Effective dynamics using conditional expectations

Frédéric Legoll¹,² and Tony Lelièvre²,³

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 and lelievre@cermics.enpc.fr

Received 26 June 2009, in final form 30 April 2010
Published 29 July 2010
Online at stacks.iop.org/Non/23/2131

Recommended by B Leimkuhler

Abstract

The question of coarse-graining is ubiquitous in molecular dynamics. In this paper, 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 $\xi$ is a given 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, with precise error bounds. We check numerically on several 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.

Mathematics Subject Classification: 35B40, 82C31, 60H10

(Some figures in this article are in colour only in the electronic version)

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, such as 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 as

\[
\int_{\mathbb{R}^n} \Phi(x) \, d\mu,
\]

where \( \Phi : \mathbb{R}^n \to \mathbb{R} \) is the observable of interest and \( \mu \) is the Boltzmann–Gibbs measure,

\[
d\mu = Z^{-1} \exp(-\beta V(x)) \, dx,
\]

(1)

where \( V \) is the potential energy of the system, \( \beta \) is proportional to the inverse of the system temperature and \( Z = \int_{\mathbb{R}^n} \exp(-\beta V(x)) \, dx \) 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 \( \mathbb{T} = \mathbb{R}/Z \) 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 paper, we assume that

[H1] \( \xi \) is a smooth scalar function such that,

for all \( x \in \mathbb{R}^n \), \( 0 < m \leq |\nabla \xi(x)| \leq M < \infty \).

It is not clear to us whether the results of this paper can be generalized to the case of a multi-dimensional reaction coordinate.

To this function \( \xi \) is naturally associated an effective energy \( A \), called the free energy, such that

\[
d(\xi \ast \mu) = \exp(-\beta A(z)) \, dz,
\]

where \( \xi \ast \mu \) denotes the image of the measure \( \mu \) by \( \xi \). In other words, for any test function \( \Phi : \mathbb{R} \to \mathbb{R} \),

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

(2)

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

The interpretation of (2) is that, when \( X \) is distributed according to the Boltzmann measure (1), 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 paper.
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 $\xi$ 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),

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

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 (1). Hence, for $\mu$-almost all initial conditions $X_0$,

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

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 dynamics (3). Typically, the wells are fully described through $\xi$ (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 paper, our aim is twofold. First, we would like to propose a one-dimensional dynamics of the form

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

where $B_t$ is a standard one-dimensional Brownian motion and $b$ and $\sigma$ are scalar functions, such that $(y_t)_{0 \leq t \leq T}$ is a good approximation (in a sense to be made precise below) of $(\xi(X_t))_{0 \leq t \leq T}$. Hence, dynamics (4) can be thought of as a coarse-grained, or effective, dynamics for the quantity of interest. A natural requirement is that (4) 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 (4), 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 (4) and the coarse-grained dynamics

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

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 $\{x \in \mathbb{R}^n; \xi(x) \leq z_0\}$) 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 (5), then the transition state theory says that residence times are a function of the free energy barriers [19, 27], 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 [7, 8, 10, 24] and also [22]). A standard framework is then the Mori–Zwanzig projection formalism, which is described in detail in [16]. Note also that, since we are interested in reproducing only some output function of $X_t$ (namely $\xi(X_t)$), tools from control theory may be used. Such an idea has been followed in [20, 21].

If a time-scale separation is present in the system, then memory effects may be neglected. In the following, 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 type (4). 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 (the time-scale separation is encoded in the constants entering the logarithmic Sobolev inequalities). In addition, within this framework, we can handle any given reaction coordinate, which may be a nonlinear function 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. One may then apply an averaging principle (see [22] and the references therein for more details along this idea; see also [32] 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 stiffnesses, as an example of application of our general result (see the potential energy (44)).

Other strategies are to try and identify fast and slow modes in the dynamics (see, e.g. [25, 39]) or to postulate a parametric form for the effective dynamics and to identify its coefficients by numerical simulation of the complete system [33, 41].

Let us emphasize that the effective dynamics we will derive (see equations (21), (22) and (23)) has already been obtained, using different arguments, in [14, equation (81)] and [29, equation (17)]. Reference [14, section 11] also contains a discussion of the relevance of the coarse-grained dynamics in the asymptotic of a very large time-scale separation, and in the case when a precise reaction coordinate (namely the so-called committor function) is chosen. The interest of our approach is to get an error estimate, for any coarse-grained variable, and which is not an asymptotic result. Of course, the error estimate we get certainly yields a large error bound in some cases (in particular if $\xi$ is not well chosen or when no time-scale separation is present in the dynamics), and this will be checked numerically in the following.

We complete this section by a discussion of the choice of the full dynamics. We chose the overdamped Langevin dynamics (3). 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 paper (such as proposition 3.1) can be generalized to the case of the Langevin dynamics. From a practical viewpoint, it is however possible to use the same strategy starting from the Langevin dynamics to write an effective low-dimensional dynamics. 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 [15, 36, 37].

1.3. Statement of the main results and outline

We propose a way to derive an effective dynamics of the form (4). This defines a process \((y_t)_{t \geq 0}\), which we compare with \((\xi(X_t))_{t \geq 0}\), where \(X_t\) satisfies (3). 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: \(\mathbb{E}(\sup_{t \in [0, T]} |\xi(X_t) - y_t|^2)\),

- [D2] convergence of the laws of paths: \(\|\mathcal{L}(\xi(X_{t_{0} \leq t \leq T})) - \mathcal{L}(y_{t_{0} \leq t \leq T})\|_{TV}\),

- [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 \(\nu\) is defined by \(\|\nu\|_{TV} = \sup_{f \in L^{\infty}, \|f\|_{L^{\infty}} \leq 1} \int f \, d\nu\). If \(\nu\) 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 (3) and (4)), 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: practitioners 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 potential energy 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 (see section 2.1). As pointed out above, such an estimate does not guarantee convergence in the sense of [D1] and [D2], which would be needed for example to get accurate estimation of correlations in times, to compute e.g. transport coefficients. Therefore, in section 2.2 we investigate, for some particular forms of potentials, sufficient conditions to get an accurate effective dynamics (in the sense of [D3]). We next check numerically if these conditions are indeed sufficient and necessary to get an accurate effective dynamics, in terms of residence times. We also investigate, both theoretically and numerically, on a toy two-dimensional situation, the relevance of these conditions in terms of pathwise convergence (namely [D1]).

The paper 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 (4), where the functions \(b\) and \(\sigma\) can be computed in practice (see equations (21), (22) and (23)). In section 3, under some time-scale separation assumptions, we prove that the solution \(y_t\) of the effective dynamics (23) 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, 17]. In section 4, we present numerical results obtained on several test cases, where we compare residence times in the potential energy wells as predicted by the reference dynamics (3) and by the one-dimensional reduced dynamics (23).

Section 5 is dedicated to establishing error estimates in the sense [D1] of pathwise convergence,
2. A ‘natural’ coarse-graining procedure

2.1. Notation

We gather here some useful notation and results. Let \( \Sigma_z \) 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 \( \mu_{\Sigma_z} \), which is the probability measure \( \mu \) conditioned at a fixed value of the reaction coordinate:

\[
d\mu_{\Sigma_z} = \exp(-\beta V) \frac{d|\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{Z \exp(-\beta V) |\nabla \xi|^{-1} d\sigma_{\Sigma_z}},
\]

where the measure \( \sigma_{\Sigma_z} \) is the Lebesgue measure on \( \Sigma_z \) 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 [9]):

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

\[
A'(z) = \int_{\Sigma_z} F d\mu_{\Sigma_z},
\]

where \( F \) is the so-called local mean force:

\[
F = \frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \text{div} \left( \frac{\nabla \xi}{|\nabla \xi|^2} \right).
\]

In view of (7), note that (6) reads as

\[
d\mu_{\Sigma_z} = \frac{\exp(-\beta V) |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{Z \exp(-\beta A)}.
\]

These expressions can be obtained by the co-area formula [13], which we now recall:

**Lemma 2.1.** For any smooth function \( \Phi : \mathbb{R}^n \rightarrow \mathbb{R} \), consider

\[
\Phi_A(x) = \int_{\Sigma_z} \Phi(x) |\nabla \xi(x)| \, dx = \int_{\mathbb{R}} \int_{\Sigma_z} \Phi d\sigma_{\Sigma_z} \, dz.
\]

**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 \( \mu_{\Sigma_z} \) defined by (6). The measure \( |\nabla \xi|^{-1} d\sigma_{\Sigma_z} \) is sometimes denoted by \( \delta_{\xi(x)=z}(dx) \) in the literature.

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

**Lemma 2.2.** For any smooth function \( \chi : \mathbb{R}^n \rightarrow \mathbb{R} \), consider

\[
\chi\xi = \int_{\Sigma_z} \chi |\nabla \xi|^{-1} d\sigma_{\Sigma_z}.
\]

The derivative of \( \chi\xi \) reads as

\[
\frac{d\chi\xi}{dz}(z) = \int_{\Sigma_z} \left[ \frac{\nabla \xi \cdot \nabla \chi}{|\nabla \xi|^2} + \chi \text{div} \left( \frac{\nabla \xi}{|\nabla \xi|^2} \right) \right] |\nabla \xi|^{-1} d\sigma_{\Sigma_z}.
\]
Proof. For any smooth test function $g : \mathbb{R} \rightarrow \mathbb{R}$, we obtain, using the co-area formula (11), that

$$\int_{\mathbb{R}} \chi_\xi(z) g'(z) \, dz = \int_{\mathbb{R}} \int_{\Sigma_1} \chi_\xi |\nabla \xi|^{-1} g'(z) \, d\sigma \, dz = \int_{\mathbb{R}^n} \chi(x) g'(\xi(x)) \, dx.$$ 

Hence,

$$\int_{\mathbb{R}} \chi_\xi(z) g'(z) \, dz = \int_{\mathbb{R}^n} \chi(x) g'(\xi(x)) \, dx = -\int_{\mathbb{R}^n} g(\xi) \text{div} \left( \chi |\nabla \xi|^2 \nabla \xi \right) \, dx,$$

which yields the result. □

2.2. A non-closed equation

Consider $X_t$ that solves (3). By a simple Itô computation, we have

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

(12)

where $B_t$ is the one-dimensional Brownian motion

$$d B_t = \frac{\nabla \xi}{|\nabla \xi|} (X_t) \cdot dW_t.$$ 

(13)

