.13 (where* $\mathcal{F}$ *is general).*

Remark 2.16. A direct consequence of the Csizàr-Kullback inequality (2.1) combined with either Theorem 2.11 or Theorem 2.13 is that for all $t \geq 0$

$$\|\pi_t - \pi_\infty^{\mathcal{F}}\|_{TV} \leq \sqrt{2K} e^{-\frac{1}{2}\Lambda t},$$

where $K, \Lambda \geq 0$ are given by either Theorem 2.11 (where $\mathcal{F} = -\nabla V$) or 2.13 (where $\mathcal{F}$ is general).

Theorem 2.13 shows the exponential convergence to a unique stationary state for the ABF and PABF algorithms even for non-conservative forces. Notice that the rate of convergence obtained in Theorem 2.11 for conservative forces is better than the rate of convergence in Theorem 2.13. It would be interesting to further investigate the sharpness of these rates.

The rest of this paper is devoted to the proofs of the results stated in this section. From now on, and without loss of generality, we will assume that $\beta = 1$. Note that the assumption of Theorem 2.13 now becomes $M < 2\rho$. An adequate change of variable to then deduce the results for $\beta \neq 1$ is: $\tilde{t} = \beta^{-1} t$, $\tilde{\mathcal{F}}(x,y) = \beta \mathcal{F}(x,y)$, $\tilde{W}^1(x) = \beta W^1(x)$, $\tilde{W}^2(y) = \beta W^2(y)$, and $\tilde{\pi}_t(x,y) = \pi_t(x,y)$, for all $t \geq 0$ and for all $(x,y) \in \mathbb{T}^n$.

3. Law of the transition coordinate

After proving in Section 3.1 the long-time entropic convergence of the density π_t^ξ towards the Lebesgue measure λ, we prove in Section 3.2 its long-time L^∞-convergence, by relying on a Nash inequality on the n-dimensional torus and on the proof of ([3], Thm 6.3.1).

3.1. Proof of Proposition 2.4

Proof. One has:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{H}(\pi_t^\xi | \lambda) = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{T}^m} \pi_t^\xi \ln \pi_t^\xi.$$

Considering $\mathcal{L}_t \mu = \nabla \cdot (\nabla \mu - (B_t - G_t)\mu)$:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{T}^m} \pi_t^\xi = \int_{\mathbb{T}^m} \mathcal{L}_t \pi_t^\xi = 0.$$

One gets, using integration by parts,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(\pi_t^\xi|\lambda) &= \int_{\mathbb{T}^m} \ln \pi_t^\xi \mathcal{L}_t \pi_t^\xi + \int_{\mathbb{T}^m} \mathcal{L}_t \pi_t^\xi \\
&= -\int_{\mathbb{T}^m} \frac{|\nabla \pi_t^\xi|^2}{\pi_t^\xi} + \int_{\mathbb{T}^m} (B_t - G_t)\nabla \pi_t^\xi \\
&= -\int_{\mathbb{T}^m} \frac{|\nabla \pi_t^\xi|^2}{\pi_t^\xi} \quad (\text{since } \nabla \cdot (B_t - G_t) = 0) \\
&= -\int_{\mathbb{T}^m} |\nabla \ln \left(\frac{\pi_t^\xi}{\lambda}\right)|^2 \, \pi_t^\xi \\
&= -I(\pi_t^\xi|\lambda).
\end{aligned} \tag{3.1}$$

Since the Lebesgue measure λ satisfies a log-Sobolev inequality of constant $4\pi^2$ Proposition 5.7.5(ii) from [3], we have:

$$\partial_t \mathcal{H}(\pi_t^\xi|\lambda) \leq -2(4\pi^2)\mathcal{H}(\pi_t^\xi|\lambda),$$

which concludes the proof of Proposition 2.4, denoting by $\pi_\infty^\xi \equiv \lambda$ the long-time limit of π_t^ξ. □

3.2. Proof of Proposition 2.5

We first state a Nash inequality on the n-dimensional torus.

Lemma 3.1. *For all $n \in \mathbb{N}^*$, there exists $a = a(n) > 0$ such that for all function $u \in H^1(\mathbb{T}^n)$:*

$$\|u\|_2^2 \leq 2\|u\|_1^2 + a\|\nabla u\|_2^{\frac{2n}{n+2}} \|u\|_1^{\frac{4}{n+2}}. \tag{3.2}$$

Proof. Let us recall that $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$. We consider $L^2(\mathbb{T}^n)$ equipped with the inner product $\langle u, v\rangle := \int_{\mathbb{T}^n} u(x)\bar{v}(x)\,\mathrm{d}x$. The sequence $\{e^{2\pi i k x}\}_{k\in\mathbb{Z}^n}$ is an orthonormal basis of $L^2(\mathbb{T}^n)$. Now given a function $u \in L^2(\mathbb{T}^n)$ and its Fourier coefficients

$$c_k = \int_{\mathbb{T}^n} u(x) e^{-2\pi i k x}\,\mathrm{d}x, \qquad \forall k \in \mathbb{Z}^n,$$

denoting by $k = (k_1, \ldots, k_n)$ a vector in $\mathbb{Z}^n$, and $|k| = \sqrt{\sum_{j=1}^n |k_j|^2}$, the Parseval identity yields

$$\|u\|_2^2 = \sum_{k\in\mathbb{Z}^n} |c_k|^2\,, \qquad \|\nabla u\|_2^2 = \sum_{k\in\mathbb{Z}^n} |k|^2\,|c_k|^2.$$

Let $\rho > 0$ to be fixed later on. One has, considering $\|k\|_\infty = \max_{j\in[\![1,n]\!]}\{|k_j|\}$:

$$\begin{aligned}
\|u\|_2^2 = \sum_{k\in\mathbb{Z}^n} |c_k|^2 &= \sum_{\|k\|_\infty \leq \rho} |c_k|^2 + \sum_{\|k\|_\infty > \rho} |c_k|^2 \\
&\leq \sum_{\|k\|_\infty \leq \rho} |c_k|^2 + \frac{1}{\rho^2} \sum_{\|k\|_\infty > \rho} \|k\|_\infty^2 |c_k|^2 \\
&\leq \sum_{\|k\|_\infty \leq \rho} |c_k|^2 + \frac{1}{\rho^2} \sum_{\|k\|_\infty > \rho} |k|^2 |c_k|^2 \\
&\leq \sum_{\|k\|_\infty \leq \rho} |c_k|^2 + \frac{1}{\rho^2} \|\nabla u\|_2^2.
\end{aligned}$$

And:

$$\sum_{\|k\|_\infty \le \rho} |c_k|^2 = \sum_{\|k\|_\infty \le \rho} \left| \int_{\mathbb{T}^n} u(x) e^{-2\pi i k x} \, \mathrm{d}x \right|^2 \ \le \ \|u\|_1^2 \sum_{\|k\|_\infty \le \rho} 1 \ \le \ (2\rho+1)^n \|u\|_1^2.$$

Consequently:

$$\|u\|_2^2 \le 3^n (\rho \vee 1)^n \|u\|_1^2 + \frac{1}{\rho^2} \|\nabla u\|_2^2. \tag{3.3}$$

We now distinguish between two cases:

(i) If $3^n \|u\|_1^2 \le \|\nabla u\|_2^2$, by choosing

$$\rho = 3^{-\frac{n}{n+2}} \frac{\|\nabla u\|_2^{\frac{2}{n+2}}}{\|u\|_1^{\frac{2}{n+2}}} \ge 1,$$

inequality (3.3) yields:

$$\begin{aligned} \|u\|_2^2 &\le 3^n 3^{-\frac{n^2}{n+2}} \|\nabla u\|_2^{\frac{2n}{n+2}} \|u\|_1^{\frac{4}{n+2}} + 3^{\frac{2n}{n+2}} \|\nabla u\|_2^{\frac{2n}{n+2}} \|u\|_1^{\frac{4}{n+2}} \\ &= 2 \cdot 3^{\frac{2n}{n+2}} \|\nabla u\|_2^{\frac{2n}{n+2}} \|u\|_1^{\frac{4}{n+2}}. \end{aligned} \tag{3.4}$$

(ii) If $3^n \|u\|_1^2 \ge \|\nabla u\|_2^2$, one wishes to rely on the Poincaré-Wirtinger inequality on the torus $\mathbb{T}^n$. The optimal Poincaré constant in $H_0^1(\mathbb{T}^n)$ being equal to λ_1^{-1}, where $\lambda_1 = 4\pi^2$ is the first non trivial eigenvalue of the negative Laplacian $-\Delta$, one can consider the following Poincaré-Wirtinger inequality:

$$\|u - \bar{u}\|_2^2 \le \frac{1}{4\pi^2} \|\nabla u\|_2^2, \qquad \forall u \in H^1(\mathbb{T}^n), \tag{3.5}$$

where $\bar{u} = \displaystyle\int_{\mathbb{T}^n} u(x) \, \mathrm{d}x$. One consequently gets:

$$\begin{aligned} \|u\|_2^2 &\le 2\bar{u}^2 + 2\|u - \bar{u}\|_2^2 \\ &\le 2\|u\|_1^2 + 2\frac{1}{4\pi^2} \|\nabla u\|_2^2 \\ &= 2\|u\|_1^2 + \frac{1}{2\pi^2} \|\nabla u\|_2^{\frac{2n}{n+2}} \|\nabla u\|_2^{\frac{4}{n+2}} \\ &\le 2\|u\|_1^2 + \frac{1}{2\pi^2} 3^{\frac{2n}{n+2}} \|\nabla u\|_2^{\frac{2n}{n+2}} \|u\|_1^{\frac{4}{n+2}}. \end{aligned} \tag{3.6}$$

Combining (3.4) and (3.6), one obtains:

$$\|u\|_2^2 \le 2\|u\|_1^2 + 3^{\frac{2n}{n+2}} \left(\frac{1}{2\pi^2} \vee 2 \right) \|\nabla u\|_2^{\frac{2n}{n+2}} \|\nabla u\|_1^{\frac{4}{n+2}}$$

which yields (3.2), with $a = 2 \cdot 3^{\frac{2n}{n+2}}$. □

We are now in position to prove Proposition 2.5.

Proof of Proposition 2.5. We will rely on the idea of the proof of ([3], Thm. 6.3.1). Let us start with two preliminary results. Let $\varphi \in \mathcal{C}^\infty(\mathbb{T}^m)$ be a test function and consider:

$$\forall z \in \mathbb{T}^m, \qquad \varphi_t(z) = \mathbb{E}_z[\varphi(Z_t)] = \mathbb{E}[\varphi(Z_t) \,|\, Z_0 = z],$$

where $(Z_t)_{t\geq 0}$ satisfies the following dynamics:

$$\mathrm{d}Z_t = (B_t - G_t)(Z_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t,$$

where $(W_t)_{t\geq 0}$ is a n-dimensional Brownian motion and $\nabla \cdot (B_t - G_t) = 0$. Let ν_Z be the invariant measure of this dynamics, $\mathcal{L} = (B_t - G_t) \cdot \nabla + \Delta$ its infinitesimal generator, and $\mathcal{L}^* = -\nabla \cdot (B_t - G_t) + \Delta$ its adjoint in $L^2(\nu_Z)$. Using Itô calculus, φ_t satisfies:

$$\varphi_0 = \varphi, \quad \partial_t \varphi_t = \Delta\varphi_t + (B_t - G_t) \cdot \nabla\varphi_t \tag{3.7}$$

which is equivalent to

$$\varphi_0 = \varphi, \quad \partial_t \varphi_t = \Delta\varphi_t + \nabla \cdot ((B_t - G_t)\varphi_t).$$

Given the result of Lemma 2.1, $\pi_t^\xi - 1$ satisfies:

$$\partial_t \left(\pi_t^\xi - 1\right) = \Delta \left(\pi_t^\xi - 1\right) - \nabla \cdot \left((B_t - G_t)(\pi_t^\xi - 1)\right). \tag{3.8}$$

For a fixed $t > 0$, one as, for all $0 \leq s \leq t$:

$$\frac{\mathrm{d}}{\mathrm{d}s} \int_{\mathbb{T}^n} \varphi_{t-s} \left(\pi_s^\xi - 1\right) = -\int_{\mathbb{T}^n} \mathcal{L}\varphi_{t-s} \left(\pi_s^\xi - 1\right) + \int_{\mathbb{T}^n} \varphi_{t-s}\mathcal{L}^* \left(\pi_s^\xi - 1\right) = 0.$$

Integrating between $s = 0$ and $s = t$ yields

$$\int_{\mathbb{T}^m} \varphi_t(\pi_0^\xi - 1) = \int_{\mathbb{T}^m} \varphi(\pi_t^\xi - 1), \qquad \forall t \geq 0. \tag{3.9}$$

Second, for all $t \geq 0$,

$$\|\varphi_t\|_1 \leq \|\varphi\|_1. \tag{3.10}$$

Indeed, one has on the torus $\mathbb{T}^m$:

$$\|\varphi_t\|_1 \leq \int_{\mathbb{T}^m} \psi(t, z)\,\mathrm{d}z$$

where, for all $t \geq 0$ and $z \in \mathbb{T}^n$, $\psi(t, z) = \mathbb{E}[\,|\varphi(Z_t)|\,|\, Z_0 = z] \geq 0$ satisfies (3.7) with initial condition $\psi(0, .) = |\varphi(.)| \geq 0$ on $\mathbb{T}^m$. Integrating by parts and using that $\nabla\cdot(B_t - G_t) = 0$ one can check that $\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{T}^m} \psi = 0$, so that:

$$\int_{\mathbb{T}^m} \psi(t, z)\,\mathrm{d}z = \int_{\mathbb{T}^m} \psi(0, z)\,\mathrm{d}z = \|\varphi\|_1, \quad \forall t \geq 0,$$

hence the result.

