Effective dynamics using conditional expectations

Frédéric Legoll[1,2] **and Tony Lelièvre**[3,2]

[1] Université Paris-Est, Institut Navier, LAMI, École des Ponts, 6 et 8 avenue Blaise Pascal, 77455 Marne-La-Vallée Cedex 2, France

[2] INRIA Rocquencourt, MICMAC Team-Project, Domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, France

[3] Université Paris-Est, CERMICS, École des Ponts, 6 et 8 avenue Blaise Pascal, 77455 Marne-La-Vallée Cedex 2, France

E-mail: legoll@lami.enpc.fr, lelievre@cermics.enpc.fr

Abstract. The question of coarse-graining is ubiquitous in molecular dynamics. In this article, we are interested in deriving effective properties for the *dynamics* of a coarse-grained variable $\xi(x)$, where x describes the configuration of the system in a high-dimensional space $\mathbb{R}^n$, and ξ is a smooth function with value in $\mathbb{R}$ (typically a reaction coordinate). It is well known that, given a Boltzmann-Gibbs distribution on $x \in \mathbb{R}^n$, the equilibrium properties on $\xi(x)$ are completely determined by the free energy. On the other hand, the question of the effective dynamics on $\xi(x)$ is much more difficult to address. Starting from an overdamped Langevin equation on $x \in \mathbb{R}^n$, we propose an effective dynamics for $\xi(x) \in \mathbb{R}$ using conditional expectations. Using entropy methods, we give sufficient conditions for the time marginals of the effective dynamics to be close to the original ones. We check numerically on some toy examples that these sufficient conditions yield an effective dynamics which accurately reproduces the residence times in the potential energy wells. We also discuss the accuracy of the effective dynamics in a pathwise sense, and the relevance of the free energy to build a coarse-grained dynamics.

AMS classification scheme numbers: 35B40, 82C31, 60H10

1. Motivation

In molecular dynamics, two types of quantities are typically of interest: averages with respect to the canonical ensemble (thermodynamic quantities, such as stress, or heat capacity), and averages of functionals over paths (dynamic quantities, like viscosity, diffusion coefficients or rate constants). In both cases, the question of coarse-graining is relevant, in the sense that the considered functionals typically depend only on a few variables of the system (collective variables, or reaction coordinates) so that it would be interesting to obtain coarse-grained models on these variables.

1.1. Coarse-graining of thermodynamic quantities

Computing canonical averages is a standard task in molecular dynamics. For a molecular system whose atom positions are described by a vector $x \in \mathbb{R}^n$, these quantities read

$$\int_{\mathbb{R}^n} \Phi(x)\, d\mu \tag{1}$$

where $\Phi : \mathbb{R}^n \to \mathbb{R}$ is the observable of interest and μ is the Boltzmann-Gibbs measure,

$$d\mu = Z^{-1} \exp(-\beta V(x))\, dx, \tag{2}$$

where V is the potential energy of the system, β is proportional to the inverse of the system temperature, and $Z = \displaystyle\int_{\mathbb{R}^n} \exp(-\beta V(x))\, dx$ is a normalizing constant. Typically, x represents the position of N three-dimensional particles, hence $x \in \mathbb{R}^n$ with $n = 3N$. All the results we prove are also satisfied if $x \in \mathbb{T}^n$, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denotes the one-dimensional torus.

As mentioned above, observables of interest are often function of only part of the variable x. For example, x denotes the positions of *all* the atoms of a protein and of the solvent molecules around, and the quantity of interest is only a particular angle between some atoms in the protein, because this angle characterizes the conformation of the protein (and thus the potential energy well in which the system is is completely determined by the knowledge of this quantity of interest). We thus introduce the so-called *reaction coordinate*

$$\xi : \mathbb{R}^n \to \mathbb{R},$$

which contains all the information we are interested in ‡. Throughout this article, we assume that

$$\textbf{[H1]} \qquad \begin{array}{c} \xi \text{ is a smooth } \textit{scalar} \text{ function such that,} \\ \text{for all } x \in \mathbb{R}^n,\ 0 < m \le |\nabla \xi(x)| \le M < \infty. \end{array}$$

We have supposed that ξ is a scalar function. It is not clear to us whether the results of this article can be generalized to the case of a multi-dimensional reaction coordinate.

‡ In this article, we do not address the difficult question of how to find a good reaction coordinate. See for instance [24] for some discussion on that point.

To this function ξ is naturally associated an effective energy A, called the *free energy*, such that

$$d(\xi \star \mu) = \exp(-\beta A(z))\, dz, \tag{3}$$

where $\xi \star \mu$ denotes the image of the measure μ by ξ. In other words, for any test function $\Phi : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} \Phi(\xi(x))\, Z^{-1} \exp(-\beta V(x))\, dx = \int_{\mathbb{R}} \Phi(z)\, \exp(-\beta A(z))\, dz. \tag{4}$$

Expressions of A and its derivative are given below (see Section 2.1).

The interpretation of (4) is that, when X is distributed according to the Boltzmann measure (2), then $\xi(X)$ is distributed according to the measure $\exp(-\beta A(z))\, dz$. Hence, the free energy A is a relevant quantity for computing thermodynamic quantities, namely canonical averages.

In conclusion, the question of coarse-graining thermodynamic quantities amounts to computing the free energy, and there are several efficient methods to perform such calculations (see for example [6]). There are also interesting questions related to computing approximations of the free energy, especially when the number of reaction coordinates is large, for example in polymer science, but this is not the subject of this article.

1.2. Coarse-graining of dynamical quantities

The objective of this work is to address some issues related to the *dynamics* of the system, and how to coarse-grain it. In short, we aim at designing a dynamics that approximates the path $t \mapsto \xi(X_t)$, where ξ is the above reaction coordinate.

To make this question precise, we first have to *choose* the full dynamics, which will be the reference one. In the following, we consider the overdamped Langevin dynamics on state space $\mathbb{R}^n$ (we will discuss this choice below),

$$dX_t = -\nabla V(X_t)\, dt + \sqrt{2\beta^{-1}}\, dW_t, \quad X_{t=0} = X_0, \tag{5}$$

where W_t is a standard n-dimensional Brownian motion. Under suitable assumptions on V, this dynamics is ergodic with respect to the Boltzmann-Gibbs measure (2). Hence, for μ-almost all initial conditions X_0,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \Phi(X_t)\, dt = \int_{\mathbb{R}^n} \Phi(x)\, d\mu \tag{6}$$

almost surely. In practice, this convergence is very slow, due to some metastabilities in the dynamics: X_t samples a given well of the potential energy for a long time, before hoping to some other well of V.

An important dynamical quantity we will consider below is the average residence time, that is the mean time that the system spends in a given well, before hoping to another one, when it follows the dynamics (5). Typically, the wells are fully described through ξ (x is in a given well if and only if $\xi(x)$ is in a given interval), so that these times can be obtained from the knowledge of the time evolution $\xi(X_t)$, which is expensive to compute since it means simulating the full system.

In this article, our aim is twofold. First, we would like to propose a one-dimensional dynamics of the form

$$dy_t = b(y_t)\,dt + \sqrt{2\beta^{-1}}\,\sigma(y_t)\,dB_t, \tag{7}$$

where B_t is a standard one-dimensional Brownian motion and b and σ are scalar functions, such that $(y_t)_{0\le t\le T}$ is a good approximation (in a sense to be made precise below) of $(\xi(X_t))_{0\le t\le T}$. Hence, the dynamics (7) can be thought of as a coarse-grained, or *effective*, dynamics for the quantity of interest. A natural requirement is that (7) preserves equilibrium quantities, *i.e.* it is ergodic with respect to $\exp(-\beta A(z))\,dz$, the equilibrium measure of $\xi(X)$, but we typically ask for more than that. For example, we would like to be able to recover residence times in the wells from (7), hence bypassing the expensive simulation of $\xi(X_t)$ (see Section 4 for some numerical results on that quantity).

Second, we would like to investigate the relation between (7) and the coarse-grained dynamics

$$d\overline{y}_t = -A'(\overline{y}_t)\,dt + \sqrt{2\beta^{-1}}\,dB_t, \tag{8}$$

which is indeed a one-dimensional dynamics, driven by the free energy, and ergodic for $\exp(-\beta A(z))\,dz$. In other words, what is the dynamical content of the free energy? This second question stems from the fact that practitioners often look at the free energy profile (*i.e.* the function $z \mapsto A(z)$) to get an idea of the dynamics of transition (typically the transition time) between one region indexed by the reaction coordinate (say for example $\{x \in \mathbb{R}^n;\ \xi(x) \le z_0\}$) and another one (for example $\{x \in \mathbb{R}^n;\ \xi(x) > z_0\}$). If $\xi(X_t)$ follows a dynamics which is close to (8), then the Transition State Theory says that residence times are a function of the free energy barriers [22, 14], and then it makes sense to look at the free energy to compute some dynamical properties. It is thus often assumed that there is some dynamical information in the free energy A.

The difficulty of the question we address stems from the fact that, in general, $t \to \xi(X_t)$ is not a Markov process: this is a closure problem. A first possibility is to try and approximate $\xi(X_t)$ by a process which has some memory in time, typically a generalized Langevin equation (see for instance [8, 19], and also [15]). A standard framework is then the Mori-Zwanzig projection formalism, which is described in details in [11]. Note also that, since we are interested in reproducing only some output function of X_t (namely $\xi(X_t)$), tools from the control theory may be used. Such an idea has been followed in [17, 16].

If a time-scale separation is present in the system, then memory effects may be neglected. In the sequel, we make such time-scale separation assumptions (see assumptions [**H2**] and [**H3**] of Proposition 3.1), which allow us to approximate $\xi(X_t)$ by a Markov process of the type (7). We use the framework of logarithmic Sobolev inequalities to write these assumptions. It has the advantage that we do not assume to *a priori* know how to split x between fast and slow modes, or to split the potential energy V between fast and slow terms (otherwise stated, the time scale separation is encoded in the constants entering the logarithmic Sobolev inequalities, and not inserted *a priori* in the model). In addition, within this framework, we can handle reaction coordinates that are nonlinear functions of x, the natural cartesian coordinates of the system (see the numerical simulations reported in Section 4).

Another possibility is to start from a dynamics which includes an *explicit* small parameter, representing a time scale separation. One may then apply an averaging principle (see [15] and the references therein for more details along this idea; see also [26] for a comprehensive review of the averaging principle, when applied to deterministic and stochastic differential equations). In Section 3.2, we consider such a case of potential energy being the sum of two terms of different stiffness, as an *example* of application of our general result (see the potential energy (47)). Note that, even if we explicitly insert a small parameter in V, our model differs from the one considered in [34], where a small parameter appears in the potential energy *and* in the diffusion coefficient.

Other strategies are to try and identify fast and slow modes in the dynamics (see e.g. [33, 20]), or to postulate a parametric form for the effective dynamics and to identify its coefficients by numerical simulation on the complete system [27, 36].

We finish this section by a discussion of the choice of the full dynamics. We chose the overdamped Langevin dynamics (5). Other choices can be made, in particular the Langevin dynamics, which is closer to a Hamiltonian dynamics and can also be seen as a method to sample the canonical measure (see [5] for a review of sampling methods of the canonical ensemble, along with a theoretical and numerical comparison of their performances for molecular dynamics). From the analysis standpoint, the dynamics we chose is much simpler, since the diffusion is non-degenerate (in contrast to the Langevin dynamics, which is an hypoelliptic equation). We do not know whether the theoretical results presented in this article (such as Proposition 3.1) can be generalized to the case of the Langevin dynamics. From a practical viewpoint, it may be possible to use the same strategy starting from the Langevin dynamics to write another low-dimensional dynamics. We have not pursued in this direction. As an alternative to continuous time processes, one can also model the dynamics of a molecular system by a discrete time Markov chain, for instance in a discrete state space, where each state represents a different metastable configuration of the system [31, 32]. In that setting, the question of estimating the accuracy of a coarse-grained dynamics has been addressed in [30], where similar bounds as those derived in this article are obtained.

1.3. Statement of the main results and outline

We propose a way to derive an effective dynamics of the form (7). This defines a process $(y_t)_{t\geq 0}$, which we compare with $(\xi(X_t))_{t\geq 0}$, where X_t satisfies (5). Three quantities can be typically considered to estimate the distance between y_t and $\xi(X_t)$ (on the time interval $[0,T]$):

- [D1] pathwise convergence: $\mathbb{E}\left(\sup_{t\in(0,T)} |\xi(X_t) - y_t|^2\right)$,
- [D2] convergence of the laws of paths: $\|\mathcal{L}(\xi(X_t)_{0\le t\le T}) - \mathcal{L}((y_t)_{0\le t\le T})\|_{TV}$,
- [D3] convergence of time marginals: $\sup_{t\in(0,T)} \|\mathcal{L}(\xi(X_t)) - \mathcal{L}(y_t)\|_{TV}$.

In the above estimators, we have arbitrarily chosen to measure distances between probability measures by the total variation (TV) norm, but other choices could be made. Recall that the total variation of a signed measure ν is defined by $\|\nu\|_{TV} = \sup_{f\in L^\infty, \|f\|_{L^\infty}\le 1} \int f d\nu$. If ν is a measure on $\mathbb{R}^n$ which has a density with respect to the Lebesgue measure, then its total variation is just the L^1 norm of its density.

It is clear that a bound in the sense of [D1] implies a bound in the sense of [D2], which implies a bound in the sense of [D3]. Conversely, by the Skorohod theorem, a bound in the sense of [D2] implies a bound in the sense of [D1], for some well chosen realizations of W_t and B_t (the brownian motions in (5) and (7)), but this theorem is not constructive. The most relevant criterion in practice is [D2]. Indeed, the criterion [D3] does not account for the correlations in time of the process, which are important to understand its dynamical properties. On the other hand, the pathwise convergence criterion [D1] is too strong: practionners in molecular dynamics are rarely interested in the trajectory *per se*. Moreover, [D2] implies the convergence of the law of escape times (hence of residence times in the wells), at least if the escape time is (almost surely) a continuous function of paths, which holds under some regularity assumptions (see [3, Exercise 3.9.10]).

Our first objective is to propose, in a general case, some sufficient conditions on the reaction coordinate for a bound of type [D3] to be satisfied. We are actually able to derive an estimate of the difference between the time marginals which is *uniform in time*. Next, on a toy-model, we investigate, both theoretically and numerically, if these conditions are sufficient and necessary for [D1] and [D2] to hold.

The article is organized as follows. In Section 2, after introducing some notation and recalling some basic relations concerning the free energy, we propose a natural coarse-graining procedure, which enables us to obtain an effective dynamics of type (7), where the functions b and σ can easily be computed (see Equations (24), (25) and (26)). In Section 3, we prove that the solution y_t of the effective dynamics (26) is indeed a good approximation of $\xi(X_t)$, in the sense [D3]. Our argument relies on entropy techniques, and is very much inspired by [12, 9]. In Section 4, we present some numerical results obtained on a simple model, where we compare residence times in the potential energy wells as predicted by the reference dynamics (5) and by the one-dimensional reduced dynamics (26). Section 5 is dedicated to establishing error estimates in the sense [D1] of pathwise convergence, in a specific case. These estimates are illustrated by numerical simulations.