Of course, equation (12) is not closed. Following Győngy [18], a simple closing procedure is to consider $\tilde{\gamma}_t$ solution to

$$d \tilde{\gamma}_t = \tilde{b}(t, \tilde{\gamma}_t) \, dt + \sqrt{2\beta^{-1}} \tilde{\sigma}(t, \tilde{\gamma}_t) \, dB_t,$$ 

(14)

where

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

(15)

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

(16)

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

As shown in [18], 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 $\psi^\xi$ of $\xi(X_t)$, where $X_t$ satisfies (3), satisfies the Fokker–Planck equation associated with (14):

$$\partial_t \psi^\xi = \partial_x \left( -\tilde{b} \psi^\xi + \beta^{-1} \partial_x (\tilde{\sigma}^2 \psi^\xi) \right).$$ 

(17)

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

$$\partial_t \psi = \text{div} \left( \nabla \psi + \beta^{-1} \nabla \psi \right).$$ 

(18)
In view of remark 2.1, the probability distribution function \( \psi^\xi(t, z) \) of \( \xi(X_t) \) is given by

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

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

\[
\partial_z \psi^\xi(t, z) = \int_{\Sigma_t} \nabla \cdot \nabla \psi(t, \cdot) |\nabla \xi|^{-1} d\sigma_{\Sigma_t}.
\]

By definition, we have the following expressions for \( \tilde{b} \) and \( \tilde{\sigma}^2 \) in terms of \( \psi \):

\[
\tilde{b}(t, z) = \frac{1}{\psi(t, z)} \int_{\Sigma_t} (-\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi) |\nabla \xi|^{-1} \psi \, d\sigma_{\Sigma_t},
\]

\[
\tilde{\sigma}^2(t, z) = \frac{1}{\psi(t, z)} \int_{\Sigma_t} |\nabla \xi| \psi \, d\sigma_{\Sigma_t}.
\]

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

\[
\partial_z (\tilde{\sigma}^2 \psi^\xi) = \partial_z \int_{\Sigma_t} |\nabla \xi|^2 \psi \, d\sigma_{\Sigma_t} = \int_{\Sigma_t} (\nabla \cdot |\nabla \psi + \psi \Delta \xi| |\nabla \xi|^{-1} d\sigma_{\Sigma_t}.
\]

Let us now prove a variational formulation of (17). For any test function \( g \), we have

\[
\frac{d}{dt} \int_R \psi^\xi(t, z) g(z) \, dz = \frac{d}{dt} \int_R \psi(t, x) g(\xi(x)) \, dx
\]

\[
= \int_R \text{div} \left( \psi \nabla V + \beta^{-1} \nabla \psi \right) g \circ \xi \, dx
\]

\[
= -\int_R \left( \psi \nabla V + \beta^{-1} \nabla \psi \right) \cdot \nabla \xi \, g' \circ \xi \, dx
\]

\[
= -\int_R \int_{\Sigma_t} |\nabla \xi|^{-1} \left( \psi \nabla V \cdot \nabla \xi + \beta^{-1} \nabla \psi \cdot \nabla \xi \right) d\sigma_{\Sigma_t} \, g'(z) \, dz
\]

\[
= -\beta^{-1} \int_R \int_{\Sigma_t} \partial_z (\tilde{\sigma}^2 \psi^\xi) g'(z) \, dz
\]

\[
+ \int_R \int_{\Sigma_t} |\nabla \xi|^{-1} \left( -\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi \right) \psi \, d\sigma_{\Sigma_t} \, g'(z) \, dz
\]

\[
= -\beta^{-1} \int_R \int_{\Sigma_t} \partial_z (\tilde{\sigma}^2 \psi^\xi) g'(z) \, dz + \int_R \tilde{b} \psi^\xi g'(z) \, dz.
\]

This shows that \( \psi^\xi \) satisfies (17). \( \square \)

2.3. A closed effective dynamics

The problem with equation (14) is that the functions \( \tilde{b} \) and \( \tilde{\sigma} \) are very complicated to compute, since they involve the full knowledge of \( \psi \). Therefore, one cannot consider (14) as a reasonable closure. A natural simplification is to consider a time-independent approximation of the functions \( \tilde{b} \) and \( \tilde{\sigma} \). Considering (15) and (16), we introduce (\( \mathbb{E}_\mu \) denoting a mean with respect to the measure \( \mu \))

\[
b(z) = \mathbb{E}_\mu \left[ (-\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi)(X) \mid \xi(X) = z \right]
\]

\[
b = \int_{\Sigma_t} (-\nabla V \cdot \nabla \xi + \beta^{-1} \Delta \xi) \, d\mu_{\Sigma_t}.
\]
Effective dynamics using conditional expectations

\[ \sigma^2(z) = \mathbb{E}_\mu \left[ |\nabla \xi|^2(X) \mid \xi(X) = z \right] \\
= \int_{\Sigma_z} |\nabla \xi|^2 \, d\mu_z, \quad (22) \]

where \( \mu_z \) is defined by (6). 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 \( \Sigma_z \). This is quantified in the following.

In the spirit of (14), we next introduce the coarse-grained dynamics

\[ dy_t = b(y_t) \, dt + \sqrt{2} \beta^{-1} \sigma(y_t) \, dB_t, \quad y_{t=0} = \xi(X_0). \quad (23) \]

The Fokker–Planck equation associated with the above dynamics will be useful. It reads as

\[ \partial_t \phi = \partial_z \left[ -b \phi + \beta^{-1} \partial_z (\sigma^2 \phi) \exp(-\beta A) \right]. \quad (24) \]

Let us first prove that dynamics (23) is ergodic for the equilibrium measure \( \xi \star \mu_z \). The distance between \( y_t \) and \( \xi(X_t) \) is estimated in section 3.

In view of assumption [H1] and of (22), we observe that the diffusion coefficient of (23) satisfies \( \sigma(y) \geq m > 0 \) for any \( y \). Hence, the process defined by (23) 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 (23). Hence, the process \( y_t \) defined by (23) is ergodic with respect to this probability (see [23, 26] and the references therein).

**Lemma 2.4.** The measure \( \xi \star \mu_z \) on \( \mathbb{R} \), which has the density \( \exp(-\beta A) \), is a stationary measure for (23).

**Proof.** We infer from (22) and (10) that

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

Using lemma 2.2 with \( \chi = |\nabla \xi|^2 \exp(-\beta V) \), we obtain

\[ \begin{align*}
\beta^{-1} \partial_z (\sigma^2 \exp(-\beta A)) &= \beta^{-1} Z^{-1} \int_{\Sigma_z} \left[ \nabla \xi \cdot \nabla (\exp(-\beta V)) + \exp(-\beta V) \Delta \xi \right] |\nabla \xi|^{-1} \, d\sigma_z \\
&= Z^{-1} \int_{\Sigma_z} \left[ -\nabla \xi \cdot \nabla V + \beta^{-1} \Delta \xi \right] \exp(-\beta V) |\nabla \xi|^{-1} \, d\sigma_z \\
&= b \exp(-\beta A).
\end{align*} \]

As a consequence of the above equation, (24) can be recast as

\[ \begin{align*}
\partial_t \phi &= \partial_z \left[ -b \phi + \beta^{-1} \partial_z (\sigma^2 \exp(-\beta A) \exp(\beta A) \phi) \right] \\
&= \beta^{-1} \partial_z \left[ \sigma^2 \partial_\phi (\exp(\beta A)) \exp(-\beta A) \right].
\end{align*} \]

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

In view of (26), we observe that \( \phi = \exp(-\beta A) \) is not only a stationary measure for (23) but also satisfies a detailed balance condition (e.g. \( y_t \) is a reversible process with respect to \( \exp(-\beta A(z)) \, dz \)). This is the subject of the next remark.
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 (26) 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

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

(27)

The above weak formulation should be compared with the weak formulation of the Fokker–Planck equation (18) associated with (3):

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

(28)

where $f = \psi \exp(\beta V)$, $\psi$ is the probability distribution function of $X_t$ satisfying (3), and $g : \mathbb{R}^n \to \mathbb{R}$ is a (time-independent) test function. We observe that (27) is (28) for functions which depend on $x$ only through $\xi(x)$.

We now discuss the relation between dynamics (23) that we propose and dynamics (5). If the function $\xi$ is such that $|\nabla \xi| = 1$, then $\sigma = 1$, and in view of (8), (9) and (21), we have $b = -A'$. Hence, in this case, the effective dynamics (23) is exactly (5). The fact that $|\nabla \xi| = 1$ is equivalent to say that $\xi$ 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 $\xi$ is such that $\sigma = 1$. Then, in view of (25), we have $b = -A'$, and again (23) is exactly (5). Note however that, in general, $\sigma$ is not a constant function, and (23) differs from (5). We will confirm in section 4 that (23) and (5) 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)

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

which is exactly $b = -A'$.

Actually, using the fact that $\xi$ 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 (see also [14, section 10 and equation (89)]) 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 \( \zeta \) is exactly the same as the one associated with \( \xi \) 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 \( \zeta \) is

\[ dy_t = -A'(y_t) \, dt + \sqrt{2} \beta^{-1} \, dB_t, \quad (29) \]

where \( A \) is the free energy associated with \( \zeta \). We hence obtain a dynamics of type (5), with an appropriate noise (that is, \( dB_t \) in (29) and \( dW_t \) in (3) are linked by (13)).

Another possibility is to keep \( \xi \) as the reaction coordinate, and to consider, instead of (3), 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 \( \mu \) is also invariant for this dynamics. Then, following the same coarse-graining procedure, based on the reaction coordinate \( \xi \), 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 with \( \xi \). This is exactly (5), again with an appropriate noise.

We conclude this discussion by noting that (see also [14]), if we make a one-to-one change of variable \( \tilde{y} = h(y) \) on dynamics (23), then we obtain exactly the same effective dynamics as if we were starting from the full dynamics (3), and were using the reaction coordinate \( \zeta = h \circ \xi \) to derive the effective dynamics. In contrast, if we make a one-to-one change of variable \( \tilde{y} = h(\tilde{y}) \) on dynamics (5), we do not obtain a dynamics of type (5), driven by the free energy associated with the reaction coordinate \( \zeta = h \circ \xi \).