Step 1: Now let us show that there exists $\mathcal{C} > 0$ such that, for all $t > 0$,

$$\|\varphi_t\|_2^2 \leq \left(\mathcal{C}t^{-\frac{m}{2}} + 2\right) \|\varphi\|_1^2.$$

To do so, consider for all $t \geq 0$,

$$\Lambda(t) = \int_{\mathbb{T}^m} |\varphi_t|^2.$$

Since $\nabla \cdot (B_t - G_t) = 0$ one can show from (3.7) that:

$$\Lambda'(t) = -2\int_{\mathbb{T}^m} |\nabla\varphi_t|^2.$$

Knowing that $\|\varphi_t\|_1 \le \|\varphi\|_1$ for all time $t \ge 0$, we use the inequality (3.2) given by Lemma 3.1 to obtain:

$$\Lambda(t) \le 2\|\varphi\|_1^2 + a\left[-\frac{1}{2}\Lambda'(t)\right]^{\frac{m}{m+2}} \|\varphi\|_1^{\frac{4}{m+2}}.$$

Consider for all $t \ge 0$, $g(t) = \Lambda(t) - 2\alpha$, where $\alpha = \|\varphi\|_1^2$. By construction, g is decreasing on $\mathbb{R}^+$. We distinguish between three cases:

(i) Assume that $g(0) \le 0$. In this case, $g(t) \le 0$ for all $t \ge 0$ and, for all $t \ge 0$:

$$\|\varphi_t\|_2^2 \le 2\|\varphi\|_1^2.$$

(ii) Assume that $g(t) > 0$ for all $t \ge 0$. Then:

$$\begin{aligned} g(t) \le a\alpha^{\frac{2}{m+2}}\left[-\frac{1}{2}g'(t)\right]^{\frac{m}{m+2}} &\Leftrightarrow g(t)^{\frac{m+2}{m}} \le -\frac{1}{2}a^{\frac{m+2}{m}}\alpha^{\frac{2}{m}}g'(t)\\ &\Leftrightarrow g'(t) \le -2\cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}g(t)^{\frac{m+2}{m}}\\ &\Leftrightarrow g'(t)g^{-\frac{m+2}{m}} \le -2\cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}\\ &\Leftrightarrow -\frac{m}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(g(t)^{-\frac{2}{m}}\right) \le -2\cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}\\ &\Leftrightarrow \frac{\mathrm{d}}{\mathrm{d}t}\left(g(t)^{-\frac{2}{m}}\right) \ge \frac{4}{m}\cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}. \end{aligned}$$

Integrating between 0 and t yields:

$$\begin{aligned} g(t)^{-\frac{2}{m}} \ge g(0)^{-\frac{2}{m}} + \frac{4}{m}\cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}t &\Leftrightarrow g(t)^{\frac{2}{m}} \le \frac{1}{g(0)^{-\frac{2}{m}} + \frac{4}{m}\cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}t}\\ &\Leftrightarrow g(t)^{\frac{1}{m}} \le \frac{1}{\left(g(0)^{-\frac{2}{m}} + \frac{4}{m}\cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}t\right)^{\frac{1}{2}}}\\ &\Rightarrow g(t)^{\frac{1}{m}} \le \frac{1}{\left(\frac{4}{m}\cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}t\right)^{\frac{1}{2}}}\\ &\Leftrightarrow g(t) \le 2^{-m}a^{\frac{m+2}{2}}m^{\frac{m}{2}}\alpha t^{-\frac{m}{2}}. \end{aligned}$$

Eventually for all $t \ge 0$:

$$\|\varphi_t\|_2^2 \le \left(\mathcal{C}t^{-\frac{m}{2}} + 2\right)\|\varphi\|_1^2$$

with $\mathcal{C} = 2^{-m}a^{\frac{m+2}{2}}m^{\frac{m}{2}} > 0$.

(iii) Assume that $g(0) > 0$ and let us assume that $t^* > 0$ is the smallest time t such that $g(t^*) \le 0$. In this case, using the above reasonings, one obtains:

a) For all $t \ge t^*$, $g(t) \le g(t^*) < 0$ and thus

$$\|\varphi_t\|_2^2 \le 2\|\varphi\|_1^2.$$

b) For all $t \in [0, t^*[$, $g(t) > 0$ and thus:

$$\|\varphi_t\|_2^2 \le \left(\mathcal{C}t^{-\frac{m}{2}} + 2\right)\|\varphi\|_1^2.$$

Hence, for all $t \geq 0$, $\|\varphi_t\|_2^2 \leq \left(\mathcal{C}t^{-\frac{m}{2}} + 2\right) \|\varphi\|_1^2$.

Step 2: Now, for all $t \geq 0$, equation (3.9) yields:

$$\left|\int_{\mathbb{T}^m} \varphi \left(\pi_t^\xi - 1\right)\right|^2 = \left|\int_{\mathbb{T}^m} \varphi_t \left(\pi_0^\xi - 1\right)\right|^2.$$

Hence, for all $t \geq 0$:

$$\begin{aligned}\left|\int_{\mathbb{T}^m} \varphi \left(\pi_t^\xi - 1\right)\right|^2 &\leq \|\varphi_t\|_2^2 \|\pi_0^\xi - 1\|_2^2 \\ &\leq \left(\mathcal{C}t^{-\frac{m}{2}} + 2\right) \|\varphi\|_1^2 \|\pi_0^\xi - 1\|_2^2 \quad \text{(using Inequality (3.10))}\end{aligned}$$

Since this is true for any function $\varphi \in L^1(\mathbb{T}^m)$, by duality, for all $t \geq 0$:

$$\|\pi_t^\xi - 1\|_\infty \leq \sqrt{\left(\mathcal{C}t^{-\frac{m}{2}} + 2\right)} \|\pi_0^\xi - 1\|_2. \tag{3.11}$$

Step 3: Considering the equation satisfied by π_t^ξ given in Lemma 2.1, with initial condition π_s^ξ with $s \geq 0$, and using inequality (3.11) over the time interval $[s, s+1]$, there exists $\mathcal{K} = \mathcal{K}(m) > 0$ such that:

$$\|\pi_{s+1}^\xi - 1\|_\infty \leq \mathcal{K} \|\pi_s^\xi - 1\|_2.$$

Denote by $H_0^1(\mathbb{T}^n)$ the closure of the space $\mathcal{C}_0^\infty(\mathbb{T}^n)$ of indefinitely differentiable functions with compact support, with respect to the Sobolev norm $\|\cdot\|_{H^1}$. Using the same reasoning as in the proof of Proposition 2.4, since $\int_{\mathbb{T}^m}(\pi_t^\xi - 1) = 0$, $(\pi_t^\xi - 1)$ belongs in $H_0^1(\mathbb{T}^n)$, and, using equation (3.8) and the Poincaré-Wirtinger inequality (3.5), one has:

$$\|\pi_t^\xi - 1\|_2 \leq \|\pi_0^\xi - 1\|_2 e^{-4\pi^2 t}, \quad \forall t \geq 0. \tag{3.12}$$

Eventually, for all $t \geq 1$:

$$\|\pi_t^\xi - 1\|_\infty \leq \mathcal{K} \|\pi_{t-1}^\xi - 1\|_2 \leq \mathcal{K} e^{-4\pi^2(t-1)} \|\pi_0^\xi - 1\|_2,$$

which concludes the proof with $\mathcal{C} = \mathcal{K} e^{4\pi^2}$. □

Remark 3.2. Note that one could use the maximum principle for times $t \in [0,1]$ in order to replace the right-hand term $\|\pi_0^\xi - 1\|_2$ by the L^∞-norm $\|\pi_0^\xi - 1\|_\infty$. Indeed, since by Assumption 2.3, π_0^ξ is continuous on $\mathbb{T}^m$, one has a uniform bound on $\pi_0^\xi - 1$. Nevertheless, considering an L^2-bound highlights the fact that the uniform bound at time 0 is not essential to the proof of Proposition 2.5, which could be useful for possible generalizations to non-bounded state space cases.

4. Existence of a stationary measure

In Section 4.1 we state and prove preliminary estimates on the invariant probability measures of homogeneous diffusions. We then proceed in Section 4.2 to prove Theorem 2.7, which gives the existence of a stationary state to (1.8) in the general case, where the force $\mathcal{F}$ can be non-conservative. Eventually, one can find in Section 4.3 the proof of Proposition 2.10 where one establishes bounds on the bias of the free energy estimation and on the bias on the computations of averages with respect to μ.

4.1. Preliminary estimates for homogeneous diffusions

The next section is concerned with the sensitivity of the equilibrium measure of a diffusion with respect to its drift, when this drift is in L^p for some p. Consider the following process on $\mathbb{T}^n$, with $n \geq 1$:

$$\mathrm{d}X_t = a(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t \tag{4.1}$$

with $(W_t)_{t\geq 0}$ a classical n-dimensional Brownian motion on the torus $\mathbb{T}^n$ and $a \in L^p(\mathbb{T}^n, \mathbb{R}^n)$ for $p \geqslant 2$ with $p > n$. We refer to [17] for a probabilist study of this SDE (existence, strong Markov and Feller properties, existence and Hlder continuity of the transition kernel, etc.). In the following we take a PDE point of view, namely we are interested in the existence, uniqueness and properties of a solution ν in $H^1(\mathbb{T}^n)$ such that $\int_{\mathbb{T}^n} \nu(x)\,\mathrm{d}x = 1$ of the following equation:

$$\forall \varphi \in H^1(\mathbb{T}^n), \qquad \int_{\mathbb{T}^n} \left(a(z)\cdot\nabla\varphi(z)\nu(z) - \nabla\varphi(z)\cdot\nabla\nu(z)\right)\mathrm{d}z \ = \ 0\,. \tag{4.2}$$

This implies in particular that $\int_{\mathbb{T}^n}(\mathcal{L}\varphi)\nu = 0$ for all $\varphi \in \mathcal{C}^2(\mathbb{T}^n)$ with $\mathcal{L}$ being the generator of (4.1).

Remark 4.1. Note that the Sobolev embedding $H^1 \hookrightarrow L^q$ for some q such that $\frac{1}{q} > \frac{1}{2} - \frac{1}{n}$ and the assumption that $p > n$ ensure that the integrals in (4.2) are well defined for all ν, φ in $H^1(\mathbb{T}^n)$ and all $a \in L^p(\mathbb{T}^n, \mathbb{R}^n)$.

Proposition 4.2. *Let $\mathfrak{M} > 0$ and $p > n \geqslant 1$ with $p \geqslant 2$. There exists $C > 0$ which depends solely on $\mathfrak{M}, p$ and n, such that the following holds. For all $a \in L^p(\mathbb{T}^n, \mathbb{R}^n)$ such that $\|a\|_{L^p(\mathbb{T}^n)} \leqslant \mathfrak{M}$, there exists a unique probability density $\nu_a \in H^1(\mathbb{T}^n)$ that solves (4.2), and which is such that*

$$\|\nu_a\|_\infty + \|1/\nu_a\|_\infty + \|\nu_a\|_{H^1(\mathbb{T}^n)} \ \leqslant \ C\,.$$

Moreover, if ν_b is the solution of (4.2) with a replaced by $b \in L^p(\mathbb{T}^n, \mathbb{R}^n)$ with $\|b\|_{L^p(\mathbb{T}^n)} \leqslant \mathfrak{M}$, then

$$\|\nu_a - \nu_b\|_{L^2(\mathbb{T}^n)} \leq C\|a - b\|_{L^2(\mathbb{T}^n)}.$$

Remark 4.3. In the case of a gradient drift $a = -\nabla\mathcal{A}$, the invariant measure ν_a is explicit: for all $z \in \mathbb{T}^n$,

$$\nu_a(z) = \frac{1}{Z_{\mathcal{A}}}e^{-\mathcal{A}(z)}, \qquad Z_{\mathcal{A}} = \int_{\mathbb{T}^n} e^{-\mathcal{A}(z)}\,\mathrm{d}z,$$

and the $\mathcal{L}^\infty$-bound of Proposition 4.2 amounts to the continuous injection given by Morrey's inequality Theorem IX.12 from [5],

$$W^{1,p}(\mathbb{T}^n) \hookrightarrow L^\infty(\mathbb{T}^n), \qquad \forall p > n. \tag{4.3}$$

Indeed, if $\mathcal{A} \in W^{1,p}(\mathbb{T}^n)$, then $\mathcal{A} \in L^\infty(\mathbb{T}^n)$ and ν_a is bounded from above and below (and conversely if ν_a is bounded above and below then $\mathcal{A}$ is bounded). In particular, since this injection is false for $p \leqslant n$, we see that the condition $p > n$ is necessary in Proposition 4.2 .

Proof. **Step 1:**
First assume that $a \in \mathcal{C}^\infty(\mathbb{T}^n, \mathbb{R}^n)$. By Theorem 5.11 from [8], there exists a Markov process $(\tilde{X}_t)_{t\geq 0}$ on $\mathbb{R}^n$ whose transition probability density is given by the fundamental solution of the equation $\partial_t\tilde{f}_t = -\mathrm{div}\left(a\tilde{f}_t - \nabla\tilde{f}_t\right)$, where a is seen as a 1-periodic function on $\mathbb{R}^n$. Note that by Theorem 0.5 and Condition 0.24.A1 from [9], the density $\tilde{f}_t$ is strictly positive and depends continuously on the initial condition. Moreover, ([8], Thms. 11.4 and 11.5) yield that $(\tilde{X}_t)_{t\geq 0}$ solves the stochastic differential equation (4.1) on $\mathbb{R}^n$. Now, consider $(X_t)_{t\geqslant 0}$ the image of $(\tilde{X}_t)_{t\geqslant 0}$ by the canonical projection from $\mathbb{R}^n$ to $\mathbb{T}^n$. Since a is periodic, $(X_t)_{t\geqslant 0}$ solves (4.1) as an equation on $\mathbb{T}^n$, and thus, using It's formula, it is a Markov process (the proof is the same as Theorems 11.5 from [8] in $\mathbb{R}^n$). Denote by $(P_t)_{t\geqslant 0}$ the associated Markov semigroup on $L^\infty(\mathbb{T}^n)$.

The positivity and continuity in the initial condition of $\tilde{f}_t$ implies that, for all $t > 0$, there exists $r_t > 0$ such that for all $x \in \mathbb{T}^n$ and all Borel set A of $\mathbb{T}^n$, $\mathbb{P}_x(X_t \in A) \geqslant r_t \lambda(A)$, namely the process satisfies a uniform Doeblin condition. In particular, for a fixed $t > 0$, the Markov chain with transition operator P_t is recurrent and irreducible and thus, by Theorem 10.0.1 from [22], it admits a unique invariant measure ν_a. Now, for $s \geqslant 0$, $(\nu_a P_s)P_t = (\nu_a P_t)P_s = \nu_a P_s$, which means that $\nu_a P_s$ is an invariant measure for P_t. Hence by uniqueness, $\nu_a P_s = \nu_a$ for all $s \geqslant 0$. In other words, ν_a is the unique invariant measure for the semigroup $(P_t)_{t\geq 0}$.

Now, let $\varphi \in \mathcal{C}^2(\mathbb{T}^n)$. Denoting by $\mathcal{L} = a \cdot \nabla + \Delta$ the infinitesimal generator of (4.1) and using It's formula, one gets for all $t \geq 0$

$$0 \;=\; \nu_a\left(P_t(\varphi) - \varphi\right) \;=\; \int_0^t \nu_a P_s \mathcal{L}\varphi \mathrm{d}s \;=\; t\,\nu_a(\mathcal{L}\varphi)\,.$$

In other words, ν_a is a solution of the weak equation

$$\forall \varphi \in \mathcal{C}^2(\mathbb{T}^n), \qquad \nu_a(\mathcal{L}\varphi) = 0. \tag{4.4}$$

By elliptic regularity (*e.g.*, [14] applied to ν_a seen as a periodic measure on $\mathbb{R}^n$), ν_a has then a $\mathcal{C}^\infty$ density (that we still denote by ν_a) and, integrating by parts, we can write (4.4) as

$$\int_{\mathbb{T}^n} \left(a(z)\cdot\nabla\varphi(z)\nu_a(z) - \nabla\varphi(z)\cdot\nabla\nu_a(z)\right)\mathrm{d}z \;=\; 0$$

for all $\varphi \in \mathcal{C}^2(\mathbb{T}^n)$ and thus, by density, for all $\varphi \in H^1(\mathbb{T}^n)$. This is (4.2).