2. A "natural" coarse-graining procedure

2.1. Notation

We gather here some useful notation and results. Let Σ_z be the submanifold of $\mathbb{R}^n$ of positions at a fixed value of the reaction coordinate:

$$\Sigma_z = \{x \in \mathbb{R}^n;\ \xi(x) = z\}.$$

Let us introduce μ_{Σ_z}, which is the probability measure μ conditioned at a fixed value of the reaction coordinate:

$$d\mu_{\Sigma_z} = \frac{\exp(-\beta V)\, |\nabla \xi|^{-1}\, d\sigma_{\Sigma_z}}{\displaystyle\int_{\Sigma_z} \exp(-\beta V)\, |\nabla \xi|^{-1}\, d\sigma_{\Sigma_z}}, \tag{9}$$

where the measure σ_{Σ_z} is the Lebesgue measure on Σ_z induced by the Lebesgue measure in the ambient Euclidean space $\mathbb{R}^n \supset \Sigma_z$.

We recall the following expressions for the free energy A and its derivative A', also called the *mean force* (see [7]):

$$A(z) = -\beta^{-1} \ln \left(\int_{\Sigma_z} Z^{-1} \exp(-\beta V)\, |\nabla \xi|^{-1}\, d\sigma_{\Sigma_z} \right) \tag{10}$$

and

$$A'(z) = \int_{\Sigma_z} F\, d\mu_{\Sigma_z}, \tag{11}$$

where F is the so-called *local mean force*:

$$F = \frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \operatorname{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right). \tag{12}$$

In view of (10), note that (9) reads

$$d\mu_{\Sigma_z} = \frac{\exp(-\beta V)\, |\nabla \xi|^{-1}\, d\sigma_{\Sigma_z}}{Z \exp(-\beta A)}. \tag{13}$$

These expressions can be obtained by the co-area formula [10], which we now recall:

Lemma 2.1 *For any smooth function $\Phi : \mathbb{R}^n \to \mathbb{R}$,*

$$\int_{\mathbb{R}^n} \Phi(x)\, |\nabla \xi(x)|\, dx = \int_{\mathbb{R}} \int_{\Sigma_z} \Phi\, d\sigma_{\Sigma_z}\, dz. \tag{14}$$

Remark 2.1 (Co-area formula and conditioning) *The co-area formula shows that if the random variable X has law $\psi(x)\, dx$ in $\mathbb{R}^n$, then $\xi(X)$ has law $\psi^\xi(z)\, dz$, with*

$$\psi^\xi(z) = \int_{\Sigma_z} \psi\, |\nabla \xi|^{-1}\, d\sigma_{\Sigma_z}.$$

It also shows that the law of X conditioned to a fixed value z of $\xi(X)$ is μ_{Σ_z}, where μ_{Σ_z} is defined by (9). The measure $|\nabla \xi|^{-1} d\sigma_{\Sigma_z}$ is sometimes denoted by $\delta_{\xi(x)-z}$ in the literature.

From the co-area formula, we get the following result:

Lemma 2.2 *For any smooth function $\chi : \mathbb{R}^n \to \mathbb{R}$, consider*

$$\chi^\xi(z) = \int_{\Sigma_z} \chi \, |\nabla \xi|^{-1} \, d\sigma_{\Sigma_z}.$$

The derivative of χ^ξ reads:

$$\frac{d\chi^\xi}{dz}(z) = \int_{\Sigma_z} \left[\frac{\nabla \xi \cdot \nabla \chi}{|\nabla \xi|^2} + \chi \operatorname{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \right] |\nabla \xi|^{-1} \, d\sigma_{\Sigma_z}.$$

Proof: For any smooth test function $g : \mathbb{R} \to \mathbb{R}$, we obtain, using the co-area formula (14), that

$$\int_{\mathbb{R}} \chi^\xi(z) \, g'(z) \, dz = \int_{\mathbb{R}} \int_{\Sigma_z} \chi \, |\nabla \xi|^{-1} \, g'(z) \, d\sigma_{\Sigma_z} \, dz = \int_{\mathbb{R}^n} \chi(x) \, g'(\xi(x)) \, dx.$$

Hence,

$$\begin{aligned}
\int_{\mathbb{R}} \chi^\xi(z) \, g'(z) \, dz &= \int_{\mathbb{R}^n} \chi(x) \, g'(\xi(x)) \, dx \\
&= \int_{\mathbb{R}^n} \chi \, |\nabla \xi|^{-2} \, \nabla \xi \cdot \nabla (g \circ \xi) \\
&= - \int_{\mathbb{R}^n} g \circ \xi \ \operatorname{div} \ \left(\chi \, |\nabla \xi|^{-2} \, \nabla \xi \right) \\
&= - \int_{\mathbb{R}} g(z) \int_{\Sigma_z} \operatorname{div} \ \left(\chi \, |\nabla \xi|^{-2} \, \nabla \xi \right) \frac{d\sigma_{\Sigma_z}}{|\nabla \xi|} \, dz, \\
&= - \int_{\mathbb{R}} g(z) \int_{\Sigma_z} \left[\frac{\nabla \xi \cdot \nabla \chi}{|\nabla \xi|^2} + \chi \operatorname{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \right] \frac{d\sigma_{\Sigma_z}}{|\nabla \xi|} \, dz,
\end{aligned}$$

which yields the result. □

2.2. A non-closed equation

Consider X_t that solves (5). By a simple Itô computation, we have

$$d\xi(X_t) = \left(-\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi \right) (X_t) \, dt + \sqrt{2\beta^{-1}} \, |\nabla \xi|(X_t) \, dB_t \tag{15}$$

where B_t is the one-dimensional Brownian motion

$$dB_t = \frac{\nabla \xi}{|\nabla \xi|}(X_t) \cdot dW_t. \tag{16}$$

Of course, equation (15) is not closed. Following Gyöngy [13], a simple closing procedure is to consider $\widetilde{y}_t$ solution to

$$d\widetilde{y}_t = \widetilde{b}(t, \widetilde{y}_t) \, dt + \sqrt{2\beta^{-1}} \, \widetilde{\sigma}(t, \widetilde{y}_t) \, dB_t, \tag{17}$$

where

$$\widetilde{b}(t, y) = \mathbb{E} \left[\left(-\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi \right) (X_t) \mid \xi(X_t) = y \right] \tag{18}$$

and

$$\widetilde{\sigma}^2(t,y) = \mathbb{E}\left[|\nabla\xi|^2(X_t) \mid \xi(X_t) = y\right]. \tag{19}$$

Note that $\widetilde{b}$ and $\widetilde{\sigma}$ depend on t, since these are expected values conditioned on the fact that $\xi(X_t) = y$, where the probability distribution function of X_t of course depends on t.

As shown in [13], this procedure is exact from the point of view of time marginals, *i.e.* [D3] in our above classification. This is stated in the following lemma:

Lemma 2.3 *The probability distribution function ψ^ξ of $\xi(X_t)$, where X_t satisfies (5), satisfies the Fokker-Planck equation associated to (17):*

$$\partial_t \psi^\xi = \partial_z \left(-\widetilde{b}\ \psi^\xi + \beta^{-1} \partial_z (\widetilde{\sigma}^2 \psi^\xi) \right). \tag{20}$$

Proof: Let us denote $\psi(t,x)$ the probability distribution function of X_t. It satisfies the Fokker-Planck equation

$$\partial_t \psi = \operatorname{div}\ \left(\nabla V\, \psi + \beta^{-1} \nabla \psi \right). \tag{21}$$

In view of Remark 2.1, the probability distribution function $\psi^\xi(t,z)$ of $\xi(X_t)$ is given by

$$\psi^\xi(t,z) = \int_{\Sigma_z} \psi(t,\cdot)\, |\nabla\xi|^{-1}\, d\sigma_{\Sigma_z}.$$

Using Lemma 2.2 with $\chi \equiv \psi(t,\cdot)$, we obtain

$$\partial_z \psi^\xi(t,z) = \int_{\Sigma_z} \left(\frac{\nabla\xi \cdot \nabla\psi(t,\cdot)}{|\nabla\xi|^2} + \operatorname{div}\ \left(\frac{\nabla\xi}{|\nabla\xi|^2} \right) \psi(t,\cdot) \right) |\nabla\xi|^{-1}\ d\sigma_{\Sigma_z}. \tag{22}$$

By definition, we have the following expressions for $\widetilde{b}$ and $\widetilde{\sigma}$ in terms of ψ:

$$\begin{aligned} \widetilde{b}(t,z) \ &= \frac{1}{\psi^\xi(t,z)} \int_{\Sigma_z} \left(-\nabla V \cdot \nabla\xi + \beta^{-1} \Delta\xi \right) |\nabla\xi|^{-1}\ \psi\ d\sigma_{\Sigma_z}, \\ \widetilde{\sigma}^2(t,z) &= \frac{1}{\psi^\xi(t,z)} \int_{\Sigma_z} |\nabla\xi|\ \psi\ d\sigma_{\Sigma_z}. \end{aligned}$$

Using again Lemma 2.2 with $\chi \equiv |\nabla\xi|^2\ \psi(t,\cdot)$, we obtain

$$\partial_z(\widetilde{\sigma}^2\, \psi^\xi) = \partial_z \int_{\Sigma_z} |\nabla\xi|\, \psi\, d\sigma_{\Sigma_z} = \int_{\Sigma_z} \left(\nabla\xi \cdot \nabla\psi + \psi\, \Delta\xi \right) |\nabla\xi|^{-1}\ d\sigma_{\Sigma_z}. \tag{23}$$

Let us now prove a variational formulation of (20). For any test function g, we have

$$\begin{aligned}
\frac{d}{dt}\int_{\mathbb{R}} \psi^\xi(t,z)\, g(z)\, dz &= \frac{d}{dt}\int_{\mathbb{R}^n} \psi(t,x)\, g(\xi(x))\, dx \\
&= \int_{\mathbb{R}^n} \mathrm{div}\, \left(\psi \nabla V + \beta^{-1} \nabla \psi\right) g \circ \xi \ dx \\
&= -\int_{\mathbb{R}^n} \left(\psi \nabla V + \beta^{-1} \nabla \psi\right) \cdot \nabla \xi \ \ g' \circ \xi \ dx \\
&= -\int_{\mathbb{R}} \int_{\Sigma_z} |\nabla \xi|^{-1} \left(\psi \nabla V \cdot \nabla \xi + \beta^{-1} \nabla \psi \cdot \nabla \xi\right) \, d\sigma_{\Sigma_z} \ g'(z)\, dz \\
&= -\beta^{-1} \int_{\mathbb{R}} \partial_z (\widetilde{\sigma}^2 \psi^\xi)\, g'(z)\, dz \\
&\quad + \int_{\mathbb{R}} \int_{\Sigma_z} |\nabla \xi|^{-1} \left(-\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi\right) \ \psi \, d\sigma_{\Sigma_z}\, g'(z)\, dz \\
&= -\beta^{-1} \int_{\mathbb{R}} \partial_z (\widetilde{\sigma}^2 \, \psi^\xi)\, g'(z)\, dz + \int_{\mathbb{R}} \widetilde{b} \ \psi^\xi \ g'(z)\, dz.
\end{aligned}$$

This shows that ψ^ξ satisfies (20). □

2.3. A closed effective dynamics

The problem with equation (17) is that the functions $\widetilde{b}$ and $\widetilde{\sigma}$ are very complicated to compute, since they involve the full knowledge of ψ. Therefore, one cannot consider (17) as a reasonable closure. A natural simplification is to consider a time-independent approximation of the functions $\widetilde{b}$ and $\widetilde{\sigma}$. Considering (18) and (19), we introduce ($\mathbb{E}_\mu$ denoting a mean with respect to the measure μ)

$$\begin{aligned}
b(z) &= \mathbb{E}_\mu \left[\left(-\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi\right)(X) \mid \xi(X) = z \right], \\
&= \int_{\Sigma_z} \left(-\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi\right) d\mu_{\Sigma_z}, \qquad (24)
\end{aligned}$$

and

$$\begin{aligned}
\sigma^2(z) &= \mathbb{E}_\mu \left(|\nabla \xi|^2 (X) \mid \xi(X) = z \right), \\
&= \int_{\Sigma_z} |\nabla \xi|^2 \ d\mu_{\Sigma_z}, \qquad (25)
\end{aligned}$$

where μ_{Σ_z} is defined by (9). This simplification especially makes sense if $\xi(X_t)$ is a slow variable, that is if the characteristic evolution time of $\xi(X_t)$ is much larger than the characteristic time needed by X_t to sample the manifold Σ_z. This is quantified in the sequel.

In the spirit of (17), we next introduce the coarse-grained dynamics

$$\boxed{dy_t = b(y_t)\, dt + \sqrt{2\beta^{-1}} \ \sigma(y_t)\, dB_t, \quad y_{t=0} = \xi(X_0).} \qquad (26)$$

The Fokker-Planck equation associated to the above dynamics will be useful. It reads

$$\partial_t \phi = \partial_z \left(-b\, \phi + \beta^{-1} \partial_z (\sigma^2\, \phi) \right). \qquad (27)$$

Let us first prove that the dynamics (26) is ergodic for the equilibrium measure $\xi \star \mu$. The distance between y_t and $\xi(X_t)$ is estimated in Section 3.

In view of assumption **[H1]** and of (25), we observe that the diffusion coefficient of (26) satisfies $\sigma(y) \geq m > 0$ for any y. Hence, the process defined by (26) is irreducible, and admits a unique invariant probability measure. In the following lemma, we prove that $\exp(-\beta A(z))\,dz$ is a stationary measure for (26). Hence, the process y_t defined by (26) is ergodic with respect to this probability (see Has'minskii [18], Kliemann [21] and the references therein).

Lemma 2.4 *The measure $\xi \star \mu$ on $\mathbb{R}$, which has the density $\exp(-\beta A)$, is a stationary measure for (26).*

Proof: We infer from (25) and (13) that

$$\sigma^2 \exp(-\beta A) = Z^{-1} \int_{\Sigma_z} |\nabla \xi| \, \exp(-\beta V)\, d\sigma_{\Sigma_z}.$$

Using Lemma 2.2 with $\chi \equiv Z^{-1}\,|\nabla\xi|^2 \exp(-\beta V)$, we obtain

$$\begin{aligned}
&\beta^{-1}\partial_z(\sigma^2 \exp(-\beta A)) \\
&\quad = \beta^{-1} Z^{-1} \int_{\Sigma_z} [\nabla\xi \cdot \nabla(\exp(-\beta V)) + \exp(-\beta V)\Delta\xi] \; |\nabla\xi|^{-1}\, d\sigma_{\Sigma_z}, \\
&\quad = Z^{-1} \int_{\Sigma_z} \left(-\nabla\xi\cdot\nabla V + \beta^{-1}\Delta\xi\right)\exp(-\beta V)\; |\nabla\xi|^{-1}\, d\sigma_{\Sigma_z}, \\
&\quad = b\; \exp(-\beta A).
\end{aligned} \tag{28}$$

As a consequence of the above equation, (27) can be recast as

$$\begin{aligned}
\partial_t \phi &= \partial_z\left(-b\,\phi + \beta^{-1}\partial_z\left(\sigma^2 \exp(-\beta A)\, \exp(\beta A)\,\phi\right)\right) \\
&= \beta^{-1}\partial_z\left[\sigma^2\; \partial_z(\phi\exp(\beta A))\; \exp(-\beta A)\right].
\end{aligned} \tag{29}$$

It is now clear that $\phi = \exp(-\beta A)$ is a stationary solution of the above equation. □

In view of (29), we observe that $\phi = \exp(-\beta A)$ is not only a stationary measure for (26), but also satisfies a detailed balance condition ((y_t) is a reversible process with respect to $\exp(-\beta A(z))\,dz$).