3. Error estimation in terms of time marginals

In this section, we establish conditions on \( \xi \) under which the effective dynamics (23) 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 (3), and \( \phi(t, z) \) be the probability distribution function of the solution \( y_t \) to (23). 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

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

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

\[ \|\nu - \eta\|_{TV} \leq \sqrt{2H(\nu|\eta)}. \]
The second one is the Wasserstein distance with quadratic cost, which is defined, for any two probability measures $\nu$ and $\eta$ with support on a Riemannian manifold $\Sigma$, by

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

(30)

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

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

and $\Pi(v, \eta)$ denotes the set of coupling probability measures, that is probability measures $\pi$ on $\Sigma \times \Sigma$ such that their marginals are $\nu$ and $\eta$; for any test function $\Phi$,

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

In the following, we will need two functional inequalities, which we now recall [1]:

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

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

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 $\eta$ satisfies a Talagrand inequality with a constant $\rho > 0$ if, for any probability measure $\nu$,

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

We will also need the following important result (see [31, theorem 1] and [4]):

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

Logarithmic Sobolev inequalities are very useful to prove properties concerning the long-time 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, 40] for more details on this subject.

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

**Proposition 3.1.** Assume that $\xi$ satisfies [H1], and that the conditioned probability measures $\mu_{\Sigma_z}$, defined by (6), satisfy a logarithmic Sobolev inequality with a constant $\rho$ uniform in $z$; for any probability measure $v$ on $\Sigma_z$ which is absolutely continuous with respect to the measure $\mu_{\Sigma_z}$, we have

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

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

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

where $F$ is the local mean force defined by (9).
Finally, let us assume that \(|\nabla \xi|\) is close to a constant on the manifold \(\Sigma_c\) in the following sense:

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

Assume that, at time \(t = 0\), the distribution of the initial conditions of (3) and (23) are consistent one with each other: \(\psi^\xi(t = 0, \cdot) = \phi(t = 0, \cdot)\). Then we have the following estimate: for any time \(t \geq 0\),

\[
E(t) \leq \frac{M^2}{4\rho^2} \left( \lambda^2 + \frac{m^2 \beta^2 \kappa^2}{\rho^2} \right) (H(\psi(0, \cdot)|\mu) - H(\psi(t, \cdot)|\mu)), \tag{31}
\]

where \(E(t)\) is the relative entropy of the probability distribution function \(\psi^\xi\) of \(\xi(X_t)\), where \(X_t\) follows (3), with respect to the probability distribution function \(\phi\) of the solution \(y_t\) to (23):

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

Note that, for the effective dynamics to be accurate, and hence for the right-hand side of (31) to be small, the ratio \(\kappa / \rho\) needs to be small. This is a way to formalize the time-scale separation assumption that we mentioned in section 1.2.

Let us now comment on these three assumptions. We will next comment on estimate (31).

Assumption [H2] means that \(\mu_{\Sigma_c}\), which is a measure on the manifold \(\Sigma_c\), is easy to sample from. In view of (31), the interesting case is when \(\rho\) is large, and then assumption [H2] implies that there is no metastability in the manifold \(\Sigma_c\). This amounts to assuming that the overdamped dynamics with respect to \(\mu_{\Sigma_c}\) (which lives on \(\Sigma_c\)) is well mixing. Note finally that, in view of (10), the relative entropy \(H(\nu|\mu_{\Sigma_c})\) and the Fisher information \(I(\nu|\mu_{\Sigma_c})\) entering assumption [H2] read as

\[
H(\nu|\mu_{\Sigma_c}) = \int_{\Sigma_c} \ln \left( \frac{f}{\exp(-\beta V)} \right) f |\nabla \nu|^{-1} d\nu_{\Sigma_c},
\]

and

\[
I(\nu|\mu_{\Sigma_c}) = \int_{\Sigma_c} \left| \nabla_{\Sigma_c} \ln \left( \frac{f}{\exp(-\beta V)} \right) \right|^2 f |\nabla \nu|^{-1} d\nu_{\Sigma_c},
\]

where \(f\) is the density of \(\nu\) with respect to the measure \(|\nabla \nu|^{-1}\sigma_{\Sigma_c}\), i.e. \(f = d\nu / |\nabla \nu|^{-1} d\sigma_{\Sigma_c}\), and \(\nabla_{\Sigma_c}\) denotes the surface gradient:

\[
\nabla_{\Sigma_c} = P \nabla, \quad \text{where} \quad P(x) = \text{Id} - \frac{\nabla \xi \otimes \xi}{|\nabla \xi|^2}(x)
\]

is the orthogonal projector on the tangent space to \(\Sigma_c\) at point \(x \in \Sigma_c\).

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 \(F = \nabla_x V\) and \(\nabla_{\Sigma_c} F = \nabla_{x_1} V\). Requesting that \(\kappa\) is small hence amounts to requesting that \(\nabla_{x_1} V\) is small, where \(x_1\) is the reaction coordinate direction whereas \(x_2\) is the direction in \(\Sigma_c\). 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 \(\nabla_{\Sigma_c} F = 0\) if and only if the covariance \(\text{Cov}_\mu ([X_1, \ldots, X_p], (X_{p+1}, \ldots, X_n)) = 0\), 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 \(\Sigma_c\) (see also [28, 30]).

In section 3.2, we will consider an explicit example, and compute an estimation of \(\rho\) and \(\kappa\) 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 $\Sigma_i$ (which is for example the case if $\xi$ is a signed distance to a reference configuration), then $\lambda = 0$.

Before proving proposition 3.1, let us comment on estimate (31). First, this estimate is not an asymptotic result: it holds whatever the values of $\rho$ and $\kappa$. Second, note that the estimate is uniform in time. The initial conditions for (23) and (3) are such that $\phi(t = 0, \cdot) = \psi^\xi(t = 0, \cdot)$, which explains that $E(t = 0) = 0$. In the long-time limit, estimate (31) is not optimal since we know that both $\phi$ and $\psi^\xi$ 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. Our estimate hence somewhat covers the regime of intermediate times. In addition, we will show that, in some cases, it helps discriminating between several given reaction coordinates (see sections 3.2, 4.1 and 4.2).

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 (3). The probability distribution function of $\xi(X_t)$ is $\psi^\xi(t, z) = \int_{\Sigma_i} \psi(t, \cdot) |\nabla \xi|^{-1} d\sigma_{\Sigma_i}$, and satisfies

\[
\exp(-\beta A) \frac{\partial}{\partial z} \left( \psi^\xi \exp(\beta A) \right) = \int_{\Sigma_i} \nabla(\psi \exp(\beta V)) \cdot \nabla \xi |\nabla \xi|^{-1} d\sigma_{\Sigma_i} + \beta \left( A'(z) - \frac{\int F \psi |\nabla \xi|^{-1} d\sigma_{\Sigma_i}}{\psi^\xi} \right) \psi^\xi,
\]

where $A$ is the free energy (7) and $F$ is the local mean force (9).

**Proof.** Using (19), we compute

\[
e^{-\beta A} \frac{\partial}{\partial z} \left( \psi^\xi e^{\beta A} \right) = \frac{\partial}{\partial z} \psi^\xi + \beta A' \psi^\xi
\]

\[
= \int_{\Sigma_i} \left( \frac{\nabla \xi \cdot \nabla \psi}{|\nabla \xi|^2} + \text{div} \left( \frac{\nabla \xi}{|\nabla \xi|^2} \right) \right) \psi |\nabla \xi|^{-1} d\sigma_{\Sigma_i} + \beta A' \psi^\xi
\]

\[
= \int_{\Sigma_i} \nabla \xi \cdot \nabla \left( \psi \exp(\beta V) \right) |\nabla \xi|^{-1} d\sigma_{\Sigma_i} + \beta A' \psi^\xi
\]

\[
+ \int_{\Sigma_i} \left( \text{div} \left( \frac{\nabla \xi}{|\nabla \xi|^2} \right) - \beta \frac{\nabla \xi \cdot \nabla V}{|\nabla \xi|^2} \right) \psi |\nabla \xi|^{-1} d\sigma_{\Sigma_i} + \beta A' \psi^\xi,
\]

which yields (32). \qed

We are now in a position to prove proposition 3.1.

**Proof.** We know that $\phi$ satisfies the Fokker–Planck equation (26), and that $\psi^\xi$ satisfies equation (17). Thus, we have

\[
\frac{dE}{dt} = \int \frac{\partial_i \psi^\xi}{\phi} \ln \left( \frac{\psi^\xi}{\phi} \right) - \int \frac{\partial_i \phi}{\phi} \psi^\xi
\]

\[
= \int \frac{\partial_i \left( -\tilde{b} \psi^\xi + \beta^{-1} \partial_i \left( \sigma_\xi \phi \right) \phi \psi^\xi \right)}{\phi} \ln \left( \frac{\psi^\xi}{\phi} \right)
\]

\[- \beta^{-1} \int \frac{\sigma_\xi^2 \partial_i (\phi \exp(\beta A)) \exp(-\beta A)}{\phi} \psi^\xi \phi
\]

\[
= - \int \frac{\tilde{b} \psi^\xi + \beta^{-1} \partial_i \left( \sigma_\xi^2 \phi \psi^\xi \right)}{\phi} \partial_i \ln \left( \frac{\psi^\xi}{\phi} \right)
\]

\[+ \beta^{-1} \int \sigma_\xi^2 \partial_i (\phi \exp(\beta A)) \exp(-\beta A) \partial_i \left( \frac{\psi^\xi}{\phi} \right).
\]
Using (20), we have

\[
\partial_z(\tilde{\sigma}^2 \psi^z) = \int_{\Sigma_z} (\nabla \xi \cdot \nabla \psi + \psi \Delta \xi) |\nabla \xi|^2 \, d\sigma_{\Sigma_z}
\]

\[
= \int_{\Sigma_z} (\nabla \xi \cdot \nabla (\psi \exp(\beta V)) \exp(-\beta V)) |\nabla \xi|^2 \, d\sigma_{\Sigma_z}
\]

\[
+ \int_{\Sigma_z} (-\beta \nabla \xi \cdot \nabla V + \Delta \xi \psi |\nabla \xi|^2 \, d\sigma_{\Sigma_z}
\]

\[
= \int_{\Sigma_z} (\nabla \xi \cdot \nabla (\psi \exp(\beta V)) \exp(-\beta V)) |\nabla \xi|^2 \, d\sigma_{\Sigma_z} + \tilde{\beta}(t, z) \psi^z(t, z).
\]