Define $\tilde{\nu}_a$ on $\mathbb{R}^n$ by $\tilde{\nu}_a(x+k) = \nu_a(x)$ for all $k \in \mathbb{Z}^n$ and $x \in \mathbb{T}^n$ (seen as $[0,1]^n$). It is such that

$$\forall \varphi \in H^1(\mathbb{R}^n), \qquad \int_{\mathbb{R}^n} \left(a(z)\cdot\nabla\varphi(z)\tilde{\nu}_a(z) - \nabla\varphi(z)\cdot\nabla\tilde{\nu}_a(z)\right)\mathrm{d}z \;=\; 0\,,$$

where, again, a is seen as a 1-periodic function. Since $p > n$, using the notations of [4] and applying the Harnack inequality Corollary 1.7.2 from [4], with the operator $L_{I_n,a,0}$ (I_n being the identity matrix of size n) and the domain $\Omega = [-1,2]^n$ which stricly contains $[0,1]^n$, we get that there exists $C_1 > 0$ depending only on M, p and n such that:

$$\sup_{z\in[0,1]^n} \tilde{\nu}_a(z) \leq C_1 \inf_{z\in[0,1]^n} \tilde{\nu}_a(z).$$

Using that $\int_{\mathbb{T}^n} \nu_a = 1$, this implies that

$$1 \leqslant \sup_{z\in\mathbb{T}^n} \nu_a \leqslant C_1 \inf_{z\in\mathbb{T}^n} \nu_a \leqslant C_1\,. \tag{4.5}$$

Taking $\varphi = \nu_a$ in (4.2) and using the Cauchy-Schwarz inequality yields

$$\int_{\mathbb{T}^n} |\nabla\nu_a|^2 \;=\; \int_{\mathbb{T}^n} a\cdot\nabla\nu_a\nu_a \;\leqslant\; \|\nu_a\|_\infty \|a\|_{L^2(\mathbb{T}^n)} \|\nabla\nu_a\|_{L^2(\mathbb{T}^n)},$$

hence $\|\nabla\nu_a\|_{L^2(\mathbb{T}^n)} \leqslant \mathfrak{M}C_1$. Consequently, using the Poincaré-Wirtinger inequality (3.5), $\|\nu_a\|_{H^1(\mathbb{T}^n)} \leqslant C_2$ for some $C_2 > 0$ that depends only on $\mathfrak{M}, p, n$.

Step 2: Now we consider $a \in L^p(\mathbb{T}^n,\mathbb{R}^n)$, with $\|a\|_{L^p(\mathbb{T}^n)} \leq \mathfrak{M}$, and proceed to prove the existence of a solution ν_a to equation (4.2). Let $(a_k)_{k\in\mathbb{N}}$ be a sequence of $\mathcal{C}^\infty$ functions that converges to a in $L^p(\mathbb{T}^n)$ and such that $\|a_k\|_{L^p(\mathbb{T}^n)} \leqslant \mathfrak{M}$ for all $k \in \mathbb{N}$. Let $(\nu_{a_k})_{k\in\mathbb{N}}$ be the associated solutions of (4.2) given in Step 1. From Step 1, $(\nu_{a_k})_{k\in\mathbb{N}}$ is bounded in $H^1(\mathbb{T}^n)$, and thus we can consider a subsequence that converges weakly in H^1 and strongly in L^2 to some $\nu_a \in H^1(\mathbb{T}^n)$. The weak convergence in H^1 implies that ν_a solves (4.2) and $\|\nu_a\|_{H^1(\mathbb{T}^n)} \leqslant C_2$. The L^2-convergence implies that ν_a is a probability density.

Step 3: Let us now consider any solution and establish bounds similar to the previous step and a Poincaré inequality. For $a \in L^p(\mathbb{T}^n, \mathbb{R}^n)$, let $\nu_a \in H^1(\mathbb{T}^n)$ be any probability density solution of (4.2). Using again ([4], Cor. 1.7.2) and the fact that the mass of ν_a is 1, we get that $1/C_1 \leqslant \nu_a \leqslant C_1$ with the same constant C_1. From this, as in Step 1, we also get that $\|\nu_a\|_{H^1(\mathbb{T}^n)} \leqslant C_2$, with the same constant C_2. The Poincaré-Wirtinger inequality (3.5), together with the lower and upper bounds on ν_a classically yields a Poincaré inequality for ν_a. Indeed, for any $\varphi \in H^1(\nu_a)$, $\int_{\mathbb{T}^n} \varphi\nu_a$ is the minimizer in $\mathbb{R}$ of $c \mapsto \int_{\mathbb{T}^n} (\varphi - c)^2\nu_a$, so that

$$
\begin{aligned}
\int_{\mathbb{T}^n} \left(\varphi - \int_{\mathbb{T}^n} \varphi\nu_a\right)^2 \nu_a \;\leqslant\; \int_{\mathbb{T}^n} \left(\varphi - \int_{\mathbb{T}^n} \varphi\right)^2 \nu_a \;&\leqslant\; C_1 \int_{\mathbb{T}^n} \left(\varphi - \int_{\mathbb{T}^n} \varphi\right)^2 \\
&\leqslant\; \frac{C_1}{4\pi^2} \int_{\mathbb{T}^n} |\nabla\varphi|^2 \\
&\leqslant\; \frac{C_1^2}{4\pi^2} \int_{\mathbb{T}^n} |\nabla\varphi|^2 \nu_a \,.
\end{aligned} \tag{4.6}
$$

Step 4: We now proceed to the proof of the last part of the proposition, from which the uniqueness of ν_a immediately follows. Let $a, b \in L^p(\mathbb{T}^n, \mathbb{R}^n)$ with L^p norms less than $\mathfrak{M}$ and $\nu_a, \nu_b \in H^1(\mathbb{T}^n)$ be probability densities solutions of (4.2) (with respective drift a and b). From the L^∞-bounds on ν_a, $1/\nu_a$, ν_b and $1/\nu_b$ obtained in Step 3, we get that ν_b/ν_a and $(\nu_b/\nu_a)^2$ are in $H^1(\mathbb{T}^n)$. Applying (4.2) for b with $\varphi = \nu_b/\nu_a$ as a test function,

$$
\begin{aligned}
0 \;&=\; \int_{\mathbb{T}^n} b \cdot \nabla \left(\frac{\nu_b}{\nu_a}\right) \nu_b - \nabla \left(\frac{\nu_b}{\nu_a}\right) \cdot \nabla \nu_b \\
&=\; \int_{\mathbb{T}^n} b \cdot \nabla \left(\frac{\nu_b}{\nu_a}\right) \nu_b - \nabla \left(\frac{\nu_b}{\nu_a}\right) \cdot \nabla \left(\frac{\nu_b}{\nu_a}\, \nu_a\right) \\
&=\; \int_{\mathbb{T}^n} b \cdot \nabla \left(\frac{\nu_b}{\nu_a}\right) \nu_b - \left|\nabla \left(\frac{\nu_b}{\nu_a}\right)\right|^2 \nu_a - \frac{1}{2} \nabla \left(\left(\frac{\nu_b}{\nu_a}\right)^2\right) \cdot \nabla \nu_a \\
&=\; \int_{\mathbb{T}^n} b \cdot \nabla \left(\frac{\nu_b}{\nu_a}\right) \nu_b - \left|\nabla \left(\frac{\nu_b}{\nu_a}\right)\right|^2 \nu_a - \frac{1}{2} a \cdot \nabla \left(\left(\frac{\nu_b}{\nu_a}\right)^2\right) \nu_a \\
&=\; \int_{\mathbb{T}^n} b \cdot \nabla \left(\frac{\nu_b}{\nu_a}\right) \nu_b - \left|\nabla \left(\frac{\nu_b}{\nu_a}\right)\right|^2 \nu_a - a \cdot \nabla \left(\frac{\nu_b}{\nu_a}\right) \nu_b
\end{aligned}
$$

where the last term of the above equality stems from (4.2) with drift a and test function $\varphi = (\nu_b/\nu_a)^2/2$. As a consequence, using the Cauchy-Schwarz's inequality and the uniform bounds on ν_a and ν_b, one gets:

$$
\left\|\nabla \left(\frac{\nu_b}{\nu_a}\right)\right\|^2_{L^2(\nu_a)} \;=\; \int_{\mathbb{T}^n} (b - a) \cdot \nabla \left(\frac{\nu_b}{\nu_a}\right) \nu_b \;\leqslant\; C_1^2 \|b - a\|_{L^2(\mathbb{T}^n)} \left\|\nabla \left(\frac{\nu_b}{\nu_a}\right)\right\|_{L^2(\nu_a)} \,,
$$

i.e

$$
\left\|\nabla \left(\frac{\nu_b}{\nu_a}\right)\right\|_{L^2(\nu_a)} \;\leq\; C_1^2 \|b - a\|_{L^2(\mathbb{T}^n)}. \tag{4.7}
$$

Now, since

$$
\|\nu_b - \nu_a\|^2_{L^2(\mathbb{T}^n)} \leq \|\nu_a\|_\infty \int_{\mathbb{T}^n} \left|\frac{\nu_b}{\nu_a} - 1\right|^2 \nu_a \leq C_1 \left\|\frac{\nu_b}{\nu_a} - 1\right\|^2_{L^2(\nu_a)} ,
$$

using the Poincaré inequality (4.6) with $\varphi = \nu_b/\nu_a$ (so that $\int_{\mathbb{T}^n} \varphi\nu_a = 1$) yields:

$$\begin{aligned}\|\nu_b - \nu_a\|^2_{L^2(\mathbb{T}^n)} &\leq \frac{C_1^3}{4\pi^2}\left\|\nabla\left(\frac{\nu_b}{\nu_a}\right)\right\|^2_{L^2(\nu_a)}\\ &\leq \frac{C_1^7}{4\pi^2}\|b-a\|^2_{L^2(\mathbb{T}^n)} \quad \text{(using (4.7))}.\end{aligned}$$

Hence

$$\|\nu_b - \nu_a\|_{L^2(\mathbb{T}^n)} \leqslant \frac{C_1^{\frac{7}{2}}}{2\pi}\|b-a\|_{L^2(\mathbb{T}^n)}.$$

In the particular case $a = b$, we get that there is only one probability density $\nu \in H^1(\mathbb{T}^n)$ that solves (4.2). ☐

4.2. Proof of Theorem 2.7

Proof. Let us recall that one can assume, without loss of generality, that $\beta = 1$ (see the change of variables at the begining of Sect. 5). From now on, let us fix $p = n + 1$. Consider $\mathcal{P}^+$ the set of probability densities on $\mathbb{T}^n$ that are lower bounded by a positive constant.

Given a probability measure $\pi \in \mathcal{P}^+$, let

$$G_\pi(x) = \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x,y)\,\frac{\pi(x,y)}{\pi^\xi(x)}\,\mathrm{d}y, \quad \forall x \in \mathbb{T}^m,$$

where $\pi^\xi(.) = \displaystyle\int_{\mathbb{T}^{n-m}} \pi(.,y)\,\mathrm{d}y$. In the ABF case, set $B_\pi = G_\pi$ and, in the PABF case, consider the Helmholtz projection

$$B_\pi = \nabla H_\pi = \mathsf{P}_{L^2(\lambda)}(G_\pi).$$

In both cases, given Lemma 15.13 from [2], for all $p \geq 2$, there exists a constant $c^* > 0$ such that,

$$\|B_\pi\|_{L^p(\mathbb{T}^m)} \leqslant c^*\|G_\pi\|_{L^p(\mathbb{T}^m)} \leqslant c^*\|\mathcal{F}\|_\infty, \tag{4.8}$$

in other words, for every $\pi \in \mathcal{P}^+$, B_π belongs to the L^p ball $E = \{f \in L^p(\mathbb{T}^m),\ \|f\|_{L^p(\mathbb{T}^m)} \leq c^*\|\mathcal{F}\|_\infty\}$. In return, given $B \in E$, consider the infinitesimal generator $\mathcal{L}_B = (\mathcal{F} + B)\cdot\nabla + \Delta$ and denote by π_B its invariant measure, such as given in Proposition 4.2 (in particular $\pi_B \in \mathcal{P}^+$). Composing these two steps, we obtain an application from E to itself,

$$\begin{aligned} T : E &\longrightarrow E \\ f &\longmapsto B_{\pi_f}\end{aligned}.$$

The link with Theorem 2.7 is that a probability measure π is a stationary state for the non-linear dynamics (1.8) if and only if the associated bias B_π is a fixed point of T. Proving Theorem 2.7 is thus equivalent to prove that T admits a fixed point. This will be established thanks to the Schauders fixed point theorem ([10], Part 9.2.2 Thm. 3). One thus have to prove that T is continuous on $(E, \|\cdot\|_{L^p(\mathbb{T}^m)})$ and that the family $T(E) := \{T(B),\ B \in E\}$ has compact closure in L^p. We have already seen that $T(E) \subset E$, which is a bounded subset of L^p. From the Frchet-Kolmogorov theorem Theorem IV.25 from [5], compactness follows from the following condition :

$$\sup_{z\in\mathbb{R}^n, |z|\leqslant\delta}\ \sup_{f\in T(E)} \|\tau_z f - f\|_{L^p(\mathbb{T}^m)} \underset{\delta\to 0}{\longrightarrow} 0, \tag{4.9}$$

where τ_z is the translation operator, namely $\tau_z f(x) = f(x+z)$ for all $x \in \mathbb{T}^m$.