Remark 2.2 *Let us set $\overline{f}(t,z) = \phi(t,z)\exp(\beta A(z))$ and let $\overline{g} : \mathbb{R} \to \mathbb{R}$ be a (time-independent) test function. Then a weak formulation of (29) is*

$$\frac{d}{dt}\int_{\mathbb{R}} \overline{f}(t,z)\; \overline{g}(z)\exp(-\beta A(z))\,dz = -\beta^{-1}\int_{\mathbb{R}} \sigma^2\; \partial_z\overline{f}\; \partial_z\overline{g}\; \exp(-\beta A),$$

which can be rewritten as

$$\frac{d}{dt}\int_{\mathbb{R}^n} \overline{f}(t,\xi(x))\,\overline{g}(\xi(x))\exp(-\beta V(x))\,dx = -\beta^{-1}\int_{\mathbb{R}^n} \nabla(\overline{f}\circ\xi)\cdot\nabla(\overline{g}\circ\xi)\exp(-\beta V). \tag{30}$$

The above weak formulation should be compared with the weak formulation of the Fokker-Planck equation (21) associated to (5):

$$\frac{d}{dt}\int_{\mathbb{R}^n} f\,g\,\exp(-\beta V) = -\beta^{-1}\int_{\mathbb{R}^n} \nabla f\cdot\nabla g\,\exp(-\beta V), \tag{31}$$

where $f = \psi\exp(\beta V)$, ψ is the probability distribution function of X_t satisfying (5), and $g : \mathbb{R}^n \to \mathbb{R}$ is a (time-independent) test function. We observe that (30) is (31) for functions which depend on x only through $\xi(x)$.

We now discuss the relation between the dynamics (26) that we propose and the dynamics (8). If the function ξ is such that $|\nabla\xi| = 1$, then $\sigma = 1$, and in view of (11), (12) and (24), we have $b = -A'$. Hence, in this case, the effective dynamics (26) is exactly (8). The fact that $|\nabla\xi| = 1$ is equivalent to say that ξ is the signed distance to the submanifold $\Sigma_0 = \{x;\ \xi(x) = 0\}$. Examples of such reaction coordinates include $\xi(x_1, \ldots, x_n) = x_1$, or $\xi(x) = |x|$.

More generally, assume that ξ is such that $\sigma = 1$. Then, in view of (28), we have $b = -A'$, and again (26) is exactly (8). Note however that, in general, σ is not a constant function, and (26) differs from (8). We will confirm in Section 4 that (26) and (8) may lead to significantly different numerical results.

Remark 2.3 *Note that* $\sigma = 1$ *writes*

$$\int_{\Sigma_z} \exp(-\beta V)\ |\nabla\xi|\ d\sigma_{\Sigma_z} = \int_{\Sigma_z} \exp(-\beta V)\ |\nabla\xi|^{-1}\ d\sigma_{\Sigma_z}.$$

Differentiating this equality with respect to z *yields (using again Lemma 2.2)*

$$\begin{aligned}&\int_{\Sigma_z} \left(-\nabla V \cdot \nabla\xi + \beta^{-1}\Delta\xi\right) \exp(-\beta V)\ |\nabla\xi|^{-1}\ d\sigma_{\Sigma_z} \\ &= -\int_{\Sigma_z} \left(\frac{\nabla V \cdot \nabla\xi}{|\nabla\xi|^2} - \beta^{-1}\mathrm{div}\left(\frac{\nabla\xi}{|\nabla\xi|^2}\right)\right)\ \exp(-\beta V)\ |\nabla\xi|^{-1}\ d\sigma_{\Sigma_z},\end{aligned}$$

which is exactly $b = -A'$.

Actually, using the fact that ξ is a *scalar* function, it is possible to recover the case $\sigma = 1$ (for which the effective dynamics is driven by the free energy) by two different methods. It is not clear to us whether such a reformulation is also possible in the case of a multi-dimensional reaction coordinate.

A first method is to introduce the following reindexation of the foliation $(\Sigma_z)_{z\in\mathbb{R}}$. We set

$$h(x) = \int_0^x \sigma^{-1}(y)\, dy$$

and we introduce the new reaction coordinate

$$\zeta = h \circ \xi.$$

Note that the foliation associated with ζ is exactly the same as the one associated with ξ since $h : \mathbb{R} \to \mathbb{R}$ is a one-to-one function. It is then easy to check that the coarse-grained dynamics associated with the reaction coordinate ζ is

$$dy_t = -\mathcal{A}'(y_t)\, dt + \sqrt{2\beta^{-1}}\, dB_t, \tag{32}$$

where $\mathcal{A}$ is the free energy associated to ζ. We hence obtain a dynamics of the type (8), with an appropriate noise (that is, dB_t in (32) and dW_t in (5) are linked by (16)).

Another possibility is to keep ξ as the reaction coordinate, and to consider, instead of (5), the dynamics

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

The measure μ is also invariant for this dynamics. Then, following the same coarse-graining procedure, based on the reaction coordinate ξ, one ends up with the coarse-grained dynamics

$$dy_t = -A'(y_t)\,dt + \sqrt{2\beta^{-1}}\;dB_t,$$

where A is the free energy associated to ξ. This is exactly (8), again with an appropriate noise.

3. Error estimation in terms of time marginals

In this section, we establish conditions on ξ under which the effective dynamics (26) is close to the dynamics of $\xi(X_t)$, from the time marginals viewpoint ([D3] in our above classification).

3.1. Error estimation

Let $\psi^\xi(t,z)$ be the probability distribution function of $\xi(X_t)$, where X_t follows (5), and $\phi(t,z)$ be the probability distribution function of the solution y_t to (26). Our aim is to bound the distance, for any time t, between these two one-dimensional probability measures.

We already introduced the total variation norm to measure distances between measures. In the case of *probability measures*, there are two other useful quantities. The first one is the relative entropy, which is defined by

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

for any two probability measures ν and η such that ν is absolutely continuous with respect to η. The relative entropy provides an upper-bound on the total variation norm distance, by the Csiszár-Kullback inequality:

$$\|\nu-\eta\|_{TV} \le \sqrt{2H\left(\nu|\eta\right)}. \tag{33}$$

The second one is the Wasserstein distance with quadratic cost, which is defined, for any two probability measures ν and η with support on a Riemannian manifold Σ, by

$$W(\nu,\eta) = \sqrt{\inf_{\pi\in\Pi(\nu,\eta)} \int_{\Sigma\times\Sigma} d_\Sigma(x,y)^2\;d\pi(x,y)}.$$

In the above expression, $d_\Sigma(x,y)$ denotes the geodesic distance between x and y on Σ,

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

and $\Pi(\nu,\eta)$ denotes the set of coupling probability measures, that is probability measures π on $\Sigma\times\Sigma$ such that their marginals are ν and η: for any test function Φ,

$$\int_{\Sigma\times\Sigma}\Phi(x)\,d\pi(x,y) = \int_\Sigma \Phi(x)\,d\nu(x) \quad\text{and}\quad \int_{\Sigma\times\Sigma}\Phi(y)\,d\pi(x,y) = \int_\Sigma \Phi(y)\,d\eta(y).$$

In the sequel, we will need two functional inequalities, that we now recall [1]:

Definition 3.1 *A probability measure η satisfies a logarithmic Sobolev inequality with a constant $\rho > 0$ if, for any probability measure ν,*

$$H(\nu|\eta) \leq \frac{1}{2\rho} I(\nu|\eta)$$

where the Fisher information $I(\nu|\eta)$ is defined by

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

Definition 3.2 *A probability measure η satisfies a Talagrand inequality with a constant $\rho > 0$ if, for any probability measure ν,*

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

We will also need the following important result (see [25, Theorem 1] and [4]):

Lemma 3.1 *If η satisfies a logarithmic Sobolev inequality with a constant $\rho > 0$, then η satisfies a Talagrand inequality with the same constant $\rho > 0$.*

Logarithmic Sobolev inequalities are very useful to prove properties concerning the longtime behaviour of solutions to PDEs (e.g. long time convergence of the solution of a Fokker-Planck equation to the stationary measure of the corresponding SDE). We refer to [1, 2, 35] for more details on this subject.

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

Proposition 3.1 *Assume that ξ satisfies* **[H1]***, and that the conditioned probability measures μ_{Σ_z}, defined by (9), satisfy a logarithmic Sobolev inequality with a constant ρ uniform in z: for any probability measure ν on Σ_z which is absolutely continuous with respect to the measure μ_{Σ_z}, we have*

$$\textbf{[H2]} \quad H(\nu|\mu_{\Sigma_z}) \leq \frac{1}{2\rho} I(\nu|\mu_{\Sigma_z}).$$

Let us also assume that the coupling is bounded in the following sense:

$$\textbf{[H3]} \quad \kappa = \|\nabla_{\Sigma_z} F\|_{L^\infty} < \infty,$$

where F is the local mean force defined by (12).

Finally, let us assume that $|\nabla \xi|$ is close to a constant on the manifold Σ_z in the following sense:

$$\textbf{[H4]} \quad \lambda = \left\| \frac{|\nabla \xi|^2 - \sigma^2 \circ \xi}{\sigma^2 \circ \xi} \right\|_{L^\infty} < \infty.$$

Assume that, at time $t = 0$, the distribution of the initial conditions of (5) and (26) are consistent one with each other: $\psi^\xi(t=0,\cdot) = \phi(t=0,\cdot)$. Then we have the following estimate: for any time $t \geq 0$,

$$E(t) \leq \frac{M^2}{4m^2}\left(\lambda^2 + \frac{m^2\beta^2\kappa^2}{\rho^2}\right)\left(H(\psi(0,\cdot)|\mu) - H(\psi(t,\cdot)|\mu)\right), \tag{34}$$

where $E(t)$ is the relative entropy of the probability distribution function ψ^ξ of $\xi(X_t)$, where X_t follows (5), with respect to the probability distribution function ϕ of the solution y_t to (26):

$$E(t) = H\left(\psi^\xi(t,\cdot)|\phi(t,\cdot)\right) = \int_{\mathbb{R}} \ln\left(\frac{\psi^\xi(t,z)}{\phi(t,z)}\right)\psi^\xi(t,z)\,dz.$$

Let us comment on these three assumptions. Assumption **[H2]** means that μ_{Σ_z}, which is a measure on the manifold Σ_z, is easy to sample from. In view of (34), the interesting case is when ρ is large, and then assumption **[H2]** implies that there is no metastability in the manifold Σ_z. This amounts to assuming that the overdamped dynamics with respect to μ_{Σ_z} (which lives on Σ_z) is well-mixing. Note finally that, in view of (13), the relative entropy $H(\nu|\mu_{\Sigma_z})$ and the Fisher information $I(\nu|\mu_{\Sigma_z})$ entering assumption **[H2]** read

$$H(\nu|\mu_{\Sigma_z}) = \int_{\Sigma_z} \ln\left(f \Big/ \frac{Z^{-1}\exp(-\beta V)}{\exp(-\beta A(z))}\right)\ f\ |\nabla\xi|^{-1} d\sigma_{\Sigma_z}$$

and

$$I(\nu|\mu_{\Sigma_z}) = \int_{\Sigma_z} \left|\nabla_{\Sigma_z} \ln\left(\frac{f}{\exp(-\beta V)}\right)\right|^2\ f\ |\nabla\xi|^{-1} d\sigma_{\Sigma_z},$$

where f is the density of ν with respect to the measure $|\nabla\xi|^{-1}\sigma_{\Sigma_z}$, *i.e.* $f = \dfrac{d\nu}{|\nabla\xi|^{-1}d\sigma_{\Sigma_z}}$, and ∇_{Σ_z} denotes the surface gradient:

$$\nabla_{\Sigma_z} = P\nabla, \quad \text{where} \quad P(x) = \mathrm{Id} - \frac{\nabla\xi \otimes \nabla\xi}{|\nabla\xi|^2}(x)$$

is the orthogonal projector on the tangent space to Σ_z at point $x \in \Sigma_z$.

We now turn to assumption **[H3]**. Consider first the case when $x = (x_1, x_2) \in \mathbb{R}^2$, and $\xi(x) = x_1$. Then $F = \nabla_{x_1} V$ and $\nabla_{\Sigma_z} F = \nabla_{x_2} F = \nabla_{x_1 x_2} V$. Requesting that κ is small hence amounts to requesting that $\nabla_{x_1 x_2} V$ is small, where x_1 is the reaction coordinate direction whereas x_2 is the direction in Σ_z. We hence ask for the coupling of these two directions to be small. In particular, in the case when $V(x) = \frac{1}{2}x^T H x$ for some symmetric positive matrix $H \in \mathbb{R}^{n\times n}$ and $\xi(x) = \xi(x_1,\ldots,x_n) = (x_1,\ldots,x_p)$ for some $p \leq n$, we have that $\nabla_{\Sigma_z} F = 0$ if and only if the covariance $\mathrm{Cov}_\mu\left((X_1,\ldots,X_p),(X_{p+1},\ldots,X_n)\right) = 0$, where $X \in \mathbb{R}^n$ is distributed according to $d\mu = Z^{-1}\exp(-\beta V(x))\,dx$. Hence **[H3]** means that the variables $(X_1,\ldots,X_p)$, which represent the reaction coordinate directions, are decoupled from the variables $(X_{p+1},\ldots,X_n)$, which represent the directions of Σ_z.

In Section 3.2, we will consider an explicit example, and compute an estimation of ρ and κ in that case, which will help understanding the assumptions **[H2]** and **[H3]**.

The assumption **[H4]** is technical. Observe that, if $|\nabla \xi|$ is a constant number in each manifold Σ_z, then $\lambda = 0$.

Before proving Proposition 3.1, let us comment on the estimate (34). Note first that this estimate is uniform in time. The initial conditions for (26) and (5) are such that $\phi(t = 0, \cdot) = \psi^\xi(t = 0, \cdot)$, which explains that $E(t = 0) = 0$. In the longtime limit, the estimate (34) is not optimal since we know that both ϕ and ψ^ξ converge to $\xi \star \mu$ (see Lemma 2.4). This implies that $\lim_{t\to\infty} E(t) = 0$, a property that we prove in Corollary 3.1 below.