Thus, it holds

\[
\frac{dE}{dt} = -\beta^{-1} \int_{\Sigma_z} (\nabla \xi \cdot \nabla (\psi \exp(\beta V)) \exp(-\beta V)) |\nabla \xi|^2 \, d\sigma_{\Sigma_z} \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

\[
+ \beta^{-1} \int_{\Sigma_z} \psi^z \partial_z (\partial_x \exp(\beta A)) \exp(-\beta A) \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

\[
= -\beta^{-1} \int_{\Sigma_z} \left( \nabla \xi \cdot \nabla (\psi \exp(\beta V)) \exp(-\beta V) \right) |\nabla \xi|^2 \, d\sigma_{\Sigma_z} \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

\[
- \beta^{-1} \int_{\Sigma_z} \psi^z \partial_z (\partial_x \exp(\beta A)) \exp(-\beta A) \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

\[
+ \beta^{-1} \int_{\Sigma_z} \partial_z (\partial_x \exp(\beta A)) \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

We next use (32) to get

\[
\frac{dE}{dt} = -\beta^{-1} \int_{\Sigma_z} \left( \nabla \xi \cdot \nabla (\psi \exp(\beta V)) \exp(-\beta V) \right) |\nabla \xi|^2 \, d\sigma_{\Sigma_z} \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

\[
- \beta^{-1} \int_{\Sigma_z} \psi^z \partial_z (\partial_x \exp(\beta A)) \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

\[
+ \beta^{-1} \int_{\Sigma_z} \partial_z (\partial_x \exp(\beta A)) \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

\[
= -\beta^{-1} \int_{\Sigma_z} \left( \nabla \xi \cdot \nabla (\psi \exp(\beta V)) \exp(-\beta V) \right) |\nabla \xi|^2 \, d\sigma_{\Sigma_z} \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

\[
+ \beta^{-1} \int_{\Sigma_z} \partial_z (\partial_x \exp(\beta A)) \partial_z \ln \left( \frac{\psi^z}{\phi} \right)
\]

\[
- \beta^{-1} \int_{\Sigma_z} \partial_z \ln \left( \frac{\psi^z}{\phi} \right)^2.
\]
We now use two Young inequalities, with \( \varepsilon_1 > 0 \) and \( \varepsilon_2 > 0 \) to be fixed later on:

\[
\frac{dE}{dr} \leq \frac{\beta - 1}{2\varepsilon_1} \int_{\Sigma} \left( \nabla \cdot \left( \nabla (\psi e^{\beta V}) \right) - \psi \right) \left| \nabla \xi \right|^2 \leq \frac{1}{\sigma^2 \psi^2}
\]

\[
\frac{\beta}{2\varepsilon_2} \int \sigma^2 \left( A'(z) - \frac{\int_{\Sigma} F \psi \left| \nabla \xi \right|^{-1} d\sigma_z}{\psi} \right)^2 \psi^2
\]

\[
- \beta^{-1} \left( 1 - \frac{\varepsilon_1 + \varepsilon_2}{2} \right) \int \sigma^2 \psi \left| \partial_z \ln \left( \frac{\psi}{\phi} \right) \right|^2.
\]

Let us first consider the second term of (33). Introducing \( \psi_{|z} \), which is the measure \( \psi(t,x) dx \) conditioned to \( \xi(x) = z \),

\[
d\psi_{|z} = \frac{\psi \left| \nabla \xi \right|^{-1} d\sigma_z}{\psi^2},
\]

we write, using (8) and [H3], that

\[
\left( A'(z) - \frac{\int_{\Sigma} F \psi \left| \nabla \xi \right|^{-1} d\sigma_z}{\psi} \right)^2 = \left( \int_{\Sigma} F d\mu_{\xi} - \int_{\Sigma} F d\psi_{\xi} \right)^2
\]

\[
= \left( \int_{\Sigma} F(\alpha) \pi(\alpha, d\gamma) - \int_{\Sigma} F(\gamma) \pi(\alpha, d\gamma) \right)^2
\]

\[
\leq \left\| \nabla_{\xi} F \right\|_{2}^2 \left( \int_{\Sigma} \pi(\alpha, \gamma) \right)^2
\]

\[
\leq \left\| \nabla_{\xi} F \right\|_{2}^2 W(\psi_{\xi}, \mu_{\xi}),
\]

where \( \pi \in \Pi(\mu_{\xi}, \psi_{\xi}) \) is any coupling measure (the last inequality has been obtained by taking the infimum over such \( \pi \), see equation (30)). Since \( \mu_{\xi} \) satisfies a logarithmic Sobolev inequality (assumption [H2]), it also satisfies a Talagrand inequality (see lemma 3.1), hence

\[
W(d\psi_{\xi}, d\mu_{\xi}) \leq \frac{2}{\rho} H(d\psi_{\xi}, d\mu_{\xi}) \leq \frac{1}{\rho^2} I(d\psi_{\xi}, d\mu_{\xi}).
\]

Gathering the above inequality with (34), we obtain

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

Using [H1], we thus bound the second term of (33):

\[
\int_{\mathbb{R}} \sigma^2 \left( A'(z) - \frac{\int_{\Sigma} F \psi \left| \nabla \xi \right|^{-1} d\sigma_z}{\psi} \right)^2 \psi \leq \frac{M^2 \kappa^2}{\rho^2} \int_{\mathbb{R}} I(d\psi_{\xi}, d\mu_{\xi}) \psi^2
\]

\[
= \frac{M^2 \kappa^2}{\rho^2} \int_{\mathbb{R}} \left| \nabla_{\xi} \ln \left( \frac{\psi}{\exp(-\beta V)} \right) \right|^2 \psi \left| \nabla \xi \right|^{-1} d\sigma_z
\]

\[
= \frac{M^2 \kappa^2}{\rho^2} \int_{\mathbb{R}} \nabla_{\xi} \ln \left( \frac{\psi}{\exp(-\beta V)} \right)^2 \psi.
\]
We now bound the first term of (33) using a Cauchy–Schwarz inequality, [H4] and [H1]:

\[
\int \int_{\Sigma_t} \left( \frac{\nabla \psi \cdot \nabla (\psi \exp(\beta V))}{|\nabla \psi|^2} \right) \exp(-\beta V) \left( |\nabla \psi|^2 - \sigma^2(z) \right) |\nabla \psi^{-1}|^{-1} \, d\sigma_{\Sigma_t} \frac{1}{\sigma^2 \psi^2} \\
= \int \int_{\Sigma_t} \frac{\nabla \psi \cdot \nabla \ln(\psi \exp(\beta V))}{|\nabla \psi|^2} \left( |\nabla \psi|^2 - \sigma^2(z) \right) \psi \, |\nabla \psi^{-1}|^{-1} \, d\sigma_{\Sigma_t} \frac{1}{\sigma^2} \\
\leq \int \int_{\Sigma_t} \frac{\nabla \psi \cdot \nabla \ln(\psi \exp(\beta V))}{|\nabla \psi|^2} \left( |\nabla \psi|^2 - \sigma^2(z) \right) \psi \, |\nabla \psi^{-1}|^{-1} \, d\sigma_{\Sigma_t} \frac{1}{\sigma^2} \\
\leq \lambda^2 \int \int_{\mathbb{R}^2} \frac{\nabla \psi \cdot \nabla \ln(\psi \exp(\beta V))}{|\nabla \psi|^2} \psi \, |\nabla \psi^{-1}|^{-1} \, d\sigma_{\Sigma_t} \, \sigma^2 \\
\leq \lambda^2 M^2 \int_{\mathbb{R}^2} \frac{\nabla \psi \cdot \nabla \ln(\psi \exp(\beta V))}{|\nabla \psi|^2} \psi. \tag{36}
\]

We infer from (33) and the bounds (35) and (36) that

\[
\frac{dE}{dt} \leq \frac{\beta^{-1}}{2 \epsilon_1} \lambda^2 M^2 \int_{\mathbb{R}^2} \frac{\nabla \psi \cdot \nabla \ln(\psi \exp(\beta V))}{|\nabla \psi|^2} \psi \\
+ \frac{\beta}{2 \epsilon_2} \frac{M^2 \kappa^2}{\rho^2} \int_{\mathbb{R}^2} |\nabla \ln(\psi \exp(\beta V))|^2 \psi \\
- \beta^{-1} \left( 1 - \frac{\epsilon_1 + \epsilon_2}{2} \right) \int \sigma^2 \psi \frac{d}{d\xi} \ln \left( \frac{\psi \xi}{\phi} \right)^2.
\]

Note that

\[
|\nabla \ln(\psi \exp(\beta V))|^2 = \left| \frac{\nabla \psi \cdot \nabla \ln(\psi \exp(\beta V))}{|\nabla \psi|} \right|^2 + \left| \nabla \ln(\psi \exp(\beta V)) \right|^2.
\]

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

\[
\frac{dE}{dt} \leq \frac{\beta^{-1}}{2 \epsilon_1} \lambda^2 M^2 \int_{\mathbb{R}^2} |\nabla \ln(\psi \exp(\beta V))|^2 \psi + \frac{\beta}{2 \epsilon_2} \frac{M^2 \kappa^2}{\rho^2} \int_{\mathbb{R}^2} |\nabla \ln(\psi \exp(\beta V))|^2 \psi \\
- \beta^{-1} \left( 1 - \frac{\epsilon_1 + \epsilon_2}{2} \right) \int \sigma^2 \psi \frac{d}{d\xi} \ln \left( \frac{\psi \xi}{\phi} \right)^2.
\]

We now optimize on $\epsilon_1$ and $\epsilon_2$ by choosing them such that $\epsilon_1 + \epsilon_2 = 2$ and $(\beta^{-1}/2\epsilon_1)(\lambda^2 M^2/\rho^2) = (\beta/2\epsilon_2)(M^2 \kappa^2/\rho^2)$. This yields $\epsilon_1 = 2 \lambda^2 \rho^2 / (\lambda^2 \rho^2 + m^2 \beta^2 \kappa^2)$, thus

\[
\frac{dE}{dt} \leq \frac{\beta^{-1} M^2}{4m^2} \left( \frac{\lambda^2 + m^2 \beta^2 \kappa^2}{\rho^2} \right) \int_{\mathbb{R}^2} |\nabla \ln(\psi \exp(\beta V))|^2 \psi \\
= \frac{\beta^{-1} M^2}{4m^2} \left( \frac{\lambda^2 + m^2 \beta^2 \kappa^2}{\rho^2} \right) I(\psi | \mu) \\
= - \frac{M^2}{4m^2} \left( \frac{\lambda^2 + m^2 \beta^2 \kappa^2}{\rho^2} \right) \frac{d}{dt} H(\psi | \mu).
\]

We next integrate this equation between 0 and $t$ and use the fact that $E(0) = 0$ to obtain (31).