Let us recall that from Proposition 4.2 there exists a constant $C > 0$ such that for all $B \in E$,

$$\|\pi_B\|_{H^1(\mathbb{T}^n)} + \|\pi_B\|_\infty + \|1/\pi_B\|_\infty \leqslant C \tag{4.10}$$

and for all $B_1, B_2 \in E$,

$$\|\pi_{B_1} - \pi_{B_2}\|_{L^2(\mathbb{T}^n)} \ \leq \ C\|B_1 - B_2\|_{L^2(\mathbb{T}^m)}\,. \tag{4.11}$$

Continuity of T. Let $B_1, B_2 \in E$ and to alleviate notations, denote by $\pi_1 = \pi_{B_1}$, $\pi_2 = \pi_{B_2}$ the associated invariant measures. In both the ABF and PABF cases, using the same arguments as in (4.8) one gets:

$$\|T(B_1) - T(B_2)\|_{L^p(\mathbb{T}^m)} \leqslant c^*\|G_{\pi_1} - G_{\pi_2}\|_{L^p(\mathbb{T}^m)}.$$

Moreover, relying on inequalities (4.10) and (4.11), one has, for all $x \in \mathbb{T}^m$,

$$\begin{aligned}
|G_{\pi_1}(x) - G_{\pi_2}(x)| &\leq \|\mathcal{F}\|_\infty \int_{\mathbb{T}^{n-m}} \left| \frac{\pi_1(x,y)}{\pi_1^\xi(x)} - \frac{\pi_2(x,y)}{\pi_2^\xi(x)} \right| \mathrm{d}y \\
&\leq \|\mathcal{F}\|_\infty \int_{\mathbb{T}^{n-m}} \frac{|\pi_1(x,y) - \pi_2(x,y)|}{\pi_1^\xi(x)} + \frac{\pi_2(x,y)|\pi_1^\xi(x) - \pi_2^\xi(x)|}{\pi_1^\xi(x)\pi_2^\xi(x)} \mathrm{d}y \\
&\leq \|\mathcal{F}\|_\infty C^3 \int_{\mathbb{T}^{n-m}} |\pi_1(x,y) - \pi_2(x,y)| + |\pi_1^\xi(x) - \pi_2^\xi(x)| \mathrm{d}y \\
&\leq 2\|\mathcal{F}\|_\infty C^3 \int_{\mathbb{T}^{n-m}} |\pi_1(x,y) - \pi_2(x,y)| \mathrm{d}y \\
&\leq 2\|\mathcal{F}\|_\infty C^3 \|\pi_1 - \pi_2\|_{L^2(\mathbb{T}^n)} \\
&\leq 2\|\mathcal{F}\|_\infty C^4 \|B_1 - B_2\|_{L^2(\mathbb{T}^m)}\,.
\end{aligned}$$

As a consequence, since $p \geq 2$, by Sobolev embedding,

$$\|T(B_1) - T(B_2)\|_{L^p(\mathbb{T}^m)} \leq c^*\|G_{\pi_1} - G_{\pi_2}\|_{L^p(\mathbb{T}^m)} \leq \ 2c^*\|\mathcal{F}\|_\infty C^4 \|B_1 - B_2\|_{L^p(\mathbb{T}^m)}\,,$$

which proves that T is a Lipschitz function on $(E, \|\cdot\|_{L^p(\mathbb{T}^m)})$.

Remark 4.4. In the particular case where $\|\mathcal{F}\|_\infty$ is small enough so that $2c^*\|\mathcal{F}\|_\infty C^4 < 1$, we directly get that T is a contraction of the L^p-norm, which yields the existence and uniqueness of a fixed-point.

Compactness. Fix $B \in E$ and let $\pi = \pi_B$ to alleviate notations. For $z \in \mathbb{R}^m$, τ_z commutes with the Helmholtz projection so that, using Lemma 15.13 from [2],

$$\|\tau_z \mathsf{P}_{L^2(\lambda)}(G_\pi) - \mathsf{P}_{L^2(\lambda)}(G_\pi)\|_{L^p(\mathbb{T}^m)} \ = \ \|\mathsf{P}_{L^2(\lambda)}(\tau_z G_\pi - G_\pi)\|_{L^p(\mathbb{T}^m)} \ \leq \ c^*\|\tau_z G_\pi - G_\pi\|_{L^p(\mathbb{T}^m)}\,.$$

Hence, in both the ABF and PABF cases, for all $z \in \mathbb{R}^m$,

$$\|\tau_z T(B) - T(B)\|_{L^p(\mathbb{T}^m)} \ \leq \ c^*\|\tau_z G_\pi - G_\pi\|_{L^p(\mathbb{T}^m)}\,.$$

Now, for all $x \in \mathbb{T}^m$ and $z \in \mathbb{R}^m$, using the same argument as in the proof of the continuity of T,

$$\begin{aligned}
|G_\pi(x+z) - G_\pi(x)| &= \left| \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x+z,y) \frac{\pi(x+z,y)}{\pi^\xi(x+z)} \mathrm{d}y - \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x,y) \frac{\pi(x,y)}{\pi^\xi(x)} \mathrm{d}y \right| \\
&\leq \left| \int_{\mathbb{T}^{n-m}} \left(-\mathcal{F}_1(x+z,y) + \mathcal{F}_1(x,y)\right) \frac{\pi(x+z,y)}{\pi^\xi(x+z)} \mathrm{d}y \right| \\
&\quad + \left| \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x,y) \left(\frac{\pi(x+z,y)}{\pi^\xi(x+z)} - \frac{\pi(x,y)}{\pi^\xi(x)} \right) \mathrm{d}y \right| \\
&\leq |z| \|\nabla \mathcal{F}\|_\infty + \|\mathcal{F}\|_\infty \int_{\mathbb{T}^{n-m}} \left| \frac{\pi(x+z,y)}{\pi^\xi(x+z)} - \frac{\pi(x,y)}{\pi^\xi(x)} \right| \mathrm{d}y \\
&\leq |z| \|\nabla \mathcal{F}\|_\infty + 2\|\mathcal{F}\|_\infty C^3 \int_{\mathbb{T}^{n-m}} |\pi(x+z,y) - \pi(x,y)| \mathrm{d}y \\
&\leq |z| \|\nabla \mathcal{F}\|_\infty + 2\|\mathcal{F}\|_\infty C^3 \|\tau_z \pi - \pi\|_{L^2(\mathbb{T}^n)},
\end{aligned}$$

where C stems from (4.10) and (4.11). To bound the last term, write

$$\begin{aligned}\int_{\mathbb{T}^n} |\pi(x+z,y)-\pi(x,y)|^2 \mathrm{d}x\mathrm{d}y &= \int_{\mathbb{T}^n} \left| \int_0^1 z\cdot\nabla_x \pi(x+sz,y)\mathrm{d}s\right|^2 \mathrm{d}x\mathrm{d}y \\ &\leq \int_0^1 \int_{\mathbb{T}^n} |z|^2 |\nabla_x \pi(x+sz,y)|^2 \mathrm{d}x\mathrm{d}y\mathrm{d}s \\ &= |z|^2 \|\nabla_x \pi\|_2^2 \\ &\leq |z|^2 \|\nabla \pi\|_2^2 .\end{aligned}$$

As a conclusion, using (4.10):

$$\|\tau_z G_\pi - G_\pi\|_{L^p(\mathbb{T}^n)} \ \leq \ |z| \left(\|\nabla \mathcal{F}\|_\infty + 2\|\mathcal{F}\|_\infty C^4\right) \,,$$

so that (4.9) holds.

Consequently, there exists an equilibrium measure $\pi_\infty^{\mathcal{F}}$ which is continuous and positive, along with an associated bias $B_\infty^{\mathcal{F}}$. By Proposition 4.2, one has positive upper and lower bounds on $\pi_\infty^{\mathcal{F}}$ and, relying on the Holley-Stroock perturbation result Proposition 5.1.6 from [3], $\pi_\infty^{\mathcal{F}}$ satisfies $LSI(R)$ for some $R>0$ and the conditional densities $y\mapsto \pi_{\infty,x}^{\mathcal{F}}(y) := \pi_\infty^{\mathcal{F}}(x,y)/\pi_\infty^{\mathcal{F},\xi}(x)$ satisfy $LSI(\rho)$ with some $\rho>0$ uniform with respect to $x\in\mathbb{T}^m$. □

4.3. Proof of Proposition 2.10

Let us conclude Section 4 with the proof of Proposition 2.10.

Proof. Let us consider the PABF algorithm. Again, without loss of generality, we suppose that $\beta=1$. Fix $V\in\mathcal{C}^2(\mathbb{T}^n)$, and define

$$\mathfrak{F} = \{(\mathcal{F},\pi_\infty^{\mathcal{F}}) \in \mathcal{C}^1(\mathbb{T}^n,\mathbb{R}^n)\times\mathcal{P}(\mathbb{T}^n) \,|\, \|\mathcal{F}+\nabla V\|_\infty \leq 1,\ \pi_\infty^{\mathcal{F}} \text{ stationary state for } (1.8)\}.$$

In particular, for $(\mathcal{F},\pi_\infty^{\mathcal{F}})\in\mathfrak{F}$, $\pi_\infty^{\mathcal{F}}$ is the invariant measure of the diffusion (4.1) on $\mathbb{T}^n$ with drift $a=\mathcal{F}+\nabla(H_{\mathcal{F}}\circ\xi)$. Moreover,

$$\begin{aligned}\|\mathcal{F}+\nabla V\|_\infty \leq 1 &\Rightarrow \|\mathcal{F}\|_\infty \leq 1+\|\nabla V\|_\infty \\ &\Rightarrow \|G_{\mathcal{F}}\|_\infty \leq 1+\|\nabla V\|_\infty .\end{aligned}$$

By Lemma 15.13 from [2], for all $p\geq 2$, there exists $c^*>0$ such that

$$\|\nabla H_{\mathcal{F}}\|_{L^p(\mathbb{T}^m)} \leq c^* \|G_{\mathcal{F}}\|_{L^p(\mathbb{T}^m)} \leq c^*\left(1+\|\nabla V\|_\infty\right),$$

which yields, by Minkowski's inequality, for all $p\geq 2$

$$\|\mathcal{F}+\nabla(H_{\mathcal{F}}\circ\xi)\|_{L^p(\mathbb{T}^n)} \leq (c^*+1)\left(1+\|\nabla V\|_\infty\right). \tag{4.12}$$

Note on the other hand, that for all $p\geq 2$

$$\|-\nabla V+\nabla(H_{\mathcal{F}}\circ\xi)\|_{L^p(\mathbb{T}^n)} \leq (1+c^*)\|\nabla V\|_\infty + c^*. \tag{4.13}$$

Denote by $\nu_{\mathcal{F}}$ the invariant measure of the diffusion (4.1) on $\mathbb{T}^n$ with drift $a=-\nabla V+\nabla(H_{\mathcal{F}}\circ\xi)$, in other words

$$\nu_{\mathcal{F}}(x,y) \ = \ \frac{1}{Z_{\nu_{\mathcal{F}}}} e^{-V(x,y)+H_{\mathcal{F}}(x)}\,, \qquad Z_{\nu_{\mathcal{F}}} = \int_{\mathbb{T}^n} e^{-V(u,v)+H_{\mathcal{F}}(u)}\mathrm{d}u\mathrm{d}v\,.$$

In the rest of the proof $(\mathcal{F},\pi_\infty^{\mathcal{F}})\in\mathfrak{F}$ is fixed and we are careful to give bounds which are uniform over $\mathfrak{F}$. Besides, to alleviate notations, we simply denote by $\pi=\pi_\infty^{\mathcal{F}}$, $\nu=\nu_{\mathcal{F}}$, $H=H_{\mathcal{F}}$ and $G=G_{\mathcal{F}}$.

Given the bounds (4.12) and (4.13), one can apply Proposition 4.2 with a drift a equal to either $\mathcal{F}+\nabla(H_{\mathcal{F}}\circ\xi)$ or $-\nabla V+\nabla(H_{\mathcal{F}}\circ\xi)$, which are both bounded in $L^p(\mathbb{T}^n)$ for all $p\geq 1$ as shown above. As a consequence, there exists a constant $C>0$ such that for all $(\mathcal{F},\pi)\in\mathfrak{F}$,

$$\|\nu\|_\infty+\|1/\nu\|_\infty+\|\nu\|_{H^1(\mathbb{T}^n)}+\|\pi\|_\infty+\|1/\pi\|_\infty+\|\pi\|_{H^1(\mathbb{T}^n)}\ \leq\ C,$$

and

$$\|\pi-\nu\|_{L^2(\mathbb{T}^n)}\leqslant C\|\mathcal{F}+\nabla V\|_{L^2(\mathbb{T}^n)}\leqslant C\|\mathcal{F}+\nabla V\|_\infty\,. \tag{4.14}$$

Notice that ν has the same conditional laws (given x) than the Gibbs measure μ, so that

$$\nabla A(x)\ =\ \frac{\int_{\mathbb{T}^{n-m}}\nabla_x V(x,y)e^{-V(x,y)}\mathrm{d}y}{\int_{\mathbb{T}^{n-m}}e^{-V(x,y)}\mathrm{d}y}\ =\ \int_{\mathbb{T}^{n-m}}\nabla_x V(x,y)\frac{\nu(x,y)}{\nu^\xi(x)}\mathrm{d}y\,.$$

As a consequence,

$$\begin{aligned}|\nabla A(x)-G(x)|&=\left|\int_{\mathbb{T}^{n-m}}\nabla_x V(x,y)\frac{\nu(x,y)}{\nu^\xi(x)}\mathrm{d}y+\int_{\mathbb{T}^{n-m}}\mathcal{F}_1(x,y)\frac{\pi(x,y)}{\pi^\xi(x)}\mathrm{d}y\right|\\&\leq\|\mathcal{F}+\nabla V\|_\infty+\|\nabla V\|_\infty\int_{\mathbb{T}^{n-m}}\left|\frac{\nu(x,y)}{\nu^\xi(x)}-\frac{\pi(x,y)}{\pi^\xi(x)}\right|\mathrm{d}y.\end{aligned}$$

Using the same argument as in the proof of the continuity of T in Theorem 2.7 and (4.14),

$$\begin{aligned}\int_{\mathbb{T}^{n-m}}\left|\frac{\nu(x,y)}{\nu^\xi(x)}-\frac{\pi(x,y)}{\pi^\xi(x)}\right|\mathrm{d}y&\leq 2C^3\int_{\mathbb{T}^{n-m}}|\nu(x,y)-\pi(x,y)|\,\mathrm{d}y\\&\leq 2C^3\|\nu-\pi\|_{L^2(\mathbb{T}^n)}\\&\leq 2C^4\|\mathcal{F}+\nabla V\|_\infty\,.\end{aligned}$$

We have thus obtained that, uniformly over $\mathfrak{F}$,

$$\|\nabla A-G\|_{L^p(\mathbb{T}^m)}\ \leq\ (1+2\|\nabla V\|_\infty C^4)\|\mathcal{F}+\nabla V\|_\infty\,.$$

Which yields, given ([2], Lem. 15.13):

$$\|\nabla A-\nabla H\|_{L^p(\mathbb{T}^m)}=\|\mathsf{P}_{L^2(\lambda)}\left(\nabla A-G\right)\|_{L^p(\mathbb{T}^m)}\ \leq c^*\|\nabla A-G\|_{L^p(\mathbb{T}^m)}\leq K_V\|\mathcal{F}+\nabla V\|_\infty. \tag{4.15}$$

with $K_V=c^*(1+2\|\nabla V\|_\infty C^4)$. This concludes the proof of the first point of Proposition 2.10. Concerning the second point, first note that

$$\hat{I}_\psi\ =\ \frac{\int_{\mathbb{T}^n}\psi(x,y)e^{-H(x)}\pi(x,y)\mathrm{d}x\mathrm{d}y}{\int_{\mathbb{T}^n}e^{-H(x)}\pi(x,y)\mathrm{d}x\mathrm{d}y}\ =\ \frac{\int_{\mathbb{T}^n}\psi(x,y)e^{-H(x)}\pi(x,y)\mathrm{d}x\mathrm{d}y}{\int_{\mathbb{T}^m}e^{-H(x)}\mathrm{d}x},$$

where we used Proposition 2.4 to see that since π is a stationary state of (1.8), π^ξ is necessarily the uniform measure on $\mathbb{T}^m$. Notice that this expression is unchanged if H is replaced by $H+c$ for some constant $c>0$. As a consequence, for the remainder of the proof and without loss of generality, we suppose that H is normalised so that $\int_{\mathbb{T}^m}e^{-H}=1$.