To prove Proposition 3.1, we will need the following lemma:

Lemma 3.2 *Let $\psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ be the probability distribution function of X_t that solves (5). The probability distribution function of $\xi(X_t)$ is $\psi^\xi(t, z) = \int_{\Sigma_z} \psi(t, \cdot)|\nabla \xi|^{-1} d\sigma_{\Sigma_z}$, and satisfies*

$$\exp(-\beta A)\, \partial_z(\psi^\xi \exp(\beta A)) = \int \frac{\nabla(\psi \exp(\beta V)) \cdot \nabla \xi}{|\nabla \xi|^2} \exp(-\beta V)\ |\nabla \xi|^{-1}\ d\sigma_{\Sigma_z} + \beta \left(A'(z) - \frac{\int F\, \psi\, |\nabla \xi|^{-1}\, d\sigma_{\Sigma_z}}{\psi^\xi} \right) \psi^\xi, \tag{35}$$

where A is the free energy (10) and F is the local mean force (12).

Proof: Using (22), we compute

$$\begin{aligned}
&\exp(-\beta A)\, \partial_z(\psi^\xi \exp(\beta A)) \\
&\quad = \partial_z \psi^\xi + \beta\, A'\, \psi^\xi, \\
&\quad = \int_{\Sigma_z} \left(\frac{\nabla \xi \cdot \nabla \psi}{|\nabla \xi|^2} + \mathrm{div}\, \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \psi \right)\ |\nabla \xi|^{-1}\ d\sigma_{\Sigma_z} + \beta\, A'\, \psi^\xi, \\
&\quad = \int_{\Sigma_z} \frac{\nabla \xi \cdot \nabla(\psi \exp(\beta V))}{|\nabla \xi|^2} \exp(-\beta V)\ |\nabla \xi|^{-1}\ d\sigma_{\Sigma_z} \\
&\qquad + \int_{\Sigma_z} \left(\mathrm{div}\, \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) - \beta \frac{\nabla \xi \cdot \nabla V}{|\nabla \xi|^2} \right)\ \psi\ |\nabla \xi|^{-1}\ d\sigma_{\Sigma_z} + \beta\, A'\, \psi^\xi,
\end{aligned}$$

which yields (35). □

We are now in position to prove Proposition 3.1.

Proof: We know that ϕ satisfies the Fokker-Planck equation (29), and that ψ^ξ satisfies the equation (20). Thus, we have:

$$\begin{aligned}
\frac{dE}{dt} &= \int \partial_t \psi^\xi\ \ln \left(\frac{\psi^\xi}{\phi} \right) - \int \partial_t \phi\ \frac{\psi^\xi}{\phi}, \\
&= \int \partial_z \left(-\widetilde{b}\ \psi^\xi + \beta^{-1} \partial_z(\widetilde{\sigma}^2\ \psi^\xi) \right)\ \ln \left(\frac{\psi^\xi}{\phi} \right) \\
&\quad - \beta^{-1} \int \partial_z \left[\sigma^2\, \partial_z(\phi \exp(\beta A)) \exp(-\beta A) \right]\ \frac{\psi^\xi}{\phi},
\end{aligned}$$

$$\begin{aligned}
&= -\int \left(-\widetilde{b}\;\psi^\xi + \beta^{-1}\partial_z(\widetilde{\sigma}^2\psi^\xi)\right)\partial_z \ln\left(\frac{\psi^\xi}{\phi}\right)\\
&\quad + \beta^{-1}\int \sigma^2\;\partial_z(\phi\exp(\beta A))\exp(-\beta A)\;\partial_z\left(\frac{\psi^\xi}{\phi}\right).
\end{aligned}$$

Using (23), we have:

$$\begin{aligned}
\partial_z(\widetilde{\sigma}^2\,\psi^\xi) &= \int_{\Sigma_z}(\nabla\xi\cdot\nabla\psi+\psi\Delta\xi)\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\\
&= \int_{\Sigma_z}(\nabla\xi\cdot\nabla(\psi\exp(\beta V))\exp(-\beta V))\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\\
&\quad + \int_{\Sigma_z}(-\beta\nabla\xi\cdot\nabla V+\Delta\xi)\;\psi\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\\
&= \int_{\Sigma_z}(\nabla\xi\cdot\nabla(\psi\exp(\beta V))\exp(-\beta V))\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\\
&\quad + \beta\,\widetilde{b}(t,z)\,\psi^\xi(t,z).
\end{aligned}$$

Thus, it holds:

$$\begin{aligned}
\frac{dE}{dt} &= -\beta^{-1}\int\int_{\Sigma_z}(\nabla\xi\cdot\nabla(\psi\exp(\beta V))\exp(-\beta V))\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\;\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)\\
&\quad +\beta^{-1}\int\sigma^2\;\partial_z(\phi\exp(\beta A))\exp(-\beta A)\;\partial_z\left(\frac{\psi^\xi}{\phi}\right),\\
&= -\beta^{-1}\int\int_{\Sigma_z}\left(\frac{\nabla\xi\cdot\nabla(\psi\exp(\beta V))}{|\nabla\xi|^2}\;\exp(-\beta V)\right)\left(|\nabla\xi|^2-\sigma^2(z)\right)\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\;\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)\\
&\quad -\beta^{-1}\int\sigma^2(z)\int_{\Sigma_z}\left(\frac{\nabla\xi\cdot\nabla(\psi\exp(\beta V))}{|\nabla\xi|^2}\;\exp(-\beta V)\right)\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\;\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)\\
&\quad +\beta^{-1}\int\sigma^2\;\partial_z(\phi\exp(\beta A))\exp(-\beta A)\;\partial_z\left(\frac{\psi^\xi}{\phi}\right).
\end{aligned}$$

We next use (35) to get:

$$\begin{aligned}
\frac{dE}{dt} &= -\beta^{-1}\int\int_{\Sigma_z}\left(\frac{\nabla\xi\cdot\nabla(\psi\exp(\beta V))}{|\nabla\xi|^2}\;\exp(-\beta V)\right)\left(|\nabla\xi|^2-\sigma^2(z)\right)\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\;\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)\\
&\quad -\beta^{-1}\int\sigma^2\left[(\exp(-\beta A))\,\partial_z(\psi^\xi\exp(\beta A))-\beta\left(A'(z)-\frac{\int F\psi|\nabla\xi|^{-1}d\sigma_{\Sigma_z}}{\psi^\xi}\right)\psi^\xi\right]\;\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)\\
&\quad +\beta^{-1}\int\sigma^2\;\partial_z(\phi\exp(\beta A))\;\exp(-\beta A)\;\partial_z\left(\frac{\psi^\xi}{\phi}\right)\\
&= -\beta^{-1}\int\int_{\Sigma_z}\left(\frac{\nabla\xi\cdot\nabla(\psi\exp(\beta V))}{|\nabla\xi|^2}\;\exp(-\beta V)\right)\;\left(|\nabla\xi|^2-\sigma^2(z)\right)\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\;\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)\\
&\quad +\int\sigma^2\left(A'(z)-\frac{\int F\psi|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}{\psi^\xi}\right)\;\psi^\xi\;\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)\\
&\quad +\beta^{-1}\int\sigma^2\;\exp(-\beta A)\;\partial_z\left(\frac{\psi^\xi}{\phi}\right)\left[\partial_z(\phi\exp(\beta A))-\partial_z(\psi^\xi\exp(\beta A))\left(\frac{\phi}{\psi^\xi}\right)\right],\\
&= -\beta^{-1}\int\int_{\Sigma_z}\left(\frac{\nabla\xi\cdot\nabla(\psi\exp(\beta V))}{|\nabla\xi|^2}\;\exp(-\beta V)\right)\;\left(|\nabla\xi|^2-\sigma^2(z)\right)\;|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}\;\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)\\
&\quad +\int\sigma^2\left(A'(z)-\frac{\int F\psi|\nabla\xi|^{-1}d\sigma_{\Sigma_z}}{\psi^\xi}\right)\;\psi^\xi\;\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)-\beta^{-1}\int\sigma^2\;\psi^\xi\;\left|\partial_z\ln\left(\frac{\psi^\xi}{\phi}\right)\right|^2.
\end{aligned}$$

We now use two Young inequalities, with $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ to be fixed later on:

$$\begin{aligned}\frac{dE}{dt} \quad \leq \quad & \frac{\beta^{-1}}{2\varepsilon_1} \int \left| \int_{\Sigma_z} \left(\frac{\nabla\xi \cdot \nabla(\psi \exp(\beta V))}{|\nabla\xi|^2} \ \exp(-\beta V) \right) \ \left(|\nabla\xi|^2 - \sigma^2(z)\right) \ |\nabla\xi|^{-1} \ d\sigma_{\Sigma_z} \right|^2 \frac{1}{\sigma^2\psi^\xi} \\ & + \frac{\beta}{2\varepsilon_2} \int \sigma^2 \left(A'(z) - \frac{\int F\psi |\nabla\xi|^{-1} d\sigma_{\Sigma_z}}{\psi^\xi} \right)^2 \ \psi^\xi \qquad (36) \\ & - \beta^{-1} \left(1 - \frac{\varepsilon_1 + \varepsilon_2}{2}\right) \int \sigma^2 \, \psi^\xi \left| \partial_z \ln\left(\frac{\psi^\xi}{\phi}\right)\right|^2 .\end{aligned}$$

Let us first consider the second term of (36). We write, using [**H3**], that

$$\begin{aligned}\left(A'(z) - \frac{\int_{\Sigma_z} F\,\psi\,|\nabla\xi|^{-1} \ d\sigma_{\Sigma_z}}{\psi^\xi} \right)^2 &= \left(\int_{\Sigma_z} F \ d\mu_{\Sigma_z} - \int_{\Sigma_z} F \ d\psi_{\Sigma_z} \right)^2 \\ &\leq \|\nabla_{\Sigma_z} F\|^2_{L^\infty} \ W(d\psi_{\Sigma_z}, d\mu_{\Sigma_z})^2, \qquad (37)\end{aligned}$$

where ψ_{Σ_z} is the measure $\psi(t,x)\,dx$ conditioned to $\xi(x) = z$:

$$d\psi_{\Sigma_z} = \frac{\psi\,|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}{\psi^\xi}.$$

Since μ_{Σ_z} satisfies a logarithmic Sobolev inequality (assumption [**H2**]), it also satisfies a Talagrand inequality (see Lemma 3.1), hence

$$W(d\psi_{\Sigma_z}, d\mu_{\Sigma_z})^2 \leq \frac{2}{\rho} H(d\psi_{\Sigma_z}|d\mu_{\Sigma_z}) \leq \frac{1}{\rho^2} I(d\psi_{\Sigma_z}|d\mu_{\Sigma_z}).$$

Gathering the above inequality with (37), we obtain

$$\left(A'(z) - \frac{\int_{\Sigma_z} F\,\psi\,|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}{\psi^\xi} \right)^2 \leq \frac{\kappa^2}{\rho^2} \ I\left(d\psi_{\Sigma_z}|d\mu_{\Sigma_z}\right).$$

Using [**H1**], we thus bound the second term of (36):

$$\begin{aligned}\int_{\mathbb{R}} \sigma^2 & \left(A'(z) - \frac{\int_{\Sigma_z} F\,\psi\,|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}{\psi^\xi} \right)^2 \psi^\xi \\ & \leq \frac{M^2\kappa^2}{\rho^2} \int_{\mathbb{R}} I\left(d\psi_{\Sigma_z}|d\mu_{\Sigma_z}\right) \ \psi^\xi, \\ & = \frac{M^2\kappa^2}{\rho^2} \int_{\mathbb{R}} \int_{\Sigma_z} \left| \nabla_{\Sigma_z} \ln\left(\frac{\psi}{\exp(-\beta V)}\right)\right|^2 \ \psi\,|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}, \\ & = \frac{M^2\kappa^2}{\rho^2} \int_{\mathbb{R}^n} \left| \nabla_{\Sigma_z} \ln\left(\frac{\psi}{\exp(-\beta V)}\right)\right|^2 \psi. \qquad (38)\end{aligned}$$

We now bound the first term of (36) using a Cauchy-Schwarz inequality, [**H4**] and [**H1**]:

$$\begin{aligned}\int & \left| \int_{\Sigma_z} \left(\frac{\nabla\xi \cdot \nabla(\psi\exp(\beta V))}{|\nabla\xi|^2} \ \exp(-\beta V) \right) \ \left(|\nabla\xi|^2 - \sigma^2(z)\right) \ |\nabla\xi|^{-1} \ d\sigma_{\Sigma_z} \right|^2 \frac{1}{\sigma^2\psi^\xi} \\ & = \int \left| \int_{\Sigma_z} \frac{\nabla\xi \cdot \nabla\ln(\psi\exp(\beta V))}{|\nabla\xi|^2} \ \left(|\nabla\xi|^2 - \sigma^2(z)\right) \psi \ |\nabla\xi|^{-1} \ d\sigma_{\Sigma_z} \right|^2 \frac{1}{\sigma^2\psi^\xi},\end{aligned}$$

$$\begin{aligned}
&\le \int\int_{\Sigma_z} \left| \frac{\nabla\xi \cdot \nabla \ln(\psi\exp(\beta V))}{|\nabla\xi|^2} \ \left(|\nabla\xi|^2 - \sigma^2(z)\right) \right|^2 \ \psi \ |\nabla\xi|^{-1} \ d\sigma_{\Sigma_z} \frac{1}{\sigma^2}, \\
&\le \lambda^2 \int\int_{\Sigma_z} \left| \frac{\nabla\xi \cdot \nabla \ln(\psi\exp(\beta V))}{|\nabla\xi|^2} \right|^2 \ \psi \ |\nabla\xi|^{-1} \ d\sigma_{\Sigma_z} \ \sigma^2, \\
&\le \lambda^2 M^2 \int_{\mathbb{R}^n} \left| \frac{\nabla\xi \cdot \nabla \ln(\psi\exp(\beta V))}{|\nabla\xi|^2} \right|^2 \psi. \qquad (39)
\end{aligned}$$

We infer from (36) and the bounds (38) and (39) that

$$\begin{aligned}
\frac{dE}{dt} \le & \frac{\beta^{-1}}{2\varepsilon_1}\lambda^2 M^2 \int_{\mathbb{R}^n} \left| \frac{\nabla\xi \cdot \nabla \ln(\psi\exp(\beta V))}{|\nabla\xi|^2} \right|^2 \psi \\
& + \frac{\beta}{2\varepsilon_2}\frac{M^2\kappa^2}{\rho^2} \int_{\mathbb{R}^n} |\nabla_{\Sigma_z} \ln\left(\psi\exp(\beta V)\right)|^2 \psi \\
& - \beta^{-1}\left(1 - \frac{\varepsilon_1+\varepsilon_2}{2}\right) \int \sigma^2 \, \psi^\xi \left| \partial_z \ln\left(\frac{\psi^\xi}{\phi}\right)\right|^2.
\end{aligned}$$

Note that

$$|\nabla \ln\left(\psi \ \exp(\beta V)\right)|^2 = \left| \frac{\nabla\xi \cdot \nabla \ln(\psi\exp(\beta V))}{|\nabla\xi|} \right|^2 + |\nabla_{\Sigma_z} \ln\left(\psi\exp(\beta V)\right)|^2.$$

Using the lower bound on $|\nabla\xi|$ given by **[H1]**, we hence obtain

$$\begin{aligned}
\frac{dE}{dt} \le & \frac{\beta^{-1}}{2\varepsilon_1}\frac{\lambda^2 M^2}{m^2} \int_{\mathbb{R}^n} |\nabla \ln(\psi\exp(\beta V))|^2 \psi + \frac{\beta}{2\varepsilon_2}\frac{M^2\kappa^2}{\rho^2} \int_{\mathbb{R}^n} |\nabla \ln\left(\psi\exp(\beta V)\right)|^2 \psi \\
& -\beta^{-1}\left(1 - \frac{\varepsilon_1+\varepsilon_2}{2}\right) \int \sigma^2 \, \psi^\xi \left| \partial_z \ln\left(\frac{\psi^\xi}{\phi}\right)\right|^2.
\end{aligned}$$

We now optimize on ε_1 and ε_2 by choosing them such that $\varepsilon_1 + \varepsilon_2 = 2$ and $\dfrac{\beta^{-1}}{2\varepsilon_1}\dfrac{\lambda^2 M^2}{m^2} = \dfrac{\beta}{2\varepsilon_2}\dfrac{M^2\kappa^2}{\rho^2}$. This yields $\varepsilon_1 = \dfrac{2\lambda^2\rho^2}{\lambda^2\rho^2 + m^2\beta^2\kappa^2}$, thus

$$\begin{aligned}
\frac{dE}{dt} &\le \frac{\beta^{-1}M^2}{4m^2}\left(\lambda^2 + \frac{m^2\beta^2\kappa^2}{\rho^2}\right) \int_{\mathbb{R}^n} |\nabla \ln(\psi\exp(\beta V))|^2 \psi, \\
&= \frac{\beta^{-1}M^2}{4m^2}\left(\lambda^2 + \frac{m^2\beta^2\kappa^2}{\rho^2}\right) I(\psi|\mu), \\
&= -\frac{M^2}{4m^2}\left(\lambda^2 + \frac{m^2\beta^2\kappa^2}{\rho^2}\right) \frac{d}{dt} H(\psi|\mu).
\end{aligned}$$

We next integrate this equation between 0 and t and use the fact that $E(0) = 0$ to obtain (34). □

We now prove a corollary of Proposition 3.1, which strengthens its long-time limit behaviour.