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

Consider again the probability distribution function $\psi^\xi$ of $\xi(X_t)$, where $X_t$ follows (3), and the probability distribution function $\phi$ of the solution $\gamma$, to (23). They satisfy

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

for some positive constant $R$ (which depends on $m$, $\rho$, $\kappa$, and $r$), where

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

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

where $\mu^\xi$ is given by (38). We also have

$$\frac{d\mu^\xi}{dz} = \exp(-\beta A(z)) dz.$$

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 (31) that

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

where $C_1(t)$ is given by (38). We also have

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

where $d\mu^\xi = \exp(-\beta A(z)) dz$ is the equilibrium measure $\xi \ast \mu$. Let us first upper-bound $E_{CG}(t) = H(\phi|\mu^\xi) = \int_{\mathbb{R}} \ln \left( \frac{\phi}{\exp(-\beta A)} \right) \phi$. Using (26), we compute

$$\frac{dE_{CG}}{dt} = \int_{\mathbb{R}} \partial_t \ln \left( \frac{\phi}{\exp(-\beta A)} \right) \phi$$

$$= \beta^{-1} \int_{\mathbb{R}} \partial_t \left( \sigma^2 \partial_z (\phi \exp(\beta A)) \exp(-\beta A) \right) \ln \left( \frac{\phi}{\exp(-\beta A)} \right)$$

$$= -\beta^{-1} \int_{\mathbb{R}} \left( \sigma^2 \partial_z (\phi \exp(\beta A)) \exp(-\beta A) \right) \partial_z \left( \ln \left( \frac{\phi}{\exp(-\beta A)} \right) \right)^2$$

$$\leq -m^2 \beta^{-1} \int_{\mathbb{R}} \partial_z \left( \ln \left( \frac{\phi}{\exp(-\beta A)} \right) \right)^2$$

$$= -m^2 \beta^{-1} \int_{\mathbb{R}} \partial_z \left( \ln \left( \frac{\phi}{\exp(-\beta A)} \right) \right)^2$$

$$\leq -m^2 \beta^{-1} \int_{\mathbb{R}} \partial_z \left( \ln \left( \frac{\phi}{\exp(-\beta A)} \right) \right)^2$$

$$\leq -m^2 \beta^{-1} \int_{\mathbb{R}} \partial_z \left( \ln \left( \frac{\phi}{\exp(-\beta A)} \right) \right)^2$$

$$\leq -m^2 \beta^{-1} \int_{\mathbb{R}} \partial_z \left( \ln \left( \frac{\phi}{\exp(-\beta A)} \right) \right)^2$$

where we have used that $\sigma^2 \geq m^2$, which is a consequence of [H1] and (22). Since $\mu^\xi$ satisfies a logarithmic Sobolev inequality with constant $r$, we infer from the above bound that

$$\frac{dE_{CG}}{dt} \leq -2m^2 \beta^{-1} E_{CG}. $$

Using a Gronwall lemma, we obtain

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

$$= H(\phi(0, \cdot)|\mu^\xi) \exp(-2m^2 \beta^{-1} t),$$

and the Csiszár-Kullback inequality then yields

$$\|\phi - \mu^\xi\|_{TV} \leq \sqrt{2H(\phi|\mu^\xi)} \leq \sqrt{2H(\phi(0, \cdot)|\mu^\xi)} \exp(-m^2 \beta^{-1} t). \quad (42)$$

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

$$\int_{\mathbb{R}} \chi(x) \, dx = \int_{\mathbb{R}} \int_{\Sigma} \frac{|\chi|}{\sqrt{\nabla \xi}} \cdot d\sigma_{\Sigma} \, dz \geq \int_{\mathbb{R}} \int_{\Sigma} \frac{\chi}{\sqrt{\nabla \xi}} \cdot d\sigma_{\Sigma} \, dz = \int_{\mathbb{R}} \chi \, dz.$$
which also reads \( \| x \|_{TV} \geq \| x^\xi \|_{TV} \). We apply this inequality with \( x = \psi - \mu \):

\[
\| \psi^\xi - \mu^\xi \|_{TV} \leq \| \psi - \mu \|_{TV} \leq \sqrt{2H(\psi|\mu)}.
\]

Since \( \mu^\xi \) and the conditional measures \( \mu_{\Sigma^z} \) satisfy a logarithmic Sobolev inequality (see [H5] and [H2]), and under assumptions [H1] and [H3], we obtain that the measure \( \mu \) also satisfies a logarithmic Sobolev inequality with some constant \( \bar{R} > 0 \), which depends on \( m, \rho, \kappa \) and \( r \) (see [28, equation (9)]). Hence, by a computation similar to the one on \( E_{\text{CG}} \), we obtain

\[
H(\psi|\mu) \leq H(\psi(0, \cdot)|\mu) \exp(-2\bar{R} \beta^{-1} t),
\]

hence

\[
\| \psi^\xi - \mu^\xi \|_{TV} \leq \sqrt{2H(\psi(0, \cdot)|\mu)} \exp(-\bar{R} \beta^{-1} t).
\]  

(43)

Gathering (41), (42) and (43), we obtain

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

where \( C_2 \) is defined by (39). Setting \( R = \min(\bar{R}, r m^2) \), the above bound then yields

\[
\| \psi^\xi - \phi \|_{TV} \leq 2C_2 \exp(-R \beta^{-1} t),
\]

which, gathered with (40), yields (37). \( \square \)

3.2. Estimation of the upper-bound constants of (31) in a particular case

In this section, we give a very formal argument to estimate the constants \( \rho \) and \( \kappa \) entering the bound (31), 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 [35], reads as

\[
V(x) = \sum_i V_2(d_{i,i+1}) + \sum_i V_3(\theta_i) + \sum_i V_4(\phi_i) + V_{\text{non-bonded}}(x),
\]

where \( d_{i,i+1} \) is the distance between atoms \( i \) and \( i + 1 \), \( \theta_i \) is the bond angle made by atoms \( i - 1, i \) and \( i + 1 \), and \( \phi_i \) 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

\[
V_\varepsilon(x) = V_0(x) + \frac{1}{\varepsilon} q^2(x),
\]  

(44)

where \( V_0 \) and \( q \) are two scalar-valued functions that do not depend on the small parameter \( \varepsilon \) (see equation (47) and figure 1 for a precise example of type (44)). For simplicity, we assume here that the reaction coordinate \( \xi \) does not depend on \( \varepsilon \), and that its gradient is constant on the manifolds \( \Sigma_{\varepsilon} \) (in assumption [H4], \( \lambda = 0 \)). Since the relative entropy is always non-negative, estimate (31) reads as

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

We also assume that the initial condition of (3) is well adapted to the Boltzmann measure \( \mu_{\varepsilon} \), in the sense that \( H(\psi(0, \cdot)|\mu_{\varepsilon}) \) is upper-bounded by a constant independent of \( \varepsilon \). Thus the above bound reads as

\[
E(t) \leq C \frac{\kappa^2}{\rho_\varepsilon^2}
\]  

(45)

for some constant \( C \) independent of \( \varepsilon \). Our aim is to roughly estimate the coefficients \( \rho_\varepsilon \) and \( \kappa_\varepsilon \) in terms of \( \varepsilon \).
We first compute the local mean force, defined by (9):

\[
F = \frac{\nabla V_\varepsilon \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \text{div} \left( \frac{\nabla \xi}{|\nabla \xi|^2} \right) = \frac{2}{\varepsilon} \frac{\nabla q \cdot \nabla \xi}{|\nabla \xi|^2} + \frac{\nabla V_0 \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \text{div} \left( \frac{\nabla \xi}{|\nabla \xi|^2} \right).
\]

Recall that \(\xi\) does not depend on \(\varepsilon\). If \(\nabla q \cdot \nabla \xi \neq 0\), then \(F\) is of order \(1/\varepsilon\), and so is \(\kappa_\varepsilon\). In contrast, if \(\nabla q \cdot \nabla \xi = 0\), then \(F\) is of order 1 with respect to \(\varepsilon\), and so is \(\kappa_\varepsilon\).

The measure \(\mu_{\Sigma}\) is assumed to satisfy a logarithmic Sobolev inequality (assumption [H2]) with the constant \(\rho_\varepsilon\), which we wish to bound from below. In view of (44), we see that, typically, \(\rho_\varepsilon\) remains bounded from below when \(\varepsilon\) goes to zero. We hence see that the right-hand side of (45) is much smaller (and so the effective dynamics is more accurate) when \(\nabla q \cdot \nabla \xi = 0\) than when \(\nabla q \cdot \nabla \xi \neq 0\).

Note that, in some particular cases, one can show that \(\rho_\varepsilon\) is actually not only bounded from below but also of order \(1/\varepsilon\), when \(\varepsilon\) goes to zero. This is typically the case when \(\Sigma \cap \{x; q(x) = 0\}\) contains a single point, on which the measure \(\mu_{\Sigma}\) concentrates as \(\varepsilon\) goes to zero. Hence, as \(\varepsilon\) decreases to 0, the effective dynamics (23) becomes more accurate, in the sense of [D3].

In the next section, we numerically confirm the above discussion, and show that the criterion

\[
\nabla \xi \cdot \nabla q = 0
\]

has indeed a significant impact on the accuracy of the effective dynamics, in the case when the potential energy is of type (44).