Using that

$$\int_{\mathbb{T}^n}\psi\mathrm{d}\mu\ =\ \frac{Z_\nu}{Z_\mu}\int_{\mathbb{T}^n}\psi e^{-H\circ\xi}\mathrm{d}\nu\,,$$

we are led to

$$\begin{aligned}\left|\int_{\mathbb{T}^n}\psi \mathrm{d}\mu - \hat{I}_\psi\right| &= \left|\frac{Z_\nu}{Z_\mu}\int_{\mathbb{T}^n}\psi e^{-H\circ\xi}\mathrm{d}\nu - \int_{\mathbb{T}^n}\psi e^{-H\circ\xi}\mathrm{d}\pi\right| \\ &\le \left\|\psi e^{-H\circ\xi}\right\|_\infty \left(\left|\frac{Z_\nu}{Z_\mu}-1\right| + \|\nu-\pi\|_{L^1(\mathbb{T}^n)}\right) \\ &\le \left\|\psi e^{-H\circ\xi}\right\|_\infty \left(\left|\frac{Z_\nu}{Z_\mu}-1\right| + \|\nu-\pi\|_{L^2(\mathbb{T}^n)}\right). \end{aligned} \tag{4.16}$$

Besides,

$$\frac{Z_\nu}{Z_\mu} = \frac{\int_{\mathbb{T}^n} e^{-(V(x,y)-H(x))}\mathrm{d}x\mathrm{d}y}{\int_{\mathbb{T}^n} e^{-V(x,y)}\mathrm{d}x\mathrm{d}y} = \frac{\int_{\mathbb{T}^m} e^{-(A(x)-H(x))}\mathrm{d}x}{\int_{\mathbb{T}^m} e^{-A(x)}\mathrm{d}x}.$$

Again, this expression is unchanged if A is replaced by $A+c$ for some constant $c>0$. In the remaining of the proof and without loss of generality, we suppose that A is normalized so that $\int_{\mathbb{T}^m} A-H = 0$. As a consequence, by the Poincaré-Wirtinger inequality Part 5.8.1 Theorem 1 from [10], there exists a constant $\bar{K}>0$ (that depends only on m and p) such that:

$$\|A-H\|_{L^p(\mathbb{T}^m)} \le \bar{K}\|\nabla A - \nabla H\|_{L^p(\mathbb{T}^m)}.$$

Thus, using (4.15):

$$\begin{aligned}\|A-H\|_{W^{1,p}(\mathbb{T}^m)} &= \|A-H\|_{L^p(\mathbb{T}^m)} + \|\nabla A - \nabla H\|_{L^p(\mathbb{T}^m)} \\ &\le (\bar{K}+1)K_V\|\mathcal{F}+\nabla V\|_\infty.\end{aligned}$$

Now, (4.3) yields the existence of $\mathcal{K}>0$ such that $\|A-H\|_\infty \le \mathcal{K}\|A-H\|_{W^{1,p}(\mathbb{T}^m)}$, hence

$$\|A-H\|_\infty \le \tilde{K}_V\|\mathcal{F}+\nabla V\|_\infty,$$

where $\tilde{K}_V := \mathcal{K}(\bar{K}+1)K_V = \mathcal{K}(\bar{K}+1)c^*(1+2\|\nabla V\|_\infty C^4)$. Then, using that $|e^a-1| \leqslant |a|e^{|a|}$ for all $a\in\mathbb{R}$, for all $x\in\mathbb{T}^m$,

$$|e^{-A(x)+H(x)}-1| \leqslant \tilde{K}_V\|\mathcal{F}+\nabla V\|_\infty e^{\tilde{K}_V\|\mathcal{F}+\nabla V\|_\infty},$$

so that, using the fact that $\displaystyle\int_{\mathbb{T}^m} e^{-H} = 1$,

$$\begin{aligned}\left|\frac{Z_\nu}{Z_\mu}-1\right| = \left|\frac{\int_{\mathbb{T}^m} e^{-A+H}}{\int_{\mathbb{T}^m} e^{-A}}-1\right| &\leqslant \frac{\left|\int_{\mathbb{T}^m} e^{-A+H}-1\right| + \left|\int_{\mathbb{T}^m} e^{-A}-1\right|}{\int_{\mathbb{T}^m} e^{-A}} \\ &\leqslant \frac{\int_{\mathbb{T}^m}\left|e^{-A+H}-1\right| + \int_{\mathbb{T}^m} e^{-H}|e^{-A+H}-1|}{\int_{\mathbb{T}^m} e^{-H-|H-A|}} \\ &\leqslant 2\tilde{K}_V\|\mathcal{F}+\nabla V\|_\infty e^{2\tilde{K}_V\|\mathcal{F}+\nabla V\|_\infty} \\ &\le 2\tilde{K}_V\|\mathcal{F}+\nabla V\|_\infty e^{2\tilde{K}_V}.\end{aligned}$$

Combining this with (4.14) in (4.16), we have obtained that

$$\begin{aligned}\left|\int_{\mathbb{T}^n}\psi\mathrm{d}\mu - \hat{I}_\psi\right| &\leqslant \frac{\left\|\psi e^{-H\circ\xi}\right\|_\infty}{\int_{\mathbb{T}^m} e^{-H}}\left(2\tilde{K}_V e^{2\tilde{K}_V}+C\right)\|\mathcal{F}+\nabla V\|_\infty \\ &\leqslant \frac{\left\|\psi e^{-A\circ\xi}\right\|_\infty}{\int_{\mathbb{T}^m} e^{-A}} e^{2\tilde{K}_V\|\mathcal{F}+\nabla V\|_\infty}\left(2\tilde{K}_V e^{2\tilde{K}_V}+C\right)\|\mathcal{F}+\nabla V\|_\infty \\ &\leqslant \frac{\|\psi\|_\infty\left\|e^{-A}\right\|_\infty}{\int_{\mathbb{T}^m} e^{-A}} e^{2\tilde{K}_V}\left(2K_V e^{2\tilde{K}_V}+C\right)\|\mathcal{F}+\nabla V\|_\infty,\end{aligned}$$

which yields the conclusion.

□

5. Long-time convergence

In Section 5.1 one can find the proof of intermediate results that will prove useful for the proofs of Theorem 2.11, Theorem 2.13, and Corollary 2.15. Said proofs can respectively be found in Sections 5.2, 5.3 and 5.4. In all this section, to alleviate notations, we will denote by π_∞ (dropping the $\mathcal{F}$ superscript) a stationary measure given by Theorem 2.7. First and foremost, let us introduce the concept of total entropy and its macroscopic-microscopic decomposition. We define the *total entropy* as:

$$E(t) = \mathcal{H}(\pi_t|\pi_\infty).$$

In the same manner, the entropy between the marginals in $x \in \mathbb{T}^m$ (called *macroscopic entropy* henceforth) is given by:

$$E_M(t) = \mathcal{H}(\pi_t^\xi|\pi_\infty^\xi).$$

Note that accordingly, one can define the *macroscopic Fisher information*:

$$I_M(t) = I(\pi_t^\xi|\pi_\infty^\xi).$$

The entropy between the conditional measures at a given $x \in \mathbb{T}^m$ (called *local entropy* in the following) is:

$$e_m(t,x) = \mathcal{H}(\pi_{t,x}|\pi_{\infty,x}),$$

where $\pi_{t,x}(.) = \dfrac{\pi_t(x,.)}{\pi_t^\xi(x)}$ and $\pi_{\infty,x}(.) = \dfrac{\pi_\infty(x,.)}{\pi_\infty^\xi(x)}$. Now, let us introduce the so-called *microscopic entropy*:

$$E_m(t) = \int_{\mathbb{T}^m} e_m(t,x)\pi_t^\xi(x)\mathrm{d}x.$$

One has for all $t \geq 0$, $E(t) = E_m(t) + E_M(t)$ (see [19], Lem. 1).

Note that we have the following bound on the microscopic entropy:

$$\begin{aligned} E_m(t) = \int_{\mathbb{T}^m} e_m(t,x)\,\pi_t^\xi(x)\,\mathrm{d}x &= \int_{\mathbb{T}^m} \mathcal{H}(\pi_{t,x}|\pi_{\infty,x})\,\pi_t^\xi(x)\,\mathrm{d}x \\ &\leq \frac{1}{2\rho}\int_{\mathbb{T}^m} I(\pi_{t,x}|\pi_{\infty,x})\,\pi_t^\xi(x)\,\mathrm{d}x \quad \text{(using Thm. 2.7 (ii)).} \end{aligned}$$

Since $\nabla_y \ln\left(\dfrac{\pi_{t,x}}{\pi_{\infty,x}}\right) = \nabla_y \ln\left(\dfrac{\pi_t}{\pi_\infty}\right)$, this leads to

$$E_m(t) \leq \frac{1}{2\rho}\int_{\mathbb{T}^n} \left|\nabla_y \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t. \tag{5.1}$$

5.1. Intermediate results

The proofs of both Theorems 2.11 and 2.13 will rely on the following intermediate results. Assumptions 2.3 and 2.6 are enforced. Here, $\mathcal{F}$ can be either conservative ($\mathcal{F} = -\nabla V$) or not, and $(\pi_\infty, B_\infty, G_\infty)$ denotes a stationary state of (1.8), with R, ρ the corresponding constants given by Theorem 2.7.

Lemma 5.1 (Bound on $G_t(x) - G_\infty(x)$)**.** *For all $t \geq 0$ and $x \in \mathbb{T}^m$:*

$$|G_t(x) - G_\infty(x)| \leq M\sqrt{\frac{2}{\rho}e_m(t,x)}.$$

Proof. Note that, given Theorem 2.7 (ii), since $\pi_{\infty,x}$ satisfies a log-Sobolev inequality with constant $\rho > 0$, it also satisfies a Talagrand inequality with constant ρ. Now, let $x \in \mathbb{T}^m$ and $\nu_x \in \Pi(\pi_{t,x}, \pi_{\infty,x})$ be a coupling measure. Then, one has:

$$\begin{aligned} G_t(x) - G_\infty(x) &= \int_{\mathbb{T}^{n-m}} (-\mathcal{F}_1(x,y) + \mathcal{F}_1(x,y'))\, \nu_x(\mathrm{d}y, \mathrm{d}y') \\ &\leq M \int_{\mathbb{T}^{n-m}} |y - y'|\, \nu_x(\mathrm{d}y, \mathrm{d}y') \quad \text{(by Assumption 2.6)} \\ &\leq M \left(\int_{\mathbb{T}^{n-m}} |y - y'|^2\, \nu_x(\mathrm{d}y, \mathrm{d}y') \right)^{\frac{1}{2}}. \end{aligned}$$

Taking the infinimum over $\Pi(\pi_{t,x}, \pi_{\infty,x})$ yields:

$$\begin{aligned} G_t(x) - G_\infty(x) &\leq M W_2(\pi_{t,x}, \pi_{\infty,x}) \\ &\leq M \sqrt{\frac{2}{\rho} \mathcal{H}(\pi_{t,x} | \pi_{\infty,x})} \quad \text{(by the Talagrand inequality (2.2))}. \end{aligned}$$

This yields the conclusion, since $\mathcal{H}(\pi_{t,x}|\pi_{\infty,x}) = e_m(t,x)$. □

Lemma 5.2 (Total entropy)**.** *One has,*

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\int_{\mathbb{T}^n} |\nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) |^2\, \pi_t + \int_{\mathbb{T}^n} (B_t - B_\infty)(x) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t.$$

Proof. If $\mathcal{L}_t$ denotes the infinitesimal generator of (1.7) and $\mathcal{L}'_t$ its formal adjoint in $L^2(\mathbb{T}^n)$ then the Fokker-Planck equation (1.8) can be rewritten as follows:

$$\partial_t \pi_t = \mathcal{L}'_t(\pi_t).$$

Denote by $\mathcal{L}_\infty = \mathcal{F} \cdot \nabla + B_\infty \cdot \nabla_x + \Delta$ the infinitesimal generator associated to the stationary state (π_∞, B_∞). One has:

$$\begin{aligned} \frac{\mathrm{d}E}{\mathrm{d}t} &= \int_{\mathbb{T}^n} \partial_t \pi_t + \int_{\mathbb{T}^n} \partial_t \pi_t \ln \left(\frac{\pi_t}{\pi_\infty} \right) \\ &= \int_{\mathbb{T}^n} \mathcal{L}'_t(\pi_t) \ln \left(\frac{\pi_t}{\pi_\infty} \right) \qquad \left(\text{since } \int_{\mathbb{T}^n} \partial_t \pi_t = 0\right) \\ &= \int_{\mathbb{T}^n} \mathcal{L}_t \left(\ln \left(\frac{\pi_t}{\pi_\infty} \right) \right) \pi_t \\ &= \int_{\mathbb{T}^n} (\mathcal{L}_\infty + \mathcal{L}_t - \mathcal{L}_\infty) \left(\ln \left(\frac{\pi_t}{\pi_\infty} \right) \right) \pi_t \\ &= \int_{\mathbb{T}^n} \mathcal{L}_\infty \left(\ln \left(\frac{\pi_t}{\pi_\infty} \right) \right) \pi_t + \int_{\mathbb{T}^n} (\mathcal{L}_t - \mathcal{L}_\infty) \left(\ln \left(\frac{\pi_t}{\pi_\infty} \right) \right) \pi_t. \end{aligned}$$

Since $\mathcal{L}_\infty$ is the infinitesimal generator of a diffusion, it follows that, for any given functions a and f:

$$\mathcal{L}_\infty(a(f)) = a'(f) \mathcal{L}_\infty(f) + a"(f) |\nabla f|^2,$$

as mentioned in ([23], Part 2.3). Applying this with $a(.) = \ln(.)$ and $f = \dfrac{\pi_t}{\pi_\infty}$ we respectively obtain:

$$\begin{aligned}
\int_{\mathbb{T}^n} \mathcal{L}_\infty \left(\ln\left(\frac{\pi_t}{\pi_\infty}\right)\right)\pi_t &= \int_{\mathbb{T}^n} \left(\frac{\pi_\infty}{\pi_t}\cdot \mathcal{L}_\infty\left(\frac{\pi_t}{\pi_\infty}\right) - \left(\frac{\pi_\infty}{\pi_t}\right)^2 \cdot \left|\nabla \frac{\pi_t}{\pi_\infty}\right|^2\right)\pi_t \\
&= \int_{\mathbb{T}^n} \mathcal{L}_\infty\left(\frac{\pi_t}{\pi_\infty}\right)\pi_\infty - \int_{\mathbb{T}^n}\left|\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t \\
&= -\int_{\mathbb{T}^n}\left|\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t \qquad (\text{since } \pi_\infty \text{ is invariant for } \mathcal{L}_\infty)
\end{aligned}$$

and

$$\int_{\mathbb{T}^n} (\mathcal{L}_t - \mathcal{L}_\infty)\left(\ln\left(\frac{\pi_t}{\pi_\infty}\right)\right)\pi_t = \int_{\mathbb{T}^n} (B_t - B_\infty)\cdot \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t,$$

which concludes the proof. □

5.2. Proof of Theorem 2.11

Let us prove the convergence of the ABF and PABF algorithms in the conservative case, namely when $\mathcal{F} = -\nabla V$. In that case $\pi_\infty = \mu_A$ is invariant by (1.8) (recall μ_A is given by (1.5)), with a corresponding $G_\infty = \nabla A$, so that $B_\infty = \nabla A$ in both the ABF and PABF case.