Corollary 3.1 *In addition to the assumptions of Proposition 3.1, assume that*

[H5] *The measure $\xi \star \mu$ satisfies a logarithmic Sobolev inequality with a constant r.*

Consider again the probability distribution function ψ^ξ of $\xi(X_t)$, where X_t follows (5), and the probability distribution function ϕ of the solution y_t to (26). They satisfy:

$$\forall t \geq 0, \quad \|\psi^\xi(t,\cdot) - \phi(t,\cdot)\|_{TV} \leq \min\left(C_1(t), 2C_2 \exp(-R\,\beta^{-1}\,t)\right), \tag{40}$$

for some positive constant R, where

$$C_1(t) = \sqrt{\frac{M^2}{2m^2}\left(\lambda^2 + \frac{m^2\beta^2\kappa^2}{\rho^2}\right)(H(\psi(0,\cdot)|\mu) - H(\psi(t,\cdot)|\mu))}, \tag{41}$$

$$C_2 \quad = \max\left(\sqrt{2H\left(\phi(0,\cdot)|\mu^\xi\right)}, \sqrt{2H\left(\psi(0,\cdot)|\mu\right)}\right), \tag{42}$$

with $d\mu^\xi = \exp(-\beta A(z))\,dz$.

As a consequence of this corollary, we see that $\lim_{t\to\infty} \|\psi^\xi(t,\cdot) - \phi(t,\cdot)\|_{TV} = 0$.

Proof: We infer from the Csiszár-Kullback inequality and from the bound (34) that

$$\|\psi^\xi - \phi\|_{TV} \leq \sqrt{2H\left(\psi^\xi|\phi\right)} \leq C_1(t), \tag{43}$$

where $C_1(t)$ is given by (41). We also have

$$\|\psi^\xi - \phi\|_{TV} \leq \left\|\psi^\xi - \mu^\xi\right\|_{TV} + \left\|\phi - \mu^\xi\right\|_{TV}, \tag{44}$$

where $d\mu^\xi = \exp(-\beta A(z))\,dz$ is the equilibrium measure $\xi \star \mu$. Let us first upper-bound $E_{\rm CG}(t) = H\left(\phi|\mu^\xi\right) = \displaystyle\int_{\mathbb{R}} \ln\left(\frac{\phi}{\exp(-\beta A)}\right)\phi$. Using (29), we compute

$$\begin{aligned}
\frac{dE_{\rm CG}}{dt} &= \int_{\mathbb{R}} \partial_t \phi \;\ln\left(\frac{\phi}{\exp(-\beta A)}\right) \\
&= \beta^{-1} \int_{\mathbb{R}} \partial_z\left[\sigma^2\; \partial_z(\phi \exp(\beta A))\; \exp(-\beta A)\right]\; \ln\left(\frac{\phi}{\exp(-\beta A)}\right) \\
&= -\,\beta^{-1} \int_{\mathbb{R}} \left[\sigma^2\; \partial_z(\phi \exp(\beta A))\; \exp(-\beta A)\right]\; \partial_z\left[\ln\left(\frac{\phi}{\exp(-\beta A)}\right)\right] \\
&\leq -\,m^2\beta^{-1} \int_{\mathbb{R}} \phi\, \left|\partial_z\left[\ln\left(\frac{\phi}{\exp(-\beta A)}\right)\right]\right|^2 \\
&= -\,m^2\beta^{-1} I(\phi|\mu^\xi),
\end{aligned}$$

where we have used that $\sigma^2 \geq m^2$, which is a consequence of **[H1]** and (25). Since μ^ξ satisfies a logarithmic Sobolev inequality with constant r, we infer from the above bound that $\dfrac{dE_{\rm CG}}{dt} \leq -2\,r\,m^2\,\beta^{-1} E_{\rm CG}$. Using a Gronwall lemma, we obtain

$$H\left(\phi|\mu^\xi\right) = E_{\rm CG}(t) \leq E_{\rm CG}(t=0)\,\exp(-2\,r\,m^2\,\beta^{-1}\,t) = H\left(\phi(0,\cdot)|\mu^\xi\right)\,\exp(-2\,r\,m^2\,\beta^{-1}\,t),$$

and the Csiszár-Kullback inequality then yields

$$\left\|\phi - \mu^\xi\right\|_{TV} \leq \sqrt{2H\left(\phi|\mu^\xi\right)} \leq \sqrt{2H\left(\phi(0,\cdot)|\mu^\xi\right)}\,\exp(-r\,m^2\,\beta^{-1}\,t). \tag{45}$$

We now turn to the term $\left\|\psi^\xi - \mu^\xi\right\|_{TV}$. For any function $\chi : \mathbb{R}^n \to \mathbb{R}$, define $\chi^\xi(z) = \displaystyle\int_{\Sigma_z} \chi \, |\nabla \xi|^{-1} \, d\sigma_{\Sigma_z}$, and observe that

$$\int_{\mathbb{R}^n} |\chi(x)| \, dx = \int_{\mathbb{R}} \int_{\Sigma_z} \frac{|\chi|}{|\nabla \xi|} \, d\sigma_{\Sigma_z} \, dz \geq \int_{\mathbb{R}} \left| \int_{\Sigma_z} \frac{\chi}{|\nabla \xi|} \, d\sigma_{\Sigma_z} \right| dz = \int_{\mathbb{R}} \left|\chi^\xi\right| \, dz$$

which also reads $\|\chi\|_{TV} \geq \left\|\chi^\xi\right\|_{TV}$. We apply this inequality with $\chi = \psi - \mu$:

$$\left\|\psi^\xi - \mu^\xi\right\|_{TV} \leq \|\psi - \mu\|_{TV} \leq \sqrt{2 H\left(\psi|\mu\right)}.$$

Since μ^ξ and the conditional measures μ_{Σ_z} satisfy a logarithmic Sobolev inequality (see [**H5**] and [**H2**]), and under assumption [**H3**], we obtain that the measure μ also satisfies a logarithmic Sobolev inequality with some constant $R > 0$ (see [23]). Hence, by a computation similar to the one on $E_{\rm CG}$, we obtain

$$H\left(\psi|\mu\right) \leq H\left(\psi(0,\cdot)|\mu\right) \exp(-2\, R \, \beta^{-1} \, t),$$

hence

$$\left\|\psi^\xi - \mu^\xi\right\|_{TV} \leq \sqrt{2 H\left(\psi(0,\cdot)|\mu\right)} \, \exp(-R \, \beta^{-1} \, t). \tag{46}$$

Gathering (44), (45) and (46), we obtain

$$\|\psi^\xi - \phi\|_{TV} \leq C_2 \, \exp(-r \, m^2 \, \beta^{-1} \, t) + C_2 \, \exp(-R \, \beta^{-1} \, t),$$

where C_2 is defined by (42). The proof of [23, Theorem 1.2] shows that $0 < R \leq r m^2$. The above bound then yields $\|\psi^\xi - \phi\|_{TV} \leq 2 C_2 \, \exp(-R \, \beta^{-1} \, t)$, which, gathered with (43), yields (40). □

3.2. Estimation of the upper-bound constants of (34) in a particular case

In this section, we give a very formal argument to estimate the constants ρ and κ entering the bound (34), in a specific case. Potential energies in molecular dynamics are often the sum of several terms, with different stiffness. For instance, the potential energy of an alkane chain, in the United Atom model [29], reads

$$V(X) = \sum_i V_2(d_{i,i+1}) + \sum_i V_3(\theta_i) + \sum_i V_4(\phi_i) + V_{\rm non-bonded}(X),$$

where $d_{i,i+1}$ is the distance between atoms i and $i+1$, θ_i is the bond angle made by atoms $i-1$, i and $i+1$, whereas ϕ_i is the dihedral angle defined by the atoms $i+j$, $j = -1, \ldots, 2$. In general, V_2 is a much stiffer potential than V_3, which is itself much stiffer than V_4.

A simple toy-model for such potential energies is

$$V_\varepsilon(X) = V_0(X) + \frac{1}{\varepsilon} q^2(X), \tag{47}$$

where V_0 and q are two scalar-valued functions that do not depend on the small parameter ε (see Equation (49) and Figure 1 below for a precise example of type (47)). For simplicity, we assume here that the reaction coordinate ξ does not depend on ε, and that it is constant on the manifolds Σ_z (in assumption **[H4]**, $\lambda = 0$). Since the relative entropy is always non-negative, the estimate (34) reads

$$E(t) \leq \frac{M^2}{4} \frac{\beta^2 \kappa_\varepsilon^2}{\rho_\varepsilon^2} \; H(\psi(0,\cdot)|\mu_\varepsilon).$$

We also assume that the initial condition of (5) is well adapted to the Boltzmann measure μ_ε, in the sense that $H(\psi(0,\cdot)|\mu_\varepsilon)$ is upper-bounded by a constant independent of ε. Thus the above bound reads

$$E(t) \leq C \frac{\kappa_\varepsilon^2}{\rho_\varepsilon^2}$$

for some constant C independent of ε. Our aim is to roughly estimate the coefficients ρ_ε and κ_ε in terms of ε.

Since ε is small, the Boltzmann measure (2) concentrates on the manifold where $q = 0$, and locally looks like a Gaussian measure of variance ε around that manifold. The same holds for μ_{Σ_z}, that is assumed to satisfy a logarithmic Sobolev inequality (assumption **[H2]**). Hence, we typically have $\rho_\varepsilon = O\left(1/\varepsilon\right)$.

We now compute the local mean force, defined by (12):

$$\begin{aligned} F &= \frac{\nabla V_\varepsilon \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \\ &= \frac{2}{\varepsilon} \, q \, \frac{\nabla q \cdot \nabla \xi}{|\nabla \xi|^2} \; + \; \frac{\nabla V_0 \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right). \end{aligned}$$

Recall that ξ does not depend on ε. If $\nabla q \cdot \nabla \xi \neq 0$, then F is of order $O\left(1/\varepsilon\right)$, and so is κ_ε. On the contrary, if $\nabla q \cdot \nabla \xi = 0$, then F is of order $O(1)$ with respect to ε, and so is κ_ε.

Let us summarize our discussion. In the case when $\nabla q \cdot \nabla \xi = 0$, it turns out that ρ_ε is of order $1/\varepsilon$, while κ_ε is of order 1, and the estimate (34) reads

$$E(t) \leq C \varepsilon^2$$

for some constant C that does not depend on ε. Hence, as ε decreases to 0, the effective dynamics (26) becomes more accurate, in the sense of [D3]. In the case when $\nabla q \cdot \nabla \xi \neq 0$, both ρ_ε and κ_ε are of order $1/\varepsilon$, and the estimate (34) reads

$$E(t) \leq C$$

for some constant C that does not depend on ε. So the effective dynamics (26) is not particularly accurate. In the next section, we numerically confirm that the criterion

$$\nabla \xi \cdot \nabla q = 0 \tag{48}$$

has indeed a significant impact on the accuracy of the effective dynamics.

4. Numerical results: residence time estimation

Our aim here is twofold. First, we want to check the accuracy of (26) in a sense related to [D2], on a simple system, and also compare this effective dynamics with the coarse-grained dynamics (8) based on the free energy. Second, we wish to assess the relevance of the criterion (48). It seems to be an important condition for estimates in the sense of [D3] to be meaningful. Is it also a necessary and sufficient condition in order to obtain accurate dynamical properties ?

In the following numerical tests, we focus on the residence times. We have indeed already underlined that the characteristic behaviour of the dynamics (5) is to sample a given well of the potential energy, then suddenly hopes to another basin, and start over. Consequently, an important quantity is the residence time that the system spends in the well, before going to another one. In this section, we describe a numerical example where we have studied such quantities, which contain dynamical information, and are related to the estimator [D2].

Consider the two-dimensional potential energy

$$V_\varepsilon(x,y) = (x^2-1)^2 + \frac{1}{\varepsilon}(x^2+y-1)^2 \tag{49}$$

which is of the form (47), with $V_0(x,y) = (x^2-1)^2$ and $q(x,y) = x^2+y-1$. For any $\varepsilon > 0$, the potential V_ε has two local minima, at $(x,y) = (\pm 1, 0)$, and one saddle point, at $(x,y) = (0,1)$ (see Figure 1). There are thus two basins, namely $\{(x,y) \in \mathbb{R}^2;\ x < 0\}$ and $\{(x,y) \in \mathbb{R}^2;\ x > 0\}$. Since V_ε is an even function of x, the residence times in each well are equal to each other. Our aim is to compare the residence time computed when the full description of the system is used (that is, we simulate the dynamics (5)) with the residence time computed from a coarse-grained description, according to (26) or (8), for two different reaction coordinates.

In the case at hand, a natural reaction coordinate is $\xi_1(x,y) = x$, since the value of ξ_1 already gives the information that the system is in the right or the left well. In that case, $|\nabla \xi_1| = 1$, hence the effective dynamics (26) is the same as the dynamics (8), that is the dynamics driven by the free energy A_1 associated to ξ_1. This free energy reads

$$A_1(z) = (z^2-1)^2 + C(\beta) \tag{50}$$

for some constant $C(\beta)$ ensuring that $\displaystyle\int_{\mathbb{R}} \exp(-\beta A_1(z))\, dz = 1$.