4. Numerical results: residence time estimation

Our aim here is twofold. First, we want to check the accuracy of the effective dynamics (23) in terms of residence times, and also compare this effective dynamics with the coarse-grained dynamics (5) based on the free energy. Second, we wish to assess the relevance of criterion (46). In view of section 3.2, it seems to be an important condition for the effective dynamics to be accurate. We would like to test this numerically.

In the following numerical tests, we focus on residence times. We have indeed already underlined that the characteristic behaviour of dynamics (3) is to sample a given well of the potential energy, then suddenly hope 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 several numerical examples where we have studied such quantities.

4.1. A simple two-dimensional example

Consider the potential energy

\[
V_\varepsilon(x, y) = (x^2 - 1)^2 + \frac{1}{\varepsilon} (x^2 + y - 1)^2,
\]

which is of form (44), with \(V_0(x, y) = (x^2 - 1)^2\) and \(q(x, y) = x^2 + y - 1\). For any \(\varepsilon > 0\), the potential \(V_\varepsilon\) 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 \(V_\varepsilon\) 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 dynamics (3)) with the residence time computed from a coarse-grained description, according to (23) or (5), for two different reaction coordinates.

In the case at hand, a natural reaction coordinate is \( \xi_1(x, y) = x \), since the value of \( \xi_1 \) 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 (23) is the same as dynamics (5), that is the dynamics driven by the free energy \( A_1 \) associated with \( \xi_1 \). This free energy reads as

\[
A_1(z) = (z^2 - 1)^2 + C(\beta) \tag{48}
\]

for some constant \( C(\beta) \) ensuring that \( \int_{\mathbb{R}} \exp(-\beta A_1(z)) \, dz = 1 \). Note that \( \nabla \xi_1 \cdot \nabla q \neq 0 \). In view of the discussion of section 3.2, we do not expect the effective dynamics based on \( \xi_1 \) 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 (23), based on \( \xi_2 \), 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 (23) differs from dynamics (5) for \( A \equiv A_2 \), the free energy associated with \( \xi_2 \).

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

With the choice we made for \( \beta \) and \( \varepsilon \), the system is metastable. In figure 3, we plot \( x_t \) as a function of time, where \( X_t = (x_t, y_t) \) satisfies (3). 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.

The functions \( b \) and \( \sigma \), as well as the derivative of the free energy \( A_2 \) (defined by (21), (22) and (8), respectively) are plotted in figure 4, in the case of the reaction coordinate \( \xi_2 \).

Remark 4.1. For all the numerical tests reported in this paper, the complete dynamics (3) 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.
Figure 2. Crosses: plot of the trajectory $X_t = (x_t, y_t)$ solution to (3), for the parameters $\varepsilon = 0.01$ and $\beta = 3$. Dashed lines: level sets of $\xi_2$.

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

Figure 4. Plot of the functions $b$, $\sigma$ and $A_2'$, for the reaction coordinate $\xi_2$. Note that $b$ and $A_2'$ are odd functions, whereas $\sigma$ is an even function. Note the large variations of $A_2'$ in the neighbourhood of $z = 0$.

For the reaction coordinate $\xi_1$, the effective dynamics is (5), which we have numerically simulated with the same algorithm as above. We have used the analytical expression (48) of the free energy $A_1$.

For the reaction coordinate $\xi_2$, the free energy derivative $A_2'$ and the functions $b$ and $\sigma$ have been computed using the algorithm proposed in [9]. We considered the interval...
A reaction coordinate \( \xi \) can be chosen that is close to one of the reaction coordinates \( x \) or \( y \) of the complete model (2).

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

(i) we first generated 15 000 configurations \( \{(x_i, y_i) \in \mathbb{R}^2\}_{1 \leq i \leq 15000} \), distributed according to the measure \( \mu \) restricted to the right well: as a consequence, \( \xi(x_i, y_i) > \xi_1^b \).

(ii) we next ran dynamics (3) from the initial condition \( (x_i, y_i) \), and monitor the first time \( \tau_i \) at which the system reaches a point \( (x(\tau_i), y(\tau_i)) \) in the left well: \( \tau_i = \inf\{t; \xi(x(t), y(t)) < -\xi_1^b\} \).

(iii) from these \( (\tau_i)_{1 \leq i \leq 15000} \), we computed an average residence time and a confidence interval.

These figures are the reference figures.

(iv) we next consider the initial conditions \( \{(\xi(x_i, y_i) \in \mathbb{R}\}_{1 \leq i \leq 15000} \) for the effective dynamics.

By construction, these configurations are distributed according to the equilibrium measure \( \xi \ast \mu \) (that is \( \exp(-\beta A(z)) \, dz \)) restricted to the right well.

(a) from these initial conditions, we run dynamics (23) or (5) until the left well is reached, 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 \( \xi_1 \) (which satisfies the condition \( \nabla \xi_1 \cdot \nabla q = 0 \) and with the effective dynamics (23), we can reproduce the reference residence time (32.5 ± 0.5) within an excellent accuracy (see first line of table 1). If we still use the reaction coordinate \( \xi_2 \), but consider as the coarse-grained dynamics dynamics (5) 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.

Note also that the results obtained with the reaction coordinate \( \xi_1 \), which is such that \( \nabla \xi_1 \cdot \nabla q \neq 0 \), are inconsistent with the reference results (in that case, the effective dynamics (23) is the same as (5)). Actually, the coarse-grained dynamics does not depend on \( \varepsilon \) (since the free energy \( A_1 \) does not depend on \( \varepsilon \)), whereas the complete description does depend on \( \varepsilon \).

We hence observe that, in the example at hand here, good results are obtained only in the case when the effective dynamics is (23) (rather than (5)) and the reaction coordinate satisfies criterion (46).
4.2. A three atom molecule

Our aim in this section is to show that different reaction coordinates, although similar at first sight, can lead to very different results. Criterion (46) turns out to be useful to choose between them.

We consider here a molecule made of three two-dimensional particles, whose positions are $x_A$, $x_B$, and $x_C$. The potential energy of the system is

$$V(x) = \frac{1}{2\varepsilon} (r_{AB} - \ell_{eq})^2 + \frac{1}{2\varepsilon} (r_{BC} - \ell_{eq})^2 + W_3(\theta_{ABC}),$$

(49)

where $r_{AB} = \|x_A - x_B\|$ is the distance between atoms A and B, $\ell_{eq}$ is an equilibrium distance, $\theta_{ABC}$ is the angle formed by the three atoms and $W_3(\theta)$ is a three-body potential, which we choose here to be a double-well potential:

$$W_3(\theta) = \frac{k_\theta}{2} (\theta - \theta_{saddle})^2 - \delta \theta^2.$$

Wells of $W_3$ are located at $\theta = \theta_{saddle} \pm \delta \theta$. Potential (49) represents stiff bonds between particles A and B on the one hand, and B and C on the other hand, with a softer term depending on the angle $\theta_{ABC}$. To remove rigid body motion invariance, we set $x_B = 0$ and $x_A \cdot e_y = 0$. In the following, we work with the parameters $\varepsilon = 10^{-3}$, $k_\theta = 208$, $\ell_{eq} = 1$, $\theta_{saddle} = \pi/2$ and $\delta \theta = \theta_{saddle} - 1.187$. The temperature is $\beta^{-1} = 1$, and all the dynamics are integrated with the time step $\Delta t = 10^{-3}$.

We consider two reaction coordinates, both of which indicate in which well the system is: $\xi_1 = \theta_{ABC}$ and $\xi_2 = \|x_A - x_C\|^2$. Note that (49) reads as

$$V(x) = \frac{1}{2\varepsilon} (q_{AB}(x))^2 + q_{BC}(x)^2 + W_3(\theta_{ABC})$$

with $q_{AB}(x) = r_{AB} - \ell_{eq}$ and $q_{BC}(x) = r_{BC} - \ell_{eq}$. The two first terms are much stiffer than the last one. We observe that $\nabla V_{\theta_{ABC}} \cdot \nabla q_{AB} = \nabla V_{\theta_{ABC}} \cdot \nabla q_{BC} = 0$. Hence, the reaction coordinate $\xi_1$ is orthogonal to the stiff terms of the potential energy, in contrast to $\xi_2$. As in the previous section, we compute the residence time in a given well following the complete description, and compare it with the result given by a reduced description, based either on (23) or (5). Results are gathered in table 2. We observe that working with $\xi_1$ (and either (23) or (5)) leads to very accurate results. On the other hand, as in the previous example, when the reaction coordinate is not orthogonal to the stiff terms of the potential, coarse-grained dynamics turn out to be not accurate.

In the case at hand here, $\|\nabla \xi_1\|^2 = \|\nabla \theta_{ABC}\|^2 = \ell_{eq}^{-2} r_{BC}^{-2}$. This quantity is almost a constant, since the bond length potential is stiff and the temperature is small. Hence, along the trajectory, we have that $\|\nabla \xi_1\|^2 \approx \ell_{eq}^{-2} = 1$. This explains why, when choosing the reaction coordinate $\xi_1$, dynamics (23) and (5) give similar results.

### Table 2. Residence times obtained from the complete description (second column) and from the reduced descriptions (two last columns), for both reaction coordinates.

| Reaction coordinate | Reference residence time | Residence time using (23) | Residence time using (5) |
|---------------------|--------------------------|---------------------------|--------------------------|
| $\xi_1 = \theta_{ABC}$ | $0.700 \pm 0.011$        | $0.704 \pm 0.011$         | $0.710 \pm 0.011$        |
| $\xi_2 = r_{BC}$     | $0.709 \pm 0.015$        | $0.219 \pm 0.004$         | $2.744 \pm 0.056$        |

4.3. A simple model of conformational change in solution

In this section, we consider a two-dimensional system composed of $N$ particles in a periodic box of side length $\ell$. Two of these particles form a solute dimer, whereas the others are solvent
Table 3. Residence times obtained from the complete description (second column) and from the reduced descriptions (two last columns).

| Inverse temperature | Reference residence time | Residence time using (23) | Residence time using (5) |
|---------------------|--------------------------|---------------------------|--------------------------|
| $\beta = 0.5$       | 262 ± 6                  | 245 ± 5                   | 504 ± 11                 |
| $\beta = 0.25$      | 1.81 ± 0.04              | 1.68 ± 0.04               | 3.47 ± 0.08              |

particles [11, 38]. The potential energy is

$$V(x) = V_S(\|x_1 - x_2\|) + \sum_{1 \leq i < j \leq N, (i,j) \neq (1,2)} V_{WCA}(\|x_i - x_j\|),$$

(50)

where the purely repulsive WCA pair potential is defined by

$$V_{WCA}(r) = 4\varepsilon \left(\frac{\sigma^{12}}{r^{12}} - 2 \frac{\sigma^6}{r^6}\right)$$

if $r \leq 2^{1/6}\sigma$, $V_{WCA}(r) = 0$ otherwise.