Lemma 5.3. *In the conservative case ($\mathcal{F} = -\nabla V$ and $\pi_\infty = \mu_A$), for all $t \geq 0$ and $x \in \mathbb{T}^m$:*

$$G_t(x) - \nabla A(x) = \int_{\mathbb{T}^{n-m}} \nabla_x \ln\left(\frac{\pi_t(x,y)}{\pi_\infty(x,y)}\right)\frac{\pi_t(x,y)}{\pi_t^\xi(x)}\,\mathrm{d}y - \nabla_x \ln\left(\frac{\pi_t^\xi(x)}{\pi_\infty^\xi(x)}\right).$$

Proof. Knowing that $\pi_\infty^\xi = 1$, one has, for a fixed x in $\mathbb{T}^m$:

$$\begin{aligned}
&\int_{\mathbb{T}^{n-m}} \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\frac{\pi_t}{\pi_t^\xi}\,\mathrm{d}y - \nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right) \\
&= \int_{\mathbb{T}^{n-m}} \frac{\nabla_x \pi_t}{\pi_t}\cdot\frac{\pi_t}{\pi_t^\xi}\,\mathrm{d}y - \int_{\mathbb{T}^{n-m}} \frac{\nabla_x \pi_\infty}{\pi_\infty}\cdot\frac{\pi_t}{\pi_t^\xi}\,\mathrm{d}y - \frac{\nabla_x \pi_t^\xi}{\pi_t^\xi} + \partial_x \ln(1) \\
&= \frac{\nabla_x \pi_t^\xi}{\pi_t^\xi} - \int_{\mathbb{T}^{n-m}} \frac{\nabla_x \pi_\infty}{\pi_\infty}\cdot\frac{\pi_t}{\pi_t^\xi}\,\mathrm{d}y - \frac{\nabla_x \pi_t^\xi}{\pi_t^\xi} \\
&= -\int_{\mathbb{T}^{n-m}} \nabla_x\left(-V(x,y) + A(x)\right)\cdot\frac{\pi_t}{\pi_t^\xi}\,\mathrm{d}y \\
&= \int_{\mathbb{T}^{n-m}} \nabla_x V(x,y)\cdot\frac{\pi_t}{\pi_t^\xi}\,\mathrm{d}y - \int_{\mathbb{T}^{n-m}} \nabla A(x)\cdot\frac{\pi_t}{\pi_t^\xi}\,\mathrm{d}y \\
&= G_t(x) - \nabla A(x).
\end{aligned}$$

□

In the following proofs, an integral over $\mathbb{T}^n$ is with respect to $(x,y) \in \mathbb{T}^m \times \mathbb{T}^{n-m}$, an integral over $\mathbb{T}^m$ is with respect to $x \in \mathbb{T}^m$, and an integral over $\mathbb{T}^{n-m}$ is with respect to $y \in \mathbb{T}^{n-m}$.

Proof of Theorem 2.11.

Step 1: Since for all $t \geq 0$, $E(t) = E_m(t) + E_M(t)$, using (3.1) and Lemma 5.2, one has:

$$\begin{aligned}\frac{\mathrm{d}E_m}{\mathrm{d}t} &= -\int_{\mathbb{T}^n} \left|\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t + \int_{\mathbb{T}^n} (B_t - \nabla A)\cdot \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t + \int_{\mathbb{T}^m} \left|\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\right|^2 \pi_t^\xi \\ &= -\int_{\mathbb{T}^n} \left|\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t + \int_{\mathbb{T}^n} (G_t - \nabla A)\cdot \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t + \int_{\mathbb{T}^m} \left|\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\right|^2 \pi_t^\xi + J_t\end{aligned}$$

where

$$J_t := \int_{\mathbb{T}^n} (B_t - G_t)\cdot \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t.$$

Now, using Lemma 5.3, one gets:

$$\begin{aligned}\frac{\mathrm{d}E_m}{\mathrm{d}t} = &-\int_{\mathbb{T}^n} \left|\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t + \int_{\mathbb{T}^n}\int_{\mathbb{T}^{n-m}} \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\frac{\pi_t}{\pi_t^\xi}\,\mathrm{d}y \cdot \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t \\ &- \int_{\mathbb{T}^n} \nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\cdot \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t + \int_{\mathbb{T}^m} \left|\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\right|^2 \pi_t^\xi + J_t.\end{aligned}$$

On the one hand, using Cauchy-Schwarz's inequality, the first terms in the right-hand side can be bounded as follows:

$$\begin{aligned}\int_{\mathbb{T}^n}\int_{\mathbb{T}^{n-m}} \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\frac{\pi_t}{\pi_t^\xi}\,\mathrm{d}y \cdot \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t &= \int_{\mathbb{T}^m} \left|\int_{\mathbb{T}^{n-m}} \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t\right|^2 \frac{1}{\pi_t^\xi} \\ &\leq \int_{\mathbb{T}^n} \left|\nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t.\end{aligned}$$

On the other hand, factorising the two next terms in the right-hand side, and using again Lemma 5.3 gives:

$$\begin{aligned}&-\int_{\mathbb{T}^n} \nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\cdot \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t + \int_{\mathbb{T}^m} \left|\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\right|^2 \pi_t^\xi \\ &= \int_{\mathbb{T}^m} \nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\cdot\left(\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right) - \int_{\mathbb{T}^{n-m}} \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\frac{\pi_t}{\pi_t^\xi}\right)\pi_t^\xi \\ &= \int_{\mathbb{T}^m} \nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)(\nabla A - G_t)\pi_t^\xi.\end{aligned}$$

Using once again the Cauchy-Schwarz's inequality, one gets:

$$\frac{\mathrm{d}E_m}{\mathrm{d}t} \leq -\int_{\mathbb{T}^n} \left|\nabla_y \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t + \left(\int_{\mathbb{T}^m} |\nabla A - G_t|^2\,\pi_t^\xi\right)^{\frac{1}{2}} \left(\int_{\mathbb{T}^m} \left|\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\right|^2 \pi_t^\xi\right)^{\frac{1}{2}} + J_t.$$

Eventually, recalling that $\int_{\mathbb{T}^m} |\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)|^2 \pi_t^\xi = I_M(t)$ and relying on (5.1) and Lemma 5.1, one has, for all $\varepsilon > 0$ and all $t \geq 0$:

$$\begin{aligned}\frac{\mathrm{d}E_m}{\mathrm{d}t} &\leq -2\rho E_m(t) + M\sqrt{\frac{2}{\rho}}\sqrt{E_m(t)}\sqrt{I_M(t)} + J_t \\ &\leq -2\rho E_m(t) + 2\sqrt{\varepsilon\rho E_m(t)}\sqrt{\frac{M^2}{2\rho^2\varepsilon} I_M(t)} + J_t \\ &\leq -(2-\varepsilon)\rho E_m(t) + \frac{M^2}{2\rho^2\varepsilon} I_M(t) + J_t. \end{aligned} \tag{5.2}$$

Step 2: In order to set the idea of the proof, let us first treat the case of the ABF algorithm, where one simply has $J_t = 0$ for all $t \geq 0$. Inequality (5.2) yields, for all $\varepsilon > 0$ and all $t \geq 0$:

$$\frac{\mathrm{d}E_m}{\mathrm{d}t} \leq -(2-\varepsilon)\rho E_m(t) + \frac{M^2}{2\rho^2\varepsilon} I_M(t),$$

and using Gronwall's lemma, one has, for all $\varepsilon > 0$ for all $t \geq 0$, :

$$E_m(t) \leq E_m(0)e^{-(2-\varepsilon)\rho t} + \frac{M^2}{2\rho^2\varepsilon}\int_0^t I_M(s)e^{-(2-\varepsilon)\rho(t-s)}\,\mathrm{d}s.$$

Remark 5.4. Note that in the ABF case Lemma 12 from [19] or the PABF case with a Helmholtz projection done with respect to the marginal density π_t^ξ Corollary 1 from [1], one has the exponential convergence towards zero of the macroscopic Fisher information $I_M(t)$. This is not the case when one considers the classical Helmholtz projection with respect to the Lebesgue measure: indeed, the density π_t^ξ does not satisfy the heat equation anymore, but an elliptic equation (2.3) with a null-divergence drift. Having no additional information about the regularity of the drift, one cannot prove the convergence of $I_M(t)$ towards zero in the long-time limit as done in [1, 19].

By Proposition 2.4, for all $t \geq 0$, $E_M(t) \leq E_M(0)e^{-8\pi^2 t}$. Since $I_M(t) = -E'_M(t)$, one gets:

$$0 \leq F(t) := \int_t^\infty I_M(s)\,\mathrm{d}s \leq E_M(t) \leq E_M(0)e^{-8\pi^2 t}, \qquad \forall t \geq 0. \tag{5.3}$$

Consequently, relying on (5.3) one has, for all $\varepsilon > 0$, for all $t \geq 0$:

$$\begin{aligned}\int_0^t I_M(s)e^{-(2-\varepsilon)\rho(t-s)}\,\mathrm{d}s &= -e^{-(2-\varepsilon)\rho t}\int_0^t F'(s)e^{(2-\varepsilon)\rho s}\,\mathrm{d}s \\ &= e^{-(2-\varepsilon)\rho t}\left(\int_0^t F(s)(2-\varepsilon)\rho e^{(2-\varepsilon)\rho s}\,\mathrm{d}s - \left[F(s)e^{(2-\varepsilon)\rho s}\right]_0^t\right) \\ &\leq e^{-(2-\varepsilon)\rho t}\left((2-\varepsilon)\rho E_M(0)\int_0^t e^{-(8\pi^2-(2-\varepsilon)\rho)s}\,\mathrm{d}s - F(t)e^{(2-\varepsilon)\rho t} + F(0)\right) \\ &\leq e^{-(2-\varepsilon)\rho t}\left((2-\varepsilon)\rho E_M(0)\int_0^t e^{-(8\pi^2-(2-\varepsilon)\rho)s}\,\mathrm{d}s - F(t)e^{(2-\varepsilon)\rho t} + E_M(0)\right) \\ &\leq E_M(0)e^{-(2-\varepsilon)\rho t}\left((2-\varepsilon)\rho\int_0^t e^{-(8\pi^2-(2-\varepsilon)\rho)s}\,\mathrm{d}s + 1\right)\end{aligned}$$

We distinguish between two case:

(i) If $8\pi^2 = (2-\varepsilon)\rho$, one gets:

$$\int_0^t I_M(s)e^{-(2-\varepsilon)\rho(t-s)}\,\mathrm{d}s \le E_M(0)e^{-8\pi^2 t}\left(8\pi^2\, t+1\right).$$

Since for all $\delta > 0$ and all $t \ge 0$, one has $t \le \frac{e^{-1}}{\delta}e^{\delta t}$, choosing $\delta = \varepsilon$ yields, for all $t \ge 0$:

$$\int_0^t I_M(s)e^{-(2-\varepsilon)\rho(t-s)}\,\mathrm{d}s \le E_M(0)\left(\frac{8\pi^2}{e\,\varepsilon}\vee 1\right)e^{-(8\pi^2-\varepsilon)t}.$$

(ii) If $8\pi^2 \ne (2-\varepsilon)\rho$, one gets, in all cases ($8\pi^2 > (2-\varepsilon)\rho$ or $8\pi^2 < (2-\varepsilon)\rho$):

$$\int_0^t I_M(s)e^{-(2-\varepsilon)\rho(t-s)}\,\mathrm{d}s \le E_M(0)\left(\frac{(2-\varepsilon)\rho}{|8\pi^2-(2-\varepsilon)\rho|}\vee 1\right)e^{-(8\pi^2\wedge(2-\varepsilon)\rho)t}.$$

Which yields,

$$\begin{aligned}E_m(t) &\le E_m(0)e^{-(2-\varepsilon)\rho t} + \frac{M^2}{2\rho^2\varepsilon}\int_0^t I_M(s)e^{-(2-\varepsilon)\rho(t-s)}\,\mathrm{d}s\\ &\le \left(E_m(0)\vee\frac{M^2}{2\rho^2\varepsilon}E_M(0)\left(\frac{8\pi^2}{e\varepsilon}\vee\frac{(2-\varepsilon)\rho}{|8\pi^2-(2-\varepsilon)\rho|}\vee 1\right)\right)e^{-((8\pi^2-\varepsilon)\wedge(2-\varepsilon)\rho)t}.\end{aligned}$$

Conclusion: for the ABF algorithm, we have obtained that for all $\varepsilon > 0$, there exists $\mathcal{K} = \mathcal{K}(\varepsilon) > 0$ such that for all $t \ge 0$,

$$E_m(t) \le \mathcal{K}e^{-\left((8\pi^2\wedge 2\rho)-\varepsilon\right)t},$$

where $\mathcal{K} = \left(E_m(0)\vee\dfrac{M^2}{2\rho\varepsilon}E_M(0)\left(\dfrac{8\pi^2\rho}{e\varepsilon}\vee\dfrac{(2\rho-\varepsilon)}{|8\pi^2-(2\rho-\varepsilon)|}\vee 1\right)\right)$.