Note that $\nabla \xi_1 \cdot \nabla q \neq 0$. In view of the discussion of the previous section, we do not expect the effective dynamics based on ξ_1 to be very accurate.

Consider now the function $\xi_2(x,y) = x \exp(-2y)$, which satisfies $\nabla \xi_2 \cdot \nabla q = 0$. We expect the effective dynamics (26), based on ξ_2, to be accurate, at least in the sense of the estimator [D3] (time marginals). Here, we want to check its accuracy in terms of residence times (and hence in a way related to estimator [D2]). Note that, for this reaction coordinate, $|\nabla \xi_2|$ is not a constant function, hence the effective dynamics (26) differs from the dynamics (8) for $A \equiv A_2$, the free energy associated to ξ_2.

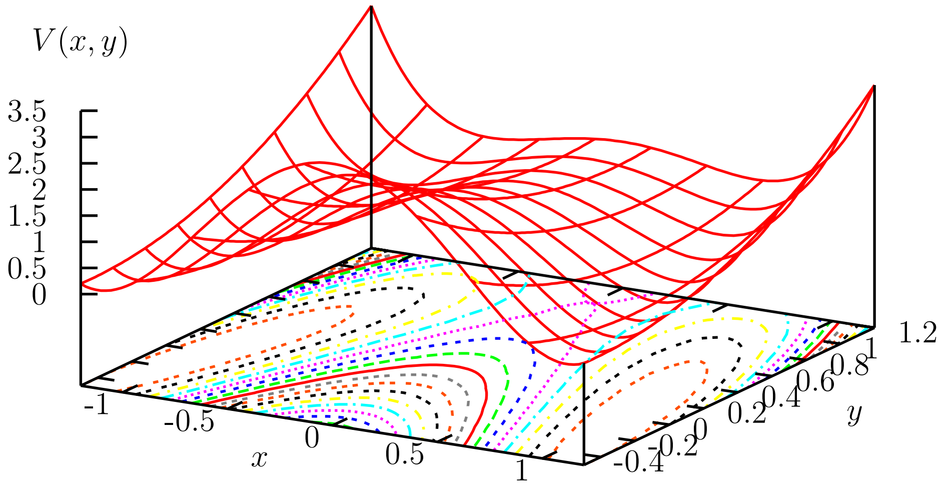

Figure 1. Plot of the double-well potential (49). For clarity of the picture, we set $\varepsilon = 1$.

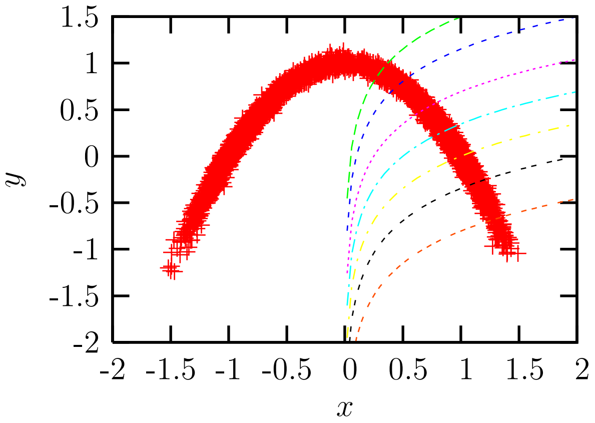

Figure 2. Crosses: plot of the trajectory $X_t = (x_t, y_t)$ solution to (5), for the parameters $\varepsilon = 0.01$ and $\beta = 3$. Dashed lines: level sets of ξ_2.

We work with the parameters $\varepsilon = 0.01$ and $\beta = 3$. On Figure 2, we plot the trajectory solution to (5), as well as the level sets of ξ_2. We can see that the trajectory remains close to the line $\{(x, y);\ q(x, y) = x^2 + y - 1 = 0\}$ (since ε is small), and that the level sets of ξ_2 are parallel to ∇q, which implies that $\nabla \xi_2$ is indeed perpendicular to ∇q.

With the choice we made for β and ε, the system is metastable. On Figure 3, we plot x_t as a function of time, where $X_t = (x_t, y_t)$ satisfies (5). We clearly see that x_t remains close to -1 (that is, the system is in the left well) for a long time before hoping to the right well.

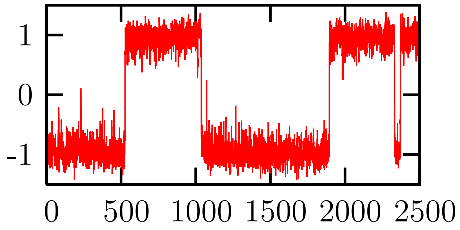

Figure 3. Time evolution $t \mapsto x_t$, for $X_t = (x_t, y_t)$ solution to (5), for the parameters $\varepsilon = 0.01$ and $\beta = 3$. We clearly see metastability.

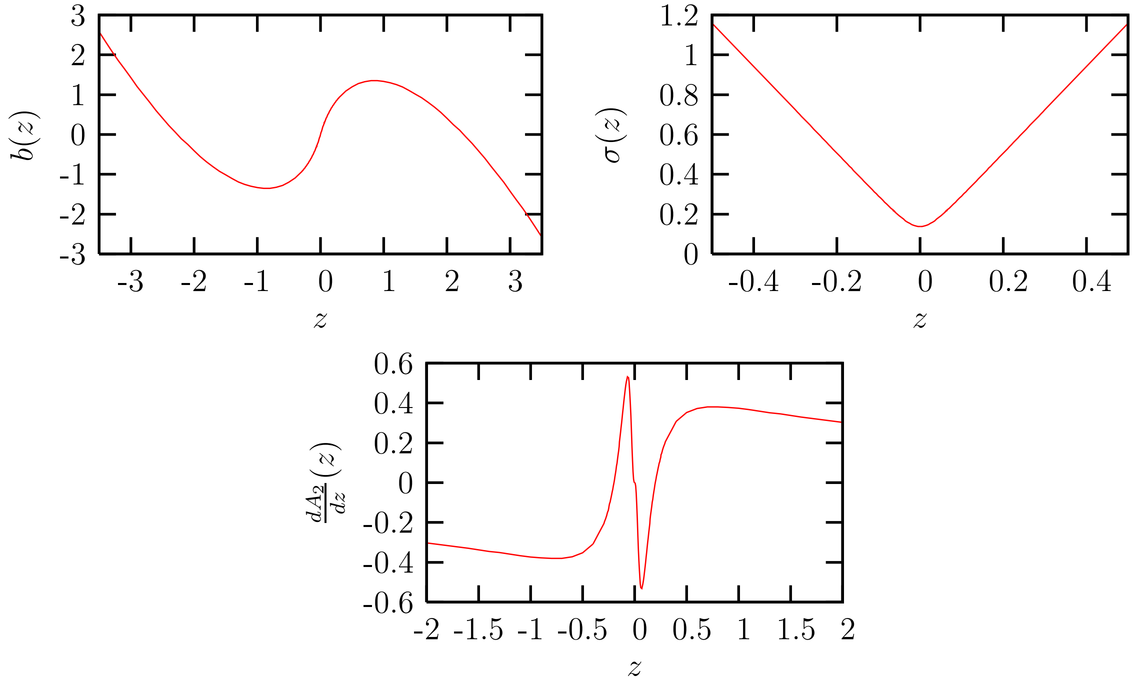

Figure 4. Plot of the functions b, σ and A_2', for the reaction coordinate ξ_2. Note that b and A_2' are odd functions, whereas σ is an even function. Note the large variations of A_2' in the neighbourhood of $z = 0$.

The functions b and σ, as well as the derivative of the free energy A_2 (respectively defined by (24), (25) and (11)) are plotted on Figure 4, in the case of the reaction coordinate ξ_2.

Remark 4.1 *For all the numerical tests reported in this article, the complete dynamics (5) has been integrated with the Euler-Maruyama scheme*

$$X_{j+1} = X_j - \Delta t \, \nabla V(X_j) + \sqrt{2 \, \Delta t \, \beta^{-1}} \; G_j,$$

where, for any j, G_j is a two-dimensional vector, whose coordinates are independent and identically distributed (i.i.d.) random variables, distributed according to a normal Gaussian law.

For the reaction coordinate ξ_1, the effective dynamics is (8), that we have numerically simulated with the same algorithm as above. We have used the analytical expression (50) of the free energy A_1.

For the reaction coordinate ξ_2, the free energy derivative A_2' and the functions b and σ have been computed using the algorithm proposed in [7]. We have chosen to work in the interval $\xi_2 \in [-200; 200]$, and computed A_2', b and σ on a grid of size $\Delta z = 0.1$ (except in the interval $[-0.3; 0.3]$, where we used a finer grid of size $\Delta z = 5.\,10^{-3}$, since the variations of A_2', b and σ are larger in the neighbourhood of 0). Values of the functions for z in-between points of that grid have been obtained by linear interpolation (see Figure 4). We have again used the Euler-Maruyama scheme to numerically integrate the dynamics (26).

All dynamics have been integrated with the time step $\Delta t = 10^{-4}$.

For the reaction coordinate ξ_i, $i = 1, 2$, the left and the right wells are defined as the sets $\left\{(x, y) \in \mathbb{R}^2;\ \xi_i(x, y) \leq -\xi_i^{\rm th}\right\}$ and $\left\{(x, y) \in \mathbb{R}^2;\ \xi_i(x, y) \geq \xi_i^{\rm th}\right\}$, respectively. We have chosen the threshold values $\xi_1^{\rm th} > 0$ and $\xi_2^{\rm th} > 0$ such that wells are more or less the same for both reaction coordinates. To compute the residence time, we proceeded as follow, for both reaction coordinates ξ_1 and ξ_2:

(i) we first generated 15 000 configurations $\{(x_i, y_i) \in \mathbb{R}^2\}_{1 \leq i \leq 15\,000}$, distributed according to the measure μ, and such that $\xi(x_i, y_i)$ belongs to the right well, that is $\xi(x_i, y_i) > \xi^{\rm th}$.

(ii) we next ran the dynamics (5) from the initial condition (x_i, y_i), and monitor the first time τ_i at which the system reaches a point $(x(\tau_i), y(\tau_i))$ in the left well: $\tau_i = \inf\left\{t;\ \xi(x(t), y(t)) < -\xi^{\rm th}\right\}$.

(iii) from these $(\tau_i)_{1 \leq i \leq 15\,000}$, we computed an average residence time and a confidence interval. These figures are the reference figures.

(iv) we next consider the initial conditions $\{\xi(x_i, y_i) \in \mathbb{R}\}_{1 \leq i \leq 15\,000}$ for the effective dynamics. By construction, these configurations are distributed according to the equilibrium measure of $\xi \star \mu$, that is $\exp(-\beta A(z))\, dz$, and are in the right well.

(v) from these initial conditions, we run the dynamics (26) or (8), until the left well is reached ($y(t) \leq -\xi^{\rm th}$ when working with (26), $\overline{y}(t) \leq -\xi^{\rm th}$ when working with (8)), and compute, as for the complete description, a residence time and its confidence interval.

The results we found for the residence time are gathered in Table 1. We see that, when we work with ξ_2 (which satisfies the condition $\nabla \xi_2 \cdot \nabla q = 0$) *and* with the effective dynamics (26), we can reproduce the reference residence time (32.5 ± 0.5) within an excellent accuracy. If we still use the reaction coordinate ξ_2, but consider as the coarse-grained dynamics the dynamics (8) driven by the free energy A_2, then we obtain results that are inconsistent with the reference results given by the complete description of the system.

Reac. Coord.	ξ^{th}	Ref. residence time	Reduced dyn. type	CG residence time
$\xi_2(x,y)$	0.13	32.5 ± 0.5	Eff. dyn. (26)	32.7 ± 0.5
$\xi_2(x,y)$	0.13	32.5 ± 0.5	Dyn. (8)	6.4 ± 0.3
$\xi_1(x,y)$	0.5	31.6 ± 0.5	Dyn. (26) = (8)	24.4 ± 0.4

Table 1. Residence times obtained from the complete description (third column) and from the reduced description (last column), for both reaction coordinates (and both dynamics (26) and (8) when applicable). The threshold values ($\xi_1^{\mathrm{th}} = 0.5$ and $\xi_2^{\mathrm{th}} = 0.13$) have been adjusted so that the reference residence times for both reaction coordinates (31.6 and 32.5, respectively) are almost equal.

Note also that the results obtained with choosing ξ_1 as reaction coordinate, which is such that $\nabla \xi_1 \cdot \nabla q \neq 0$, are inconsistent with the reference results (in that case, the effective dynamics (26) is the same as (8)). Actually, the coarse-grained dynamics does not depend on ε (since the free energy A_1 does not depend on ε), whereas the complete description does depend on ε.

5. Pathwise convergence

In this section, we prove pathwise convergence results between $\xi(X_t)$, where X_t solves (5), and y_t which solves (26), for some potential energies of the type (47). On these specific examples, we obtain stronger convergence results than in the previous sections (namely, convergence in the sense of [D1] rather than in the sense of [D3] or [D2], as in Sections 3 and 4).

Consider the dynamics (5), with the potential energy V_ε defined by (47). It reads

$$dX_t^\varepsilon = -\nabla V_0(X_t^\varepsilon)\,dt - \frac{1}{\varepsilon}\nabla(q^2)(X_t^\varepsilon)\,dt + \sqrt{2\beta^{-1}}\,dW_t, \quad X_{t=0}^\varepsilon = X_0. \quad (51)$$

Note that the initial condition is supposed to not depend on ε. The limit of X_t^ε when $\varepsilon \to 0$ has been identified in [7]: it is a process $\overline{X}_t$ solution of a SDE that we write below (see equation (54)), and that is such that $q(\overline{X}_t) = 0$ for any t.