The two first particles interact through a double-well type potential,

$$V_{S}(r) = h \left(1 - \frac{(r - r_0 - w)^2}{w^2}\right)^2,$$

which has two local minima at $r = r_0$ and $r = r_0 + 2w$, separated by a barrier of height $h$. We work with the bond length of the dimer as the reaction coordinate: $\xi(x) = \|x_1 - x_2\|$. As in the previous sections, we compute the residence times in a given well following the complete description, for two different temperatures, and compare them with the results given by a reduced description, based either on (23) or (5). We consider a system of $N = 16$ particles, in a box of side length $\ell = 8$. The potential parameters are $\varepsilon = \sigma = 1$, $h = 20$, $r_0 = 2^{1/6}\sigma$ and $w = 0.7$. We used the time step $\Delta t = 10^{-4}$ to integrate the dynamics. Results are gathered in table 3. Again, we observe that residence times are computed with a very good accuracy when using the effective dynamics (23).

5. Pathwise convergence

In this section, we prove pathwise convergence results between $\xi(X_t)$, where $X_t$ solves (3), and $y_t$ which solves (23), for some potential energies of type (44). On these specific examples, we obtain stronger convergence results than in the previous sections (namely, convergence in the sense of [D1]).

Consider dynamics (3), with the potential energy $V_\varepsilon$ defined by (44). It reads as

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

(51)

Note that the initial condition is supposed to not depend on $\varepsilon$. The limit of $X^\varepsilon_t$ when $\varepsilon \to 0$ has been identified in [9]: it is a process $\bar{X}_t$ solution of a SDE that we write below (see equation (54)), and that is such that $q(\bar{X}_t) = 0$ for any $t$.

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

$$\mathcal{M} = \{x \in \mathbb{R}^2; q(x) = 0\}.  

(52)

Assume also that the reaction coordinate $\xi$ 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, $\xi$ parametrizes $\mathcal{M}$). In that case, it is easy to build a reduced dynamics from (51), in the limit $\varepsilon \to 0$: one first lets $\varepsilon$ go to zero, writes the dynamics of $\bar{X}_t$, and then makes a one-to-one change of variable to write the dynamics in term of $\xi(\bar{X}_t)$.
Our aim is to write conditions under which the so-obtained dynamics corresponds to (23), which amounts to say that diagram (53) is a commutative diagram.

$$2\text{D dynamics (51) on } X_t^\varepsilon \rightarrow_{\varepsilon \to 0} \begin{bmatrix} \text{pathwise convergence} \end{bmatrix} \text{ID limit dynamics (54) on } \overline{X}_t$$

\[ \text{Itô computation} \]

\[ \downarrow \]

\[ \text{Nonclosed dynamics on } \xi(X_t^\varepsilon) \]

\[ \downarrow \]

\[ \text{Conditional expectations} \]

\[ \downarrow \]

\[ \text{Dynamics (23) on } y_t^\varepsilon \approx \xi(X_t^\varepsilon) : \rightarrow_{\varepsilon \to 0} \text{Dynamics (60) on } z_t \]

\[ dy_t^\varepsilon = b_\varepsilon(y_t^\varepsilon) \, dt + \sigma_\varepsilon(y_t^\varepsilon) \, dB_t \]

5.1. Limit of (51) in a pathwise sense

We now proceed in detail. For any \( x \in \mathcal{M} \), let

\[ P(x) = \text{Id} - \frac{\nabla q \otimes \nabla q}{|\nabla q|^2}(x) \]

be the projector on \( T_x\mathcal{M} \), the tangent space to \( \mathcal{M} \) at \( x \). Let us define

\[ n = \frac{\nabla q}{|\nabla q|} \quad \text{and} \quad \kappa = \text{div } n. \]

Let us now introduce the process \( \overline{X}_t \), solution to the equation

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

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 [9]), under some regularity assumptions on \( q \) and \( V_0 \), there exists a constant \( C \) that does not depend on \( \varepsilon \) such that

\[ \sup_{t \in [0, T]} E \left| X_t^\varepsilon - \overline{X}_t \right|^2 \leq C\varepsilon. \quad (55) \]

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

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

\[ \chi : x \in \mathbb{R}^2 \mapsto (\xi(x), q(x)), \]

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

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

After some tedious but not difficult computations, the above dynamics can be written as

\[ d(\xi(\overline{X}_t)) = d_1(\overline{X}_t) \, dt + d_2(\overline{X}_t) \, dt + \sqrt{2\beta^{-1}} |\nabla \xi(\overline{X}_t)| \, dW_t, \]

with \( dW_t = (\nabla \xi / |\nabla \xi|)(\overline{X}_t) \cdot dW_t \), and

\[ d_1 = -\nabla \xi \cdot \nabla V_0 + \beta^{-1} \Delta \xi, \]

\[ d_2 = -\frac{1}{\beta} \frac{\nabla q \cdot \nabla u}{|\nabla q|^2} + u \frac{\nabla q \cdot \nabla V_0}{|\nabla q|^2} - \frac{1}{\beta} \kappa \frac{u}{|\nabla q|} + \frac{1}{\beta} \frac{\nabla q^T \nabla^2 q \nabla q}{|\nabla q|^4}. \]

(58)
where we set
\[ u = \nabla \xi \cdot \nabla q. \] (59)

Since \( X_t \) satisfies the constraint \( q(X_t) = 0 \), dynamics (57) can be rewritten only in terms of \( \xi(X_t) = z_t \), in the form
\[ dz_t = \tilde{d}_1(z_t) \, dt + \tilde{d}_2(z_t) \, dt + \sqrt{2\beta^{-1}} \, \tilde{\gamma}(z_t) \, dB_t \] (60)
where, for any \( z \),
\[ \tilde{d}_1(z) = d_1 \left( \chi^{-1}(z, 0) \right), \quad \tilde{d}_2(z) = d_2 \left( \chi^{-1}(z, 0) \right), \quad \tilde{\gamma}(z) = |\nabla \xi| \left( \chi^{-1}(z, 0) \right). \] (61)

5.2. Effective dynamics associated with (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 (23), where \( b_\varepsilon \) and \( \sigma_\varepsilon \) (that depend on \( \varepsilon \) since the Gibbs measure \( \mu_\varepsilon \) depends on \( \varepsilon \)) are defined by (21) and (22) and read as
\[ b_\varepsilon(\alpha) = E_{\mu_\varepsilon} \left[ (-\nabla V_0 \cdot \nabla \xi + \beta^{-1} \Delta \xi) (X) \mid \xi(X) = \alpha \right] = \tilde{d}_1(\alpha) - \frac{2}{\varepsilon} E_{\mu_\varepsilon} \left[ (q \, \nabla q \cdot \nabla \xi) (X) \mid \xi(X) = \alpha \right]. \] (62)
\begin{align*}
\sigma_\varepsilon^2(\alpha) &= E_{\mu_\varepsilon} \left[ |\nabla \xi|^2(X) \mid \xi(X) = \alpha \right], \\
\end{align*}
where \( \tilde{d}_1(\alpha) = E_{\mu_\varepsilon} \left[ (-\nabla V_0 \cdot \nabla \xi + \beta^{-1} \Delta \xi) (X) \mid \xi(X) = \alpha \right] \). It is easy to check that, for any \( \alpha \), we have
\[ \tilde{d}_1(\alpha) = \tilde{d}_1(\alpha) + O(\varepsilon) \quad \text{and} \quad \sigma_\varepsilon(\alpha) = \tilde{\gamma}(\alpha) + O(\varepsilon). \] (63)
where \( \tilde{d}_1 \) and \( \tilde{\gamma} \) are defined by (61).

5.3. Sufficient conditions for the pathwise convergence to the effective dynamics (23)

Let us establish sufficient conditions under which equation (60) is equivalent to the effective dynamics (23), in the limit \( \varepsilon \to 0 \). We hence request that, in the limit \( \varepsilon \to 0 \), dynamics (23) 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 \( \sigma_\varepsilon \) of (23) converges to \( \tilde{\gamma} \), the diffusion coefficient of (60).

We now turn to the drift term, which is \( \tilde{d}_1 + d_2 \) in the case of (60), and \( b_\varepsilon \) given by (62) for the effective dynamics (23). In view of (63), these drift terms are equal, in the limit \( \varepsilon \to 0 \), if and only if
\[ - \tilde{d}_2(\alpha) = \lim_{\varepsilon \to 0} \frac{2}{\varepsilon} E_{\mu_\varepsilon} \left[ (q \, \nabla q \cdot \nabla \xi) (X) \mid \xi(X) = \alpha \right]. \] (64)