Step 3: Let us now concentrate on the PABF case, and let us prove an upper bound on J_t. For $t \ge 0$, recall the notation $\nabla H_t := \mathsf{P}_{L^2(\lambda)}(G_t)$, so that $B_t = \nabla H_t$. Similarly, let us introduce, for all $t \ge 0$,

$$\nabla\tilde{H}_t := \mathsf{P}_{L^2(\pi_t^\xi)}(G_t).$$

Recall that $\mathsf{P}_{L^2(\nu)}(f)$ stands for the Helmholtz projection of a vector field f with respect to the measure ν. In the conservative case one has $\pi_\infty \propto e^{-V+A}$, so that $G_\infty = \nabla A$. Since G_∞ is a gradient, one has:

$$\nabla H_\infty = \mathsf{P}_{L^2(\lambda)}(G_\infty) = \nabla A = \mathsf{P}_{L^2(\pi_\infty^\xi)}(G_\infty) = \nabla\tilde{H}_\infty.$$

On the contrary, there is no reason for ∇H_t and $\nabla\tilde{H}_t$ to be equal at a fixed time $t > 0$. Let us decompose

$$J_t = \int_{\mathbb{T}^n}(\nabla H_t - \nabla\tilde{H}_t)\cdot\nabla_x\ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t + \int_{\mathbb{T}^n}(\nabla\tilde{H}_t - G_t)\cdot\nabla_x\ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t.$$

As proven in Lemma 6 from [1], relying on the fact that since $\nabla_x\ln(\pi_\infty) = -\nabla(V-A)$, $\pi_\infty^\xi \equiv 1$ and $\nabla\tilde{H}_t = \mathsf{P}_{L^2(\pi_t^\xi)}(G_t)$, one can show that the last right-hand term is negative. One consequently has:

$$\begin{aligned}\int_{\mathbb{T}^n}(\nabla H_t - \nabla\tilde{H}_t)\nabla_x\ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t &= \int_{\mathbb{T}^n}(\nabla H_t - \nabla\tilde{H}_t)\cdot\nabla_x\ln(\pi_t)\,\pi_t - \int_{\mathbb{T}^n}(\nabla H_t - \nabla\tilde{H}_t)\cdot\nabla_x\ln(\pi_\infty)\,\pi_t\\ &= \int_{\mathbb{T}^m}(\nabla H_t - \nabla\tilde{H}_t)\cdot\nabla_x\pi_t^\xi - \int_{\mathbb{T}^n}(\nabla H_t - \nabla\tilde{H}_t)\cdot\nabla_x\ln(\pi_\infty)\,\pi_t\\ &= \int_{\mathbb{T}^m}(\nabla H_t - \nabla\tilde{H}_t)\cdot\nabla_x\ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\pi_t^\xi - \int_{\mathbb{T}^n}(\nabla H_t - \nabla\tilde{H}_t)\cdot\nabla_x\ln(\pi_\infty)\,\pi_t.\end{aligned}$$

Hence, in the PABF case,

$$
\begin{aligned}
J_t &\leq \int_{\mathbb{T}^m} (\nabla H_t - \nabla \tilde{H}_t)\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\pi_t^\xi - \int_{\mathbb{T}^n} (\nabla H_t - \nabla \tilde{H}_t)\nabla_x \ln(\pi_\infty)\,\pi_t \\
&= \int_{\mathbb{T}^m} (\nabla H_t - \nabla \tilde{H}_t)\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\pi_t^\xi - \int_{\mathbb{T}^m} (\nabla H_t - \nabla \tilde{H}_t)(\nabla A - G_t)\,\pi_t^\xi \\
&\leq \left(\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2\,\pi_t^\xi\right)^{\frac{1}{2}} \left(\left(\int_{\mathbb{T}^m} \left|\nabla_x \ln\left(\frac{\pi_t^\xi}{\pi_\infty^\xi}\right)\right|^2 \pi_t^\xi\right)^{\frac{1}{2}} + \left(\int_{\mathbb{T}^m} |\nabla A - G_t|^2\,\pi_t^\xi\right)^{\frac{1}{2}}\right) \\
&\leq \left(\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2\,\pi_t^\xi\right)^{\frac{1}{2}} \left(\sqrt{I_M(t)} + M\sqrt{\frac{2}{\rho}}\sqrt{E_m(t)}\right).
\end{aligned}
\tag{5.4}
$$

Step 4: We will now consider times such that $t \geq 1$. Since $\nabla \tilde{H}_t = \mathsf{P}_{L^2(\pi_t^\xi)}(G_t)$, one has:

$$
\int_{\mathbb{T}^m} |\nabla H_t - G_t|^2 \pi_t^\xi = \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi + \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - \nabla H_t|^2 \pi_t^\xi,
$$

which yields:

$$
\begin{aligned}
\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2\,\pi_t^\xi &= \int_{\mathbb{T}^m} |\nabla H_t - G_t|^2 \pi_t^\xi - \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi \\
&\leq \|\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |\nabla H_t - G_t|^2 - \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi \\
&\leq \|\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 - \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi \\
&\leq \|\pi_t^\xi\|_\infty \left(\left\|1/\pi_t^\xi\right\|_\infty - 1\right) \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi \\
&\leq \|\pi_t^\xi\|_\infty \left(\left\|1/\pi_t^\xi\right\|_\infty - 1\right) \int_{\mathbb{T}^m} |G_t|^2 \pi_t^\xi \\
&\leq \|\pi_t^\xi\|_\infty \left(\left\|1/\pi_t^\xi\right\|_\infty - 1\right) M^2,
\end{aligned}
$$

where we used that, under Assumption 2.6, $\|G_t\|_\infty \leq \|\nabla_x V\|_\infty \leq M$. Now, from Proposition 2.5, there exists $C \geq 0$ such that, for all $t \geq 1$:

$$
\|\pi_t^\xi\|_\infty \leq 1 + Ce^{-4\pi^2 t} \qquad \text{and} \qquad \|1/\pi_t^\xi\|_\infty \leq 1 + Ce^{-4\pi^2 t}.
$$

This yields the existence of a constant $\tilde{C} > 0$ such that, for all $t \geq 1$:

$$
\left(\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2\,\pi_t^\xi\right)^{\frac{1}{2}} \leq \tilde{C} e^{-2\pi^2 t},
$$

and, for all $\varepsilon > 0$, for all $t \geq 1$:

$$\begin{aligned}
J_t &\leq \tilde{C} e^{-2\pi^2 t} \left(\sqrt{I_M}(t) + M \sqrt{\frac{2}{\rho}} \sqrt{E_m(t)} \right) \\
&\leq \tilde{C} e^{-2\pi^2 t} \left(\sqrt{I_M}(t) + 2 \sqrt{\frac{M^2}{2\rho^2 \varepsilon}} \sqrt{\varepsilon \rho E_m(t)} \right) \\
&\leq \varepsilon \rho E_m(t) + I_M(t) + \left(\frac{\tilde{C}^2}{4} + \frac{M^2 \tilde{C}^2}{2\rho^2 \varepsilon} \right) e^{-4\pi^2 t}.
\end{aligned}$$

Hence one gets:

$$\frac{\mathrm{d}E_m}{\mathrm{d}t} \leq -(2 - 2\varepsilon)\rho E_m(t) + K_1 I_M(t) + K_2 e^{-4\pi^2 t}, \quad \forall t \geq 1,$$

with

$$K_1 = K_1(\varepsilon) = 1 + \frac{M^2}{2\rho^2 \varepsilon}, \qquad K_2 = K_2(\varepsilon) = \frac{\tilde{C}^2}{4} + \frac{M^2 \tilde{C}^2}{2\rho^2 \varepsilon}.$$

From now on, let us fix $\varepsilon \in (0,1)$ and denote by $r_\varepsilon := 2(1-\varepsilon)$. Using Gronwall's lemma yields, for all $t \geq 1$:

$$E_m(t) \leq E_m(1) e^{r_\varepsilon \rho} e^{-r_\varepsilon \rho t} + \int_1^t K_1 I_M(s) e^{-r_\varepsilon \rho (t-s)} + K_2 e^{-4\pi^2 s - r_\varepsilon \rho(t-s)} \, \mathrm{d}s.$$

• Let us first consider, for all $t \geq 1$, $I_1 := K_1 \int_1^t I_M(s) e^{-r_\varepsilon \rho (t-s)}$. As done in Step 2, relying on (5.3), one has, for all $t \geq 1$:

$$\begin{aligned}
I_1 &= -K_1 e^{-r_\varepsilon \rho t} \int_1^t F'(s) e^{r_\varepsilon \rho s} \, \mathrm{d}s \\
&= K_1 e^{-r_\varepsilon \rho t} \left(r_\varepsilon \rho \int_1^t F(s) e^{r_\varepsilon s} \, \mathrm{d}s - F(t) e^{r_\varepsilon \rho t} + F(1) e^{r_\varepsilon \rho} \right) \\
&\leq K_1 E_M(0) e^{-r_\varepsilon \rho t} \left(r_\varepsilon \rho \int_1^t e^{-(8\pi^2 - r_\varepsilon)s} \, \mathrm{d}s + e^{-(8\pi^2 - r_\varepsilon \rho)} \right).
\end{aligned}$$

We distinguish between two cases:

(i) If $8\pi^2 = r_\varepsilon \rho$, one gets, for all $t \geq 1$:

$$I_1 \leq K_1 E_M(0) e^{-8\pi^2 t} \left(8\pi^2 (t-1) + 1 \right)$$

and, since $(t-1) \leq \frac{e^{-1-\delta}}{\delta} e^{\delta t}$ for all $\delta > 0$, considering $\delta = \varepsilon$, one gets that, for all $t \geq 1$:

$$\begin{aligned}
I_1 &\leq K_1 E_M(0) e^{-8\pi^2 t} \left(8\pi^2 \frac{e^{-1-\varepsilon}}{\varepsilon} e^{\varepsilon t} + 1 \right) \\
&\leq K_1 E_M(0) \left(8\pi^2 \frac{e^{-1-\varepsilon}}{\varepsilon} \vee 1 \right) e^{-(8\pi^2 - \varepsilon)t}.
\end{aligned}$$

(ii) If $8\pi^2 \neq r_\varepsilon \rho$, one gets, for all $t \geq 1$:

$$I_1 \leq K_1 E_M(0) \left(\frac{r_\varepsilon \rho}{|8\pi^2 - r_\varepsilon \rho|} \vee e^{-(8\pi^2 - r_\varepsilon \rho)} \right) e^{-(8\pi^2 \wedge r_\varepsilon \rho)t}.$$

In any case one has, for all $t \geq 1$

$$I_1 \leq \mathcal{K}_1 e^{-\left((8\pi^2-\varepsilon)\wedge r_\varepsilon\rho\right)t},$$

where $\mathcal{K}_1 = \mathcal{K}_1(\varepsilon) = \left(1 + \frac{M^2}{2\rho^2\varepsilon}\right) E_M(0) \left(8\pi^2 \frac{e^{-1-\varepsilon}}{\varepsilon} \vee \frac{r_\varepsilon\rho}{|8\pi^2 - r_\varepsilon\rho|} \vee e^{-(8\pi^2-r_\varepsilon\rho)} \vee 1\right) > 0.$

• Now consider, for all $t \geq 1$, $I_2 := K_2 \int_1^t e^{-4\pi^2 s - r_\varepsilon\rho(t-s)}\,\mathrm{d}s$. We distinguish between two cases:

(i) If $r_\varepsilon\rho \neq 4\pi^2$ then, for all $t \geq 1$:

$$K_2 \int_1^t e^{-4\pi^2 s - r_\varepsilon\rho(t-s)}\,\mathrm{d}s \leq \frac{K_2}{|4\pi^2 - r_\varepsilon\rho|} e^{-\left(4\pi^2 \wedge r_\varepsilon\rho\right)t}.$$

(ii) If $r_\varepsilon\rho = 4\pi^2$ then, for all $t \geq 1$:

$$K_2 \int_1^t e^{-4\pi^2 s - r_\varepsilon\rho(t-s)}\,\mathrm{d}s = K_2 e^{-4\pi^2 t}(t-1),$$

and, since $(t-1) \leq \frac{e^{-1-\delta}}{\delta} e^{\delta t}$ for all $\delta > 0$, considering $\delta = \varepsilon$, one gets that, for all $t \geq 1$:

$$K_2 \int_1^t e^{-4\pi^2 s - r_\varepsilon\rho(t-s)}\,\mathrm{d}s \leq K_2 \frac{e^{-1-\varepsilon}}{\varepsilon} e^{-\left(4\pi^2-\varepsilon\right)t}.$$

In any case one has, for all $t \geq 1$:

$$I_2 \leq \mathcal{K}_2 e^{-\left((4\pi^2-\varepsilon)\wedge r_\varepsilon\rho\right)t},$$

where $\mathcal{K}_2 = \mathcal{K}_2(\varepsilon) = \left(\frac{\tilde{C}^2}{4} + \frac{M^2\tilde{C}^2}{2\rho^2\varepsilon}\right)\left(\frac{1}{|4\pi^2 - r_\varepsilon\rho|} \vee \frac{e^{-1-\varepsilon}}{\varepsilon}\right) > 0.$

Hence, recalling that $r_\varepsilon = 2(1-\varepsilon)$ one gets that for all $\varepsilon > 0$, for all $t \geq 1$,

$$\begin{aligned} E_m(t) &\leq E_m(1) e^{r_\varepsilon\rho}\, e^{-r_\varepsilon\rho t} + \mathcal{K}_1 e^{-\left((8\pi^2-\varepsilon)\wedge r_\varepsilon\rho\right)t} + \mathcal{K}_2 e^{-\left((4\pi^2-\varepsilon)\wedge r_\varepsilon\rho\right)t} \\ &\leq \mathcal{K}_3 e^{-\left((4\pi^2\wedge 2\rho)-\varepsilon\right)t}, \end{aligned}$$

for some $\mathcal{K}_3 = \mathcal{K}_3(\varepsilon) = \left(E_m(1)e^{2\rho-\varepsilon} \vee \tilde{\mathcal{K}}_1 \vee \tilde{\mathcal{K}}_2\right) > 0$, where

$$\begin{cases} \tilde{\mathcal{K}}_1 = \left(1 + \frac{M^2}{\rho\varepsilon}\right) E_M(0) \left(\frac{16\pi^2\rho\; e^{-(1+\frac{\varepsilon}{2\rho})}}{\varepsilon} \vee \frac{(2\rho-\varepsilon)}{|8\pi^2 - (2\rho-\varepsilon)|} \vee e^{-(8\pi^2-(2\rho-\varepsilon))} \vee 1\right) \\ \tilde{\mathcal{K}}_2 = \left(\frac{\tilde{C}^2}{4} + \frac{M^2\tilde{C}^2}{\rho\varepsilon}\right)\left(\frac{1}{|4\pi^2 - (2\rho-\varepsilon)|} \vee \frac{2\rho e^{-(1+\frac{\varepsilon}{2\rho})}}{\varepsilon}\right) \end{cases}.$$