Assume now that $X_t^\varepsilon \in \mathbb{R}^2$: then $\overline{X}_t$ belongs to the one-dimensional manifold

$$\mathcal{M} = \left\{ X \in \mathbb{R}^2;\ q(X) = 0 \right\}. \quad (52)$$

Assume also that the reaction coordinate ξ is such that its restriction $\xi_{|\mathcal{M}}$ on $\mathcal{M}$ is a one-to-one map from $\mathcal{M}$ to some subset of $\mathbb{R}$ (that is, ξ parameterizes $\mathcal{M}$). In that case, it is easy to build a reduced dynamics from (51), in the limit $\varepsilon \to 0$: one first lets ε go to zero, writes the dynamics of $\overline{X}_t$, and then makes a one-to-one change of variable to write the dynamics in term of $\xi\left(\overline{X}_t\right)$. Our aim is to write conditions under which the so-obtained dynamics corresponds to (26), which amounts to say that the diagram (53) is a commutative diagram.

$$\begin{tikzcd}
\text{2D dynamics (51) on } X_t^\varepsilon \arrow[r, "{\left[\begin{array}{c}\text{pathwise}\\ \text{convergence}\end{array}\right]}", "\varepsilon\to 0"'] \arrow[d] & \text{1D limit dynamics (54) on } \overline{X}_t \arrow[dd] \\
\textit{Itô computation} \arrow[d] & \\
\text{Nonclosed dynamics on } \xi(X_t^\varepsilon) \arrow[d] & \begin{array}{c}\textit{One-to-one change of variable:}\\ z_t = \xi(\overline{X}_t)\end{array} \arrow[dd] \\
\textit{Conditional expectations} \arrow[d] & \\
\begin{array}{c}\text{Dynamics (26) on } y_t^\varepsilon \approx \xi(X_t^\varepsilon) :\\ dy_t^\varepsilon = b_\varepsilon(y_t^\varepsilon)\,dt + \sigma_\varepsilon(y_t^\varepsilon)dB_t\end{array} \arrow[r, "\varepsilon\to 0"'] & \text{Dynamics (60) on } z_t
\end{tikzcd} \tag{53}$$

5.1. Limit of (51) in a pathwise sense

We now proceed in details. For any $X \in \mathcal{M}$, let

$$P(X) = \mathrm{Id} - \frac{\nabla q \otimes \nabla q}{|\nabla q|^2}(X)$$

be the projector on $\mathcal{T}_X\mathcal{M}$, the tangent space to $\mathcal{M}$ at X. Let us define

$$n = \frac{\nabla q}{|\nabla q|} \quad \text{and} \quad \kappa = \operatorname{div} n.$$

Let us now introduce the process $\overline{X}_t$ solution to the equation

$$d\overline{X}_t = -P\left(\overline{X}_t\right)\nabla\left(V_0 + \beta^{-1}\ln|\nabla q|\right)\left(\overline{X}_t\right)\,dt - \beta^{-1}\kappa\, n\,dt + \sqrt{2\beta^{-1}}\,P\left(\overline{X}_t\right)\,dW_t, \tag{54}$$

with the same initial condition $\overline{X}_{t=0} = X_0$ as (51). Let us assume that this initial condition satisfies $q(X_0) = 0$, and let us fix a time interval $[0,T]$. Then (see [7]), under some regularity assumptions on q and V_0, there exists a constant C that does not depend on ε such that

$$\sup_{t\in[0,T]} \mathbb{E}\left|X_t^\varepsilon - \overline{X}_t\right|^2 \le C\varepsilon. \tag{55}$$

Note also that $q\left(\overline{X}_t\right) = 0$ for any time t.

Assume now that there exists a one-to-one map

$$\chi : X \in \mathbb{R}^2 \mapsto (\xi(X), q(X)), \tag{56}$$

which implies that the manifold $\mathcal{M}$ defined by (52) can be parameterized by ξ. Then, equation (54) is equivalent to the dynamics

$$d\left(\xi\left(\overline{X}_t\right)\right) = \nabla\xi\left(\overline{X}_t\right)\cdot d\overline{X}_t + \beta^{-1}\,P\left(\overline{X}_t\right) : \nabla^2\xi\left(\overline{X}_t\right)\,dt.$$

After some tedious but not difficult computations, we see that the above dynamics can be written

$$d\left(\xi\left(\overline{X}_t\right)\right) = d_1\left(\overline{X}_t\right)\,dt + d_2\left(\overline{X}_t\right)\,dt + \sqrt{2\beta^{-1}}\,|\nabla\xi|\left(\overline{X}_t\right)\,dB_t, \tag{57}$$

with again $dB_t = \dfrac{\nabla \xi}{|\nabla \xi|} \left(\overline{X}_t\right) \cdot dW_t$, and

$$\begin{aligned} d_1 &= -\nabla \xi \cdot \nabla V_0 + \beta^{-1} \Delta \xi, \\ d_2 &= -\frac{1}{\beta} \frac{\nabla q \cdot \nabla u}{|\nabla q|^2} + u \, \frac{\nabla q \cdot \nabla V_0}{|\nabla q|^2} - \frac{1}{\beta} \, \kappa \, \frac{u}{|\nabla q|} + \frac{1}{\beta} \, u \, \frac{\nabla q^T \, \nabla^2 q \, \nabla q}{|\nabla q|^4}, \end{aligned} \tag{58}$$

where we set

$$u = \nabla \xi \cdot \nabla q. \tag{59}$$

Since $\overline{X}_t$ satisfies the constraint $q\left(\overline{X}_t\right) = 0$, the dynamics (57) can be rewritten only in terms of $\xi\left(\overline{X}_t\right) =: z_t$, in the form

$$dz_t = \widetilde{d_1}(z_t) \, dt + \widetilde{d_2}(z_t) \, dt + \sqrt{2\beta^{-1}} \, \widetilde{\gamma}(z_t) \, dB_t \tag{60}$$

where, for any z,

$$\widetilde{d_1}(z) = d_1\left(\chi^{-1}(z,0)\right), \quad \widetilde{d_2}(z) = d_2\left(\chi^{-1}(z,0)\right), \quad \widetilde{\gamma}(z) = |\nabla \xi|\left(\chi^{-1}(z,0)\right). \tag{61}$$

5.2. Effective dynamics associated to (51) using conditional expectations

We now follow the strategy that we have outlined in Section 2.3. Starting from (51), we first compute the time derivative of $\xi(X_t^\varepsilon)$ by an Itô computation, and next take the conditional expectations of the drift and the diffusion terms. We hence obtain the effective dynamics (26), where b_ε and σ_ε (that depend on ε since the Gibbs measure μ_ε depends on ε) are defined by (24) and (25) and read

$$\begin{aligned} b_\varepsilon(\alpha) &= \mathbb{E}_{\mu_\varepsilon}\left[\left(-\nabla V_\varepsilon \cdot \nabla \xi + \beta^{-1} \Delta \xi\right)(X) \mid \xi(X) = \alpha\right] \\ &= \mathbb{E}_{\mu_\varepsilon}\left[\left(-\nabla V_0 \cdot \nabla \xi + \beta^{-1} \Delta \xi\right)(X) \mid \xi(X) = \alpha\right] \\ &\qquad - \frac{2}{\varepsilon} \, \mathbb{E}_{\mu_\varepsilon}\left[\left(q \, \nabla q \cdot \nabla \xi\right)(X) \mid \xi(X) = \alpha\right] \\ &= \widetilde{d_1^\varepsilon}(\alpha) - \frac{2}{\varepsilon} \, \mathbb{E}_{\mu_\varepsilon}\left[\left(q \, \nabla q \cdot \nabla \xi\right)(X) \mid \xi(X) = \alpha\right] \\ \sigma_\varepsilon^2(\alpha) &= \mathbb{E}_{\mu_\varepsilon}\left(|\nabla \xi|^2(X) \mid \xi(X) = \alpha\right), \end{aligned} \tag{62}$$

where $\widetilde{d_1^\varepsilon}(\alpha) = \mathbb{E}_{\mu_\varepsilon}\left[\left(-\nabla V_0 \cdot \nabla \xi + \beta^{-1} \Delta \xi\right)(X) \mid \xi(X) = \alpha\right]$. It is easy to check that, for any α, we have

$$\widetilde{d_1^\varepsilon}(\alpha) = \widetilde{d_1}(\alpha) + O(\varepsilon) \quad \text{and} \quad \sigma_\varepsilon(\alpha) = \widetilde{\gamma}(\alpha) + O(\varepsilon), \tag{63}$$

where $\widetilde{d_1}$ and $\widetilde{\gamma}$ are defined by (61).

5.3. Sufficient conditions for the pathwise convergence to the effective dynamics (26)

Let us establish sufficient conditions under which the equation (60) is equivalent to the effective dynamics (26), in the limit $\varepsilon \to 0$. We hence request that, in the limit $\varepsilon \to 0$, the dynamics (26) and (60) have the same drift and diffusion coefficients.

We first see that this condition is satisfied for the diffusion coefficients, in view of (63): the diffusion coefficient σ_ε of (26) converges to $\widetilde{\gamma}$, the diffusion coefficient of (60).

We now turn to the drift terms, which is $\widetilde{d}_1 + \widetilde{d}_2$ in the case of (60), and b_ε given by (62) for the effective dynamics (26). In view of (63), these drift terms are equal, in the limit $\varepsilon \to 0$, if and only if

$$- \widetilde{d}_2(\alpha) = \lim_{\varepsilon \to 0} \frac{2}{\varepsilon} \, \mathbb{E}_{\mu_\varepsilon} \left[\left(q \, \nabla q \cdot \nabla \xi \right)(X) \mid \xi(X) = \alpha \right]. \tag{64}$$

In view of (58) and (61), we have

$$\widetilde{d}_2(\alpha) = d_2 \left(\chi^{-1}(\alpha, 0) \right) = d_{2a} \left(\chi^{-1}(\alpha, 0) \right) + \beta^{-1} d_{2b} \left(\chi^{-1}(\alpha, 0) \right), \tag{65}$$

where d_{2a} and d_{2b} do not depend on β:

$$d_{2a} = u \, \frac{\nabla q \cdot \nabla V_0}{|\nabla q|^2}, \tag{66}$$

$$d_{2b} = - \frac{\nabla q \cdot \nabla u}{|\nabla q|^2} - \kappa \, \frac{u}{|\nabla q|} + u \, \frac{\nabla q^T \, \nabla^2 q \, \nabla q}{|\nabla q|^4}. \tag{67}$$

On the other hand, we compute, for any α,

$$\begin{aligned} \mathbb{E}_{\mu_\varepsilon} \left[\left(q \, \nabla q \cdot \nabla \xi \right)(X) \mid \xi(X) = \alpha \right] &= \mathbb{E}_{\mu_\varepsilon} \left[q(X) \, u(X) \mid \xi(X) = \alpha \right] \\ &= \int_{\Sigma_\alpha} q \, u \; d\mu_{\varepsilon, \Sigma_\alpha}. \end{aligned}$$

A direct computation shows that

$$\mathbb{E}_{\mu_\varepsilon} \left[\left(q \, \nabla q \cdot \nabla \xi \right)(X) \mid \xi(X) = \alpha \right] = \frac{\varepsilon}{2\beta} \, \mathcal{E}(\alpha) + O(\varepsilon^{3/2}), \tag{68}$$

where $\mathcal{E}(\alpha)$ does not depend on β and reads

$$\mathcal{E}(\alpha) = \frac{\partial \widetilde{u}}{\partial q}(\alpha, 0) + \frac{\widetilde{u}(\alpha, 0)}{j(\alpha, 0)} \, \frac{\partial j}{\partial q}(\alpha, 0) + \widetilde{u}(\alpha, 0) \, \frac{\partial \widetilde{V_0}}{\partial q}(\alpha, 0),$$

where

$$\widetilde{u}(\xi, q) = u(\chi^{-1}(\xi, q)), \tag{69}$$

$\widetilde{V_0}(\xi, q) = V_0(\chi^{-1}(\xi, q))$, and $j = \det \operatorname{jac} \chi^{-1}$. Hence, (64) reads

$$- d_{2a}(\chi^{-1}(\alpha, 0)) - \frac{1}{\beta} d_{2b}(\chi^{-1}(\alpha, 0)) = \frac{1}{\beta} \, \mathcal{E}(\alpha). \tag{70}$$

We want to enforce this relation for any β. Since d_{2a}, d_{2b} and $\mathcal{E}$ do not depend on β, this yields

$$d_{2a}(\chi^{-1}(\alpha, 0)) = 0 \quad \text{and} \quad - d_{2b}(\chi^{-1}(\alpha, 0)) = \mathcal{E}(\alpha). \tag{71}$$

In view of (66), a sufficient condition for the first relation to hold is

$$\forall \alpha \in \mathbb{R}, \quad u(\chi^{-1}(\alpha, 0)) = 0, \tag{72}$$

where, we recall, $u = \nabla \xi \cdot \nabla q$ and χ is such that $\chi(X) = (\xi(X), q(X))$. In what follows, we now assume that ξ is such that (72) holds. The second relation of (71) now reads

$$\forall \alpha \in \mathbb{R}, \quad \frac{\nabla q \cdot \nabla u}{|\nabla q|^2}(\chi^{-1}(\alpha, 0)) = \frac{\partial \widetilde{u}}{\partial q}(\alpha, 0). \tag{73}$$

We have thus proved the following result:

Proposition 5.1 *Consider the two-dimensional dynamics (51), and its one-dimensional limit (54), when $\varepsilon \to 0$. On the other hand, consider the one-dimensional effective dynamics (26), obtained using conditional expectations, and pass to the limit $\varepsilon \to 0$ in the drift and diffusion coefficients.*

Under the conditions (72) and (73) (where u, χ and $\widetilde{u}$ are defined by (59), (56) and (69) respectively), these two dynamics are the same. In addition, for any $T > 0$, there exists $C > 0$ and $\varepsilon_0 > 0$ such that, for all $\varepsilon \le \varepsilon_0$, we have

$$\sup_{t\in[0,T]} \mathbb{E} \left|\xi\left(X_t^\varepsilon\right) - y_t^\varepsilon\right|^2 \le C\varepsilon, \tag{74}$$

where X_t^ε solves (51) and y_t^ε solves the effective dynamics (26).

Proof: We only have to prove the bound (74). We infer from (55) and assumption **[H1]** that

$$\sup_{t\in[0,T]} \mathbb{E} \left|\xi\left(X_t^\varepsilon\right) - \xi\left(\overline{X}_t\right)\right|^2 \le C\varepsilon. \tag{75}$$

The drift and diffusion coefficients of the effective dynamics on y_t^ε are b_ε and σ_ε. In view of (62), (63), (68), (70) and (65), the former satisfies

$$\begin{aligned} b_\varepsilon(\alpha) &= \widetilde{d}_1^\varepsilon(\alpha) - \frac{2}{\varepsilon}\, \mathbb{E}_{\mu_\varepsilon}\left[\left(q\,\nabla q \cdot \nabla\xi\right)(X) \mid \xi(X) = \alpha\right] \\ &= \widetilde{d}_1(\alpha) + O(\varepsilon) - \beta^{-1}\mathcal{E}(\alpha) + O(\sqrt{\varepsilon}) \\ &= \widetilde{d}_1(\alpha) + \widetilde{d}_2(\alpha) + O(\sqrt{\varepsilon}). \end{aligned}$$

In view of (63), the latter satisfies

$$\sigma_\varepsilon(\alpha) = \widetilde{\gamma}(\alpha) + O(\varepsilon).$$

Hence, the difference between, on the one hand, the drift and diffusion coefficients of the effective dynamics (b_ε and σ_ε) and, on the other hand, the drift and diffusion coefficients of the equation (60) on $z_t = \xi\left(\overline{X}_t\right)$ (namely $\widetilde{d}_1 + \widetilde{d}_2$ and $\widetilde{\gamma}$), is of order $O\left(\sqrt{\varepsilon}\right)$. We infer from this estimate that, on any bounded time interval,

$$\sup_{t\in[0,T]} \mathbb{E} \left|y_t^\varepsilon - \xi\left(\overline{X}_t\right)\right|^2 \le C\varepsilon.$$

Gathering that estimate with (75) yields (74). □

In Sections 3.2 and 4, we outlined the condition $\nabla\xi \cdot \nabla q = 0$ as an important condition to get a good analytical estimate in the sense of [D3], and good numerical results in terms of residence times. If $u = \nabla\xi \cdot \nabla q = 0$, then conditions (72) and (73) are satisfied, and we also get pathwise convergence (i.e. accuracy in the sense of [D1]), in the simple two-dimensional setting considered in this section.

Hence, the same condition $\nabla\xi \cdot \nabla q = 0$ appears, independently of the estimator ([D3], [D2] or [D1]) that we choose to measure the accuracy of the effective dynamics.