Step 5: It remains to treat the case where $t \in [0,1]$. We have:

$$\left(\int_{\mathbb{T}^m} |\nabla H_t - \nabla\tilde{H}_t|^2\, \pi_t^\xi\right)^{\frac{1}{2}} \leq \|\pi_t^\xi\|_{L^2(\mathbb{T}^m)}^{\frac{1}{2}} \|\nabla H_t - \nabla\tilde{H}_t\|_{L^4(\mathbb{T}^m)}, \quad \forall t \in [0,1].$$

From (3.12), there exists $C_2 > 0$ such that for all $t \in [0,1]$, $\|\pi_t^\xi\|_{L^2(\mathbb{T}^m)}^{\frac{1}{2}} \leq C_2$, and, using Lemma 15.13 from [2], there exists $C_4 > 0$ such that for all $t \in [0,1]$,

$$\|\nabla H_t\|_{L^4(\mathbb{T}^m)} \leq C_4\|G_t\|_{L^4(\mathbb{T}^m)} \leq C_4\|\mathcal{F}\|_\infty \leq C_4\|\nabla V\|_\infty < \infty.$$

Similarly, one has $\|\nabla \tilde{H}_t\|_{L^4(\mathbb{T}^m)} \leq C_4 \|\nabla V\|_\infty$. Hence inequality (5.4) becomes, for all $\varepsilon > 0$ and for all $t \in [0,1]$:

$$\begin{aligned}
J_t &\leq 2C_2C_4\|\nabla V\|_\infty \left(\sqrt{I_M}(t) + M\sqrt{\frac{2}{\rho}}\sqrt{E_m(t)}\right)\\
&\leq 2C_2C_4\|\nabla V\|_\infty \left(\sqrt{I_M}(t) + \sqrt{\frac{2M^2}{\varepsilon\rho^2}}\sqrt{\varepsilon\rho E_m(t)}\right)\\
&\leq \varepsilon\rho E_m(t) + I_M(t) + (C_2C_4\|\nabla V\|_\infty)^2\left(1+\frac{2M^2}{\varepsilon\rho^2}\right).
\end{aligned}$$

It yields, from inequality (5.2), for all $\varepsilon > 0$ and for all $t \in [0,1[$:

$$\frac{\mathrm{d}E_m}{\mathrm{d}t} \leq -r_\varepsilon \rho E_m(t) + K_1 I_M(t) + K_2,$$

with

$$K_1 = K_1(\varepsilon) = 1 + \frac{M^2}{2\varepsilon\rho^2}, \qquad K_2 = K_2(\varepsilon) = (C_2C_4\|\nabla V\|_\infty)^2\left(1+\frac{2M^2}{\varepsilon\rho^2}\right).$$

The Gronwall's lemma yields, for all $\varepsilon > 0$ and for all $t \in [0,1[$:

$$\begin{aligned}
E_m(t) &\leq E_m(0)e^{-r_\varepsilon\rho t} + K_1\int_0^t I_M(s)e^{-r_\varepsilon\rho(t-s)}\,\mathrm{d}s + K_2\int_0^t e^{-r_\varepsilon\rho(t-s)}\,\mathrm{d}s\\
&\leq E_m(0) + K_1e^0\int_0^\infty I_M(s)\,\mathrm{d}s + \frac{K_2}{r_\varepsilon\rho}\left(1-e^{-r_\varepsilon\rho t}\right)\\
&\leq E_m(0) + K_1E_M(0) + \frac{K_2}{r_\varepsilon\rho},
\end{aligned}$$

where we used (5.3). Hence, for all $\varepsilon > 0$ and for all $t \in [0,1[$

$$E_m(t)e^{\left((4\pi^2\wedge 2\rho)-\varepsilon\right)} \leq \left(E_m(0) + K_1E_M(0) + \frac{K_2}{r_\varepsilon\rho}\right)e^{\left((4\pi^2\wedge 2\rho)-\varepsilon\right)} < +\infty.$$

Conclusion: for the PABF algorithm, we have obtained that for all $\varepsilon > 0$, there exists $C = C(\varepsilon) > 0$ such that, for all $t \geq 0$,

$$E_m(t) \leq \ Ce^{-\left((4\pi^2\wedge 2\rho)-\varepsilon\right)t}.$$

Recall that by Proposition 2.4, $E_M(t) \leq E_M(0)e^{-8\pi^2 t}$ for all $t \geq 0$. The decomposition $E(t) = E_m(t) + E_M(t)$ concludes the proof. □

5.3. Proof of Theorem 2.13

Let us prove Theorem 2.13.

Proof. Using Lemma 5.2 one gets:

$$\begin{aligned}
\frac{\mathrm{d}E}{\mathrm{d}t} &= -\int_{\mathbb{T}^n}\left|\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2\pi_t + \int_{\mathbb{T}^n}(B_t - B)\cdot\nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\pi_t\\
&\leq -\int_{\mathbb{T}^n}\left|\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2\pi_t + \left(\int_{\mathbb{T}^m}|B_t - B|^2\,\pi_t^\xi\right)^{\frac{1}{2}}\left(\int_{\mathbb{T}^n}\left|\nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2\pi_t\right)^{\frac{1}{2}}.
\end{aligned} \tag{5.5}$$

Step 1: Let us first consider $t \geq 1$. In the PABF case, since an orthogonal projection contracts the corresponding norm, for all $t \geq 1$:

$$\begin{aligned}\int_{\mathbb{T}^m} |\nabla H_t - \nabla H_\infty|^2 \pi_t^\xi &\leq \|\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |\nabla H_t - \nabla H_\infty|^2 \\ &\leq \|\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |G_t - G_\infty|^2 \\ &\leq \|\pi_t^\xi\|_\infty \|1/\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |G_t - G_\infty|^2 \pi_t^\xi \\ &\leq \left(1 + Ce^{-4\pi^2 t}\right) \int_{\mathbb{T}^m} |G_t - G_\infty|^2 \pi_t^\xi,\end{aligned}$$

for some $C > 0$ according to Proposition 2.5. Together with Lemma 5.1 and the microscopic log-Sobolev inequality (5.1), we have thus obtained for all $t \geq 1$, in both the ABF case (where $B_t = G_t$ and $B_\infty = G_\infty$) and PABF case (where $B_t = \nabla H_t$ and $B_\infty = \nabla H_\infty$),

$$\begin{aligned}\left(\int_{\mathbb{T}^m} |B_t - B_\infty|^2 \pi_t^\xi\right)^{\frac{1}{2}} &\leq \sqrt{1 + Ce^{-4\pi^2 t}} M \sqrt{\frac{2}{\rho}} \sqrt{E_m(t)} \\ &\leq \sqrt{1 + Ce^{-4\pi^2 t}} M \sqrt{\frac{2}{\rho}} \frac{1}{\sqrt{2\rho}} \left(\int_{\mathbb{T}^n} \left|\nabla_y \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t\right)^{\frac{1}{2}}.\end{aligned}$$

As a consequence,

$$\begin{aligned}\frac{\mathrm{d}E}{\mathrm{d}t} &\leq -\int_{\mathbb{T}^n} \left|\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t + \frac{M}{\rho}\sqrt{1 + Ce^{-4\pi^2 t}} \left(\int_{\mathbb{T}^n} \left|\nabla_y \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t\right)^{\frac{1}{2}} \left(\int_{\mathbb{T}^n} \left|\nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t\right)^{\frac{1}{2}} \\ &\leq \left(-1 + \frac{M}{2\rho} + C'e^{-2\pi^2 t}\right) \int_{\mathbb{T}^n} |\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)|^2 \pi_t .\end{aligned}$$

with $C' = M\sqrt{C}/(2\rho)$. Since we assumed $M < 2\rho$, there exists $t_0 \geq 1$ such that for all $t \geq t_0$, the right hand side is negative:

$$-1 + \frac{M}{2\rho} + C'e^{-2\pi^2 t} := -\alpha(t) \leq 0, \quad \forall t \geq t_0.$$

And, given the logarithmic-Sobolec inequality of constant $R > 0$ satisfied by π_∞:

$$\frac{\mathrm{d}E}{\mathrm{d}t} \leq -2\alpha(t) R E(t) \quad \forall t \geq t_0.$$

Hence by Gronwall's lemma, for all $t \geq t_0$:

$$E(t) \;\leq\; E(t_0) \exp\left(-2R\int_{t_0}^t \alpha(s)\mathrm{d}s\right) \;=\; E(t_0)\exp\left(-2R\left(1 - \frac{M}{2\rho}\right)(t - t_0) + \frac{C'R}{2\pi^2}\right)$$

Step 2: As for times $t \in [0, t_0]$, as in the third step of the proof of Theorem 2.11, there exists $C_2 > 0$ and $C_4 > 0$ such that for all $t \in [0, t_0]$:

$$\left(\int_{\mathbb{T}^m} |B_t - B_\infty|^2 \,\pi_t^\xi\right)^{\frac{1}{2}} \leq \sqrt{\|\pi_t^\xi\|_2} \|B_t - B_\infty\|_4 \;\leq\; 2C_2C_4\|\mathcal{F}\|_\infty.$$

Inequality (5.5) becomes, for all $t \in [0, t_0]$:

$$\begin{aligned}\frac{\mathrm{d}E(t)}{\mathrm{d}t} &\leq -\int_{\mathbb{T}^n} \left|\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t + 2C_2C_4\|\mathcal{F}\|_\infty \left(\int_{\mathbb{T}^n} \left|\nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right)\right|^2 \pi_t\right)^{\frac{1}{2}} \\ &\leq C_2^2C_4^2\|\mathcal{F}\|_\infty^2.\end{aligned}$$

Hence, for all $t \in [0, t_0]$

$$E(t) \leq E(0) + \left(C_2C_4\|\mathcal{F}\|_\infty\right)^2 t,$$

and

$$E(t)e^{2R\left(1-\frac{M}{2\rho}\right)t} \leq \left(E(0) + \left(C_2C_4\|\mathcal{F}\|_\infty\right)^2 t_0\right) e^{2R\left(1-\frac{M}{2\rho}\right)t_0}$$

which concludes the proof, relying on the same argument as in the proof of Theorem 2.11. □

5.4. Proof of Corollary 2.15

Proof. Similarly to the previous proofs, using Lemma 5.1 and Proposition 2.5, there exists $C > 0$ such that, for all $t \geq 1$:

$$\begin{aligned}\int_{\mathbb{T}^m} |G_t - G_\infty|^2 \mathrm{d}x &\leq \|1/\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |G_t - G|^2 \pi_t^\xi \\ &\leq \ (1 + Ce^{-4\pi^2 t})\frac{2M}{\rho} E_m(t) \\ &\leq (1 + Ce^{-4\pi^2 t})\frac{2M}{\rho} K e^{-\Lambda t},\end{aligned}$$

where we used either Theorem 2.11 or 2.13.

For $t \in [0, 1]$, we simply bound

$$\int_{\mathbb{T}^m} |G_t - G|^2 \mathrm{d}x \ \leq 2\ \|\mathcal{F}\|_\infty^2 .$$

This concludes the ABF case, for which $B_t = G_t$ and $B_\infty = G_\infty$. Besides, the L^2-norm is decreased by the Helmholtz projection, which concludes the PABF case. □

Acknowledgements. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 810367), project EMC2. P. Monmarché acknowledges partial support by the projects EFI ANR-17-CE40-0030.

References

[1] H. Alrachid and T. Lelièvre, Long-time convergence of an adaptive biasing force method: variance reduction by Helmholtz projection. *J. Comput. Math.* **1** (2015) 55–82.

[2] L. Ambrosio, A. Carlotto and A. Massaccesi, Lectures on Elliptic Partial Differential Equations (2018).

[3] D. Bakry, I. Gentil and M. Ledoux, Analysis and Geometry of Markov Diffusion operators. Vol. 348 of *Grundlehren der mathematischen Wissenschaften.* Springer (2014).

[4] V.I. Bogachev, N.V. Krylov, M. Röckner and S.V. Shaposhnikov, Fokker–Planck–Kolmogorov equations. American Mathematical Society (2015).

[5] H. Brézis, Analyse fonctionnelle : théorie et applications. Masson (1987).

[6] S. Chmiela, A. Tkatchenko, H.E. Sauceda, I. Poltavsky, K. Schütt and K.-R. Müller, Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3** (2017) 05.

[7] E. Darve and A. Pohorille, Calculating free energies using average force. *J. Chem. Phys.* **115** (2001) 9169–9183.

[8] E.B. Dynkin, Markov Processes Volume I. Springer Verlag (1965).

[9] E.B. Dynkin, Markov Processes Volume II. Springer Verlag (1965).

[10] L.C. Evans, Partial Differential Equations. American Mathematical Society (2010).

[11] 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** (2016) 3506–3513.

[12] P. Gkeka, G. Stoltz, A. Barati 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. Preprint arXiv:2004.06950 (2020).

[13] J. Hénin and C. Chipot, Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* **121** (2004) 2904–2914.

[14] L. Hörmander, Hypoelliptic second order differential equations. *Acta Math.* **119** (1967) 147–171.

[15] B. Jourdain, T. Lelièvre and R. Roux, Existence, uniqueness and convergence of a particle approximation for the Adaptive Biasing Force process. *ESAIM: M2AN* **44** (2010) 831–865.

[16] X. Kong and C.L. Brooks, $\lambda-$ dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **105** (1996) 2414–2423.

[17] N. Krylov, On diffusion processes with drift in L_d. *Probab. Theory Relat. Fields* (2020).

[18] T. Lelièvre and G. Stoltz, Partial differential equations and stochastic methods in molecular dynamics. *Acta Numer.* **25** (2016).

[19] T. Lelièvre, M. Rousset and G. Stoltz, Long-time convergence of an Adaptive Biasing Force method. *Nonlinearity* **21** (2008).

[20] T. Lelièvre, M. Rousset and G. Stoltz, Free Energy Computations. Imperial College Press (2010).

[21] G. Menz and A. Schlichting, Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *Ann. Probab.* **42** (2014) 1809–1884.

[22] S.P. Meyn and R.L. Tweedie, Markov Chains and Stochastic Stability. Springer Verlag, London (1993).

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

[24] A. Niklasson, C. Tymczak and M. Challacombe, Time-reversible born-oppenheimer molecular dynamics. *Phys. Rev. Lett.* **97** (2006) 10.

[25] F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** (2000) 361–400.

[26] P. Pulay and G. Fogarasi, Fock matrix dynamics. *Chem. Phys. Lett.* **386** (2004) 03.