5.4. A standard test-case

Consider the two-dimensional potential energy

$$V_\varepsilon(x,y) = V_0(x,y) + \frac{\Omega^2(x)\,y^2}{\varepsilon}, \quad x \in \mathbb{R},\ y \in \mathbb{R}, \tag{76}$$

where Ω is bounded away from 0 and V_0 does not depend on ε, and the associated overdamped Langevin equation, which defines the process $X_t^\varepsilon = (x_t^\varepsilon, y_t^\varepsilon)$. The limit dynamics on x_t^ε when $\varepsilon \to 0$ is well-known in that case (see for instance [28]): it reads

$$dx_t = -\left(\partial_x V_0(x_t,0) + \frac{\Omega'(x_t)}{\beta\Omega(x_t)}\right) dt + \sqrt{2\,\beta^{-1}}\,dB_t, \tag{77}$$

which is the overdamped Langevin equation associated to the potential $V_{\rm eff}(x) = V_0(x,0) + \beta^{-1}\ln\Omega(x)$.

We now wish to recover that result within our approach. The potential energy (76) is of the form (47), with $q(x,y) = \Omega(x)\,y$. We wish to choose the reaction coordinate $\xi(x,y) = x$. Observe then that $u = \nabla\xi \cdot \nabla q = \Omega'(x)\,y \neq 0$. Hence the simple sufficient condition $u = 0$ (see end of Section 5.3) is not satisfied. However, it is easy to see that the less demanding conditions (72) and (73) are satisfied. Hence, in the limit $\varepsilon \to 0$, the effective dynamics (26) is accurate in the sense of pathwise convergence.

For $\varepsilon > 0$, the effective dynamics reads

$$d\xi_t = b_\varepsilon(\xi_t)\,dt + \sqrt{2\,\beta^{-1}}\,dB_t, \tag{78}$$

with $b_\varepsilon(\alpha) = -\mathbb{E}_{\mu_\varepsilon(\alpha,\cdot)}\left(\partial_x V_0(\alpha,y) + 2\frac{\Omega'(\alpha)\Omega(\alpha)y^2}{\varepsilon}\right)$. A straightforward computation shows that $\lim_{\varepsilon\to 0} b_\varepsilon(\alpha) = -\partial_x V_0(\alpha,0) - \frac{\Omega'(\alpha)}{\beta\Omega(\alpha)}$. Inserting this relation in (78), we recover (77).

Hence, taking the limit $\varepsilon \to 0$ in the effective dynamics that we propose, we recover a well-known result.

5.5. Numerical results on the example (49)

In the numerical case considered in Section 4, we showed that the reaction coordinate $\xi_2(x,y) = x\exp(-2y)$ satisfies the relation $\nabla\xi_2 \cdot \nabla q = 0$. In view of Proposition 5.1, we hence expect good results when working with ξ_2, in terms of pathwise convergence. We have checked this as follows. First, we have simulated a solution of (51) (with the potential V_ε defined by (49)), for a given realization of the two-dimensional noise, with $\varepsilon = 0.01$. From this trajectory X_t (we omit here for clarity the dependence with respect to ε), we obtain the time evolution $\xi_2(X_t)$, and we can also construct the one-dimensional noise (16). This noise is next used in the effective dynamics (26). We compare both trajectories on Figure 5: we observe an excellent agreement over 10^6 time steps (the trajectories plotted on Figure 5 have been computed with a time step $\Delta t = 10^{-4}$, hence $T = 100 = 10^6\Delta t$).

In Section 4, we also considered the reaction coordinate $\xi_1(x,y) = x$, which is such that $u_1(x,y) = \nabla\xi_1 \cdot \nabla q = 2x \neq 0$. With this choice of reaction coordinate,

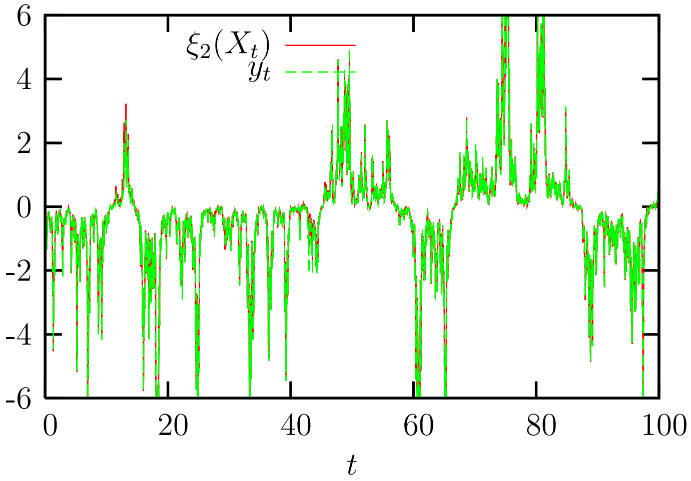

Figure 5. Comparison of $\xi_2(X_t)$, where X_t solves (5), and y_t solution of (26) with the reaction coordinate ξ_2.

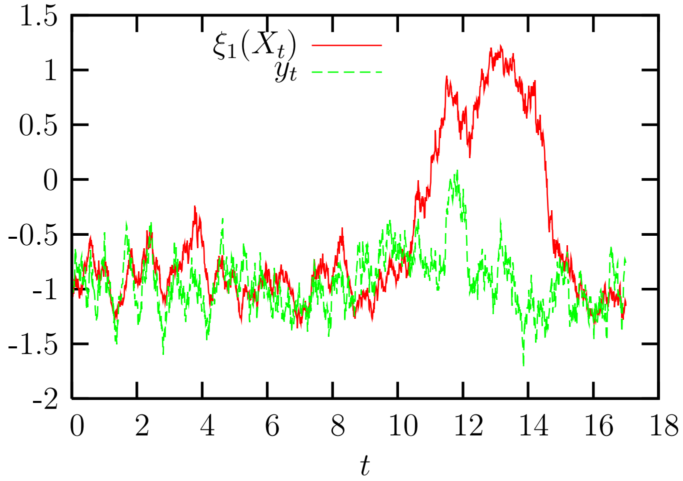

Figure 6. Comparison of $\xi_1(X_t)$, where X_t solves (5), and y_t solution of (26) with the reaction coordinate ξ_1.

$\chi^{-1}(\xi, q) = (\xi, q + 1 - \xi^2)$, hence $u_1(\chi^{-1}(\xi, 0)) = 2\xi$, so condition (72) is not satisfied. We have numerically performed the same comparison with ξ_1 as the one reported above for ξ_2. Results are shown on Figure 6: we observe that the complete dynamics (projected on the reaction coordinate) and the effective dynamics disagree, as expected. Note also the difference in time ranges between Figures 5 and 6 (the former corresponding to a time interval 5 times larger than the latter).

5.6. Numerical results on a three atom molecule

We conclude this section by considering a system closer to those considered in molecular simulation, although we acknowledge that it is still a toy-example. The system is made of three two-dimensional particles at position $r_i \in \mathbb{R}^2$, $1 \leq i \leq 3$ (hence

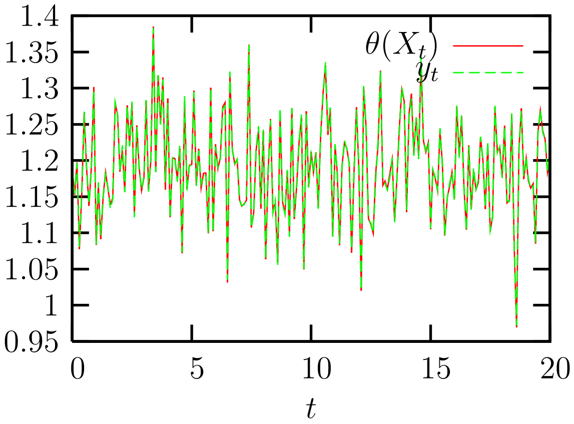

Figure 7. Comparison of $\theta(X_t)$, where X_t solves (5), and y_t solution of (26) with the reaction coordinate $X \mapsto \theta(X)$.

$X = (r_1, r_2, r_3) \in \mathbb{R}^6$), and submitted to the potential

$$\begin{aligned} V(X) &= \frac{1}{2\varepsilon}\left(\|r_1 - r_2\| - \ell_0\right)^2 + \frac{1}{2\varepsilon}\left(\|r_2 - r_3\| - \ell_0\right)^2 + \frac{1}{2}k_\theta(\theta(X) - \theta_0)^2 \\ &= \frac{1}{2\varepsilon}\left(q_1(X)^2 + q_3(X)^2\right) + \frac{1}{2}k_\theta(\theta(X) - \theta_0)^2, \end{aligned}$$

where $\theta(X)$ is the angle between the bonds (r_1, r_2) and (r_2, r_3), $q_1(X) = \|r_1 - r_2\| - \ell_0$ and $q_3(X) = \|r_2 - r_3\| - \ell_0$. In the above potential, ℓ_0 is an equilibrium length whereas θ_0 is an equilibrium angle. This potential represents stiff bonds between particles 1 and 2 on the one hand, and 2 and 3 on the other hand, with a softer term depending on the three-body angle θ. To remove rigid body motion invariance, we set $r_2 = 0$ and $r_1 \cdot e_y = 0$. Then it is easy to see that the angle $\theta(X)$ satisfies $\nabla\theta \cdot \nabla q_1 = \nabla\theta \cdot \nabla q_3 = 0$, and hence seems to be a good reaction coordinate, in view of the several discussions above.

Numerical experiments confirm this belief: choosing this reaction coordinate, we considered the effective dynamics (26), and compared its solution with the time evolution $\theta(X_t)$, where X_t solves (5). Results are shown on Figure 7 (we worked with the numerical parameters $\varepsilon = 10^{-3}$, $\ell_0 = 1$, $\theta_0 = 1.187$ and $k_\theta = 208$): again, we see a good agreement between both trajectories.

Acknowledgments

This work is supported in part by the INRIA, through the grant 'Action de Recherche Collaborative' HYBRID, and by the MEGAS non-thematic program (Agence Nationale de la Recherche, France). Part of this work was completed while the two authors were visiting the Hausdorff Research Institute for Mathematics (HIM, Bonn), whose hospitality is gratefully acknowledged. We also wish to acknowledge very enlightning discussions with F. Otto and G. Menz.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques.* Société Mathématique de France, 2000.

[2] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Part. Diff. Eq.*, 26:43–100, 2001.

[3] K. Bichteler. *Stochastic integration with jumps.* Cambridge University Press, 2002.

[4] S. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.

[5] E. Cancès, F. Legoll, and G. Stoltz. Theoretical and numerical comparison of some sampling methods for molecular dynamics. *Math. Mod. Num. Anal. (M2AN)*, 41(2):351–389, 2007.

[6] C. Chipot and A. Pohorille, editors. *Free energy calculations*, volume 86 of *Springer Series in Chemical Physics.* Springer, 2007.

[7] G. Ciccotti, T. Lelièvre, and E. Vanden-Eijnden. Projection of diffusions on submanifolds: application to mean force computation. *Comm. Pure and Applied Math.*, 61(3):371–408, 2008.

[8] E. Darve, J. Solomon, and A. Kia. Computing generalized Langevin equations and generalized Fokker-Planck equations. *Proceedings of the National Academy of Sciences*, 2009. in press.

[9] D. Dizdar. *Towards an optimal rate of convergence in the hydrodynamic limit for Kawasaki dynamics.* PhD thesis, Bonn University, 2007.

[10] L.C. Evans and R.F. Gariepy. *Measure theory and fine properties of functions.* Studies in Advanced Mathematics. CRC Press, 1992.

[11] D. Givon, R. Kupferman, and A.M. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17(6):55–127, 2004.

[12] N. Grunewald, F. Otto, C. Villani, and M.G. Westdickenberg. A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit. *Ann. Inst. H. Poincaré Probab. Statist.*, 45(2):302–351, 2009.

[13] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probab. Th. Rel. Fields*, 71:501–516, 1986.

[14] P. Hänggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Reviews of Modern Physics*, 62(2):251–342, 1990.

[15] C. Hartmann. *Model reduction in classical molecular dynamics.* PhD thesis, Freie Universität Berlin, 2007. http://www.diss.fu-berlin.de/2007/458.

[16] C. Hartmann and C. Schütte. Balancing of partially-observed stochastic differential equations. *47th IEEE conference on decision and control*, pages 4867–4872, 2008.

[17] C. Hartmann, V.-M. Vulcanov, and C. Schütte. Balanced truncation of linear second-order systems: a Hamiltonian approach. *SIAM Mult. Mod. Sim.*, submitted.

[18] R.Z. Has'minskii. *Stochastic stability of differential equations.* Sijthoff and Noordhoff, Alphen aan den Rijn, 1980.

[19] I. Horenko, C. Hartmann, C. Schütte, and F. Noe. Data-based parameter estimation of generalized multidimensional Langevin processes. *Phys. Rev. E*, 76:016706, 2007.

[20] J.A. Izaguirre and C.R. Sweet. Adaptive dimensionality reduction of stochastic differential equations for protein dynamics. *Proc. second international workshop on model reduction in reacting flows*, April 2009 (Notre Dame, IN).

[21] W. Kliemann. Recurrence and invariant measures for degenerate diffusions. *The annals of probability*, 15(2):690–707, 1987.

[22] H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[23] T. Lelièvre. A general two-scale criteria for logarithmic Sobolev inequalities. *J. Funct. Anal.*, 256(7):2211–2221, 2009.

[24] A. Michalak and T. Ziegler. First-principle molecular dynamic simulations along the intrinsic

reaction paths. *J. Phys. Chem. A*, 105:4333–4343, 2001.

[25] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[26] G.A. Pavliotis and A.M. Stuart. *Multiscale methods: averaging and homogenization.* Springer, 2007.

[27] Y. Pokern, A.M. Stuart, and P. Wiberg. Parameter estimation for partially observed hypoelliptic diffusions. *J. Royal Statistical Society, Series B*, 71(1):49–73, 2009.

[28] S. Reich. Smoothed Langevin dynamics of highly oscillatory systems. *Physica D*, 138:210–224, 2000.

[29] J.P. Ryckaert and A. Bellemans. Molecular dynamics of liquid alkanes. *Faraday Discuss.*, 66:95–106, 1978.

[30] C. Schütte. private communication.

[31] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A direct approach to conformational dynamics based on Hybrid Monte-Carlo. *J. Comp. Phys.*, 151:146–168, 1999.

[32] C. Schütte and W. Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. In P.G. Ciarlet and C. Le Bris, editors, *Handbook of Numerical Analysis (Special volume on computational chemistry)*, volume X, pages 699–744. Elsevier, 2003.

[33] C.R. Sweet, P. Petrone, V.S. Pande, and J.A. Izaguirre. Normal mode partitioning of Langevin dynamics for biomolecules. *J. Chem. Phys.*, 128(14):145101, 2008.

[34] E. Vanden-Eijnden. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Comm. Math. Sci.*, 1(2):385–391, 2003.

[35] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003.

[36] S. Yang, J.N. Onuchic, and H. Levine. Effective stochastic dynamics on a protein folding energy landcape. *J. Chem. Phys.*, 125:054910, 2006.