In view of (58) and (61), we have
\[ \tilde{d}_2(\alpha) = d_2 \left( \chi^{-1}(\alpha, 0) \right) = d_{2a} \left( \chi^{-1}(\alpha, 0) \right) + \beta^{-1} d_{2b} \left( \chi^{-1}(\alpha, 0) \right), \] (65)
where \( d_{2a} \) and \( d_{2b} \) do not depend on \( \beta \):
\[ d_{2a} = u \frac{\nabla q \cdot \nabla V_0}{|\nabla q|^2}, \] (66)
\begin{align*}
d_{2b} &= u \frac{|\nabla q|}{|\nabla q|^2} - \kappa \frac{u}{|\nabla q|} + u \frac{\nabla q^T \nabla q}{|\nabla q|^4}. \\
\end{align*}
On the other hand, we compute, for any \( \alpha \),
\[
\mathbb{E}_{\mu_x} [(q \nabla q \cdot \nabla \xi) \mid \xi(X) = \alpha] = \int_{S_x} q \, d\mu_x|_{S_x}.
\]
A direct computation shows that
\[
\mathbb{E}_{\mu_x} [(q \nabla q \cdot \nabla \xi) \mid \xi(X) = \alpha] = \frac{\varepsilon}{2\beta} \mathcal{E}(\alpha) + O(\varepsilon^{3/2}),
\]
where \( \mathcal{E}(\alpha) \) does not depend on \( \beta \) and reads as
\[
\mathcal{E}(\alpha) = \frac{\partial \tilde{u}}{\partial q}(\alpha, 0) + \tilde{u}(\alpha, 0) \cdot \frac{\partial j}{\partial q}(\alpha, 0) + \tilde{u}(\alpha, 0) \frac{\partial \tilde{V}_0}{\partial q}(\alpha, 0),
\]
where \( \tilde{V}_0(\xi, q) = V_0(\chi^{-1}(\xi, q)) \),
\[
\tilde{u}(\xi, q) = u(\chi^{-1}(\xi, q)),
\]
and \( j = \text{det \text{jac} } \chi^{-1} \), where \( \text{jac} \chi^{-1} \) is the Jacobian matrix of \( \chi^{-1} \). Hence, (64) reads as
\[
- d_{2a}(\chi^{-1}(\alpha, 0)) - 1 \beta d_{2b}(\chi^{-1}(\alpha, 0)) = 1 \beta \mathcal{E}(\alpha).
\]
We want to enforce this relation for any \( \beta \). Since \( d_{2a}, d_{2b} \) and \( \mathcal{E} \) do not depend on \( \beta \), this yields
\[
d_{2a}(\chi^{-1}(\alpha, 0)) = 0 \quad \text{and} \quad - d_{2b}(\chi^{-1}(\alpha, 0)) = \mathcal{E}(\alpha).
\]
In view of (66), a sufficient condition for the first relation to hold is
\[
\forall \alpha \in \mathbb{R}, \quad u(\chi^{-1}(\alpha, 0)) = 0,
\]
where, we recall, \( u = \nabla \xi \cdot \nabla q \) and \( \chi \) is such that \( \chi(x) = (\xi(x), q(x)) \). In what follows, we now assume that \( \xi \) is such that (71) holds. The second relation of (70) now reads as
\[
\forall \alpha \in \mathbb{R}, \quad \frac{\nabla q \cdot \nabla u}{|\nabla q|^2}(\chi^{-1}(\alpha, 0)) = \frac{\partial \tilde{u}}{\partial q}(\alpha, 0).
\]
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 (23), obtained using conditional expectations, and pass to the limit \( \varepsilon \to 0 \) in the drift and diffusion coefficients.

Under conditions (71) and (72) (where \( u, \chi \) and \( \tilde{u} \) are defined by (59), (56) and (68), 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 \( \varepsilon \leq \varepsilon_0 \), we have
\[
\sup_{t \in [0:T]} \mathbb{E} |\xi(X_t^\varepsilon) - \xi(X_t)|^2 \leq C \varepsilon,
\]
where \( X_t^\varepsilon \) solves (51) and \( y_t^\varepsilon \) solves the effective dynamics (23).

**Proof.** We only have to prove (73). We infer from (55) and assumption [H11] that
\[
\sup_{t \in [0:T]} \mathbb{E} |\xi(X_t^\varepsilon) - \xi(X_t)|^2 \leq C \varepsilon.
\]
The drift and diffusion coefficients of the effective dynamics on \( y_t^\varepsilon \) are \( b_\varepsilon \) and \( \sigma_\varepsilon \). In view of (62), (63), (67), (69) and (65), the former satisfies
\[
\begin{align*}
b_\varepsilon(\alpha) &= \tilde{d}_1(\alpha) - \frac{2}{\varepsilon} \mathbb{E}_{\mu_x} [(q \nabla q \cdot \nabla \xi) \mid \xi(X) = \alpha] \\
&= \tilde{d}_1(\alpha) + O(\varepsilon) - \beta^{-1} \mathcal{E}(\alpha) + O(\sqrt{\varepsilon}) \\
&= \tilde{d}_1(\alpha) + \tilde{d}_2(\alpha) + O(\sqrt{\varepsilon}).
\end{align*}
\]
In view of (63), the latter satisfies \( \sigma_\varepsilon(\alpha) = \tilde{\gamma}(\alpha) + O(\varepsilon) \).
Hence, the difference between, on the one hand, the drift and diffusion coefficients of the effective dynamics \((b_\varepsilon, \sigma_\varepsilon)\) and, on the other hand, the drift and diffusion coefficients of equation (60) on \(z_t = \xi(\overline{X}_t)\) (namely \(\tilde{d}_1 + \tilde{d}_2\) and \(\tilde{\gamma}\)), is of order \(O(\sqrt{\varepsilon})\). We infer from this estimate that, on any bounded time interval,

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

Gathering that estimate with (74) yields (73).

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 (71) and (72) 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

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

where \(\Omega\) is bounded away from 0 and \(V_0\) does not depend on \(\varepsilon\), and the associated overdamped Langevin equation, which defines the process \(X_\varepsilon^t = (x_\varepsilon^t, y_\varepsilon^t)\). The limit dynamics on \(x_\varepsilon^t\) when \(\varepsilon \to 0\) is well known in that case (see for instance [34]): it reads as

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

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

We now wish to recover that result within our approach. The potential energy (75) is of the form (44), with \(q(x, y) = \Omega(x) y\). We wish to choose the reaction coordinate \(\xi(x, y) = x\). Observe then that \(u = \nabla \xi \cdot \nabla q = \Omega'(x) y / \Omega(x) \neq 0\). 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 (71) and (72) are satisfied. Hence, according to proposition 5.1, in the limit \(\varepsilon \to 0\), the effective dynamics (23) is accurate in the sense of pathwise convergence.

For \(\varepsilon > 0\), the effective dynamics reads as

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

with \(b_\varepsilon(\alpha) = -E_{\mu_\varepsilon(\alpha, \cdot)} \left( \partial_x V_0(\alpha, y) + \frac{2 \Omega'(\alpha) \Omega(\alpha) y^2}{\beta \Omega(\alpha)} \right)\). A straightforward computation shows that \(\lim_{\varepsilon \to 0} b_\varepsilon(\alpha) = -\partial_x V_0(\alpha, 0) - (\Omega'(\alpha) / \beta \Omega(\alpha))\). Inserting this relation in (77), we recover (76).

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 (47)

In the numerical case considered in section 4.1, 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 \(\xi_2\), in terms of pathwise convergence. We
have checked this as follows. First, we have simulated a solution of (51) (with the potential $V_\varepsilon$ defined by (47)), for a given realization of the two-dimensional noise, with $\varepsilon = 0.01$ and $\beta = 3$. From this trajectory $X_t$ (we omit here for clarity the dependence with respect to $\varepsilon$), we obtain the time evolution $\xi_2(X_t)$, and we can also construct the one-dimensional noise (13). This noise is next used in the effective dynamics (23). We compare both trajectories in figure 5: we observe an excellent agreement over $10^6$ time steps (the trajectories plotted in figure 5 have been computed with a time step $\Delta t = 10^{-4}$, hence $T = 100 = 10^6 \Delta t$).

In section 4.1, 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, $\chi^{-1}(\xi, q) = (\xi, q + 1 - \xi^2)$, hence $u_1(\chi^{-1}(\xi, 0)) = 2\xi$, so condition (71) is not satisfied. We have numerically performed the same comparison with $\xi_1$ as the one reported above for $\xi_2$. Results are shown in figure 5: we observe that the complete dynamics (projected on the reaction coordinate) and the effective dynamics differ, as expected.

5.6. Numerical results on a three atom molecule

We consider again the three atom molecule of section 4.2, with potential (49). We now choose $W_3(\theta) = \frac{1}{2} k_0 (\theta - \theta_0)^2$.

We work with the parameters $\varepsilon = 10^{-3}$, $k_0 = 208$, $\ell_{\text{eq}} = 1$, $\theta_0 = 1.187$, $\beta = 1$ and $\Delta t = 10^{-3}$. As pointed out in section 4.2, the reaction coordinate $\xi(x) = \theta_{\text{ABC}}(x)$ is orthogonal to the stiff terms of the potential energy, and is hence expected to be a good reaction coordinate.

Numerical experiments confirm this belief: choosing this reaction coordinate, we considered the effective dynamics (23), and compared its solution with the time evolution $\theta_{\text{ABC}}(X_t)$, where $X_t$ solves (3). Results are shown in figure 6: again, we see a good agreement between both trajectories.

6. Conclusion

Following a natural strategy, we have introduced the effective dynamics (23), using conditional expectations to close the dynamics of the reaction coordinate. In general, the diffusion
coefficient $\sigma(z)$ defined by (22) is not equal to 1, and dynamics (23) is hence different from dynamics (5), driven by the free energy $A$ associated with the reaction coordinate. We have also shown (see (25)) that the drift and diffusion coefficients of (23), namely $b$ and $\sigma$, defined by (21) and (22) respectively, satisfy

$$b = -\sigma^2 A' + \beta^{-1} \partial_z (\sigma^2).$$

Hence, the computation of both $A'$ and $\sigma^2$ (and its derivative) are needed to simulate dynamics (23). This is reminiscent of a well-known fact in molecular dynamics, namely that it is important to take into account a so-called metric tensor, in addition to the free energy, when describing dynamical effects [29].

The numerical examples of section 4 clearly demonstrate that residence times in metastable basins can be accurately recovered using (23), and that (5) may lead to inaccurate results.

In addition, we have also shown that, when it is possible, the reaction coordinate should be chosen such that it is orthogonal to the stiff terms of the potential energy (see criterion (46)). As a consequence, in the example of the three atom molecule described in section 4.2, such a criterion helps discriminate between two reaction coordinates that are, a priori, equivalent.

Acknowledgments

This work is supported in part by the INRIA, under the grant ’Action de Recherche Collaborative’ HYBRID, and by the Agence Nationale de la Recherche, under grant ANR-09-BLAN-0216-01 (MEGAS). 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. The authors also wish to acknowledge very enlightening discussions with Felix Otto and Georg Menz.

References

[1] And C, Blachère S, Chafaï D, Fougeres P, Gentil I, Malrieu F, Roberto C and Scheffer G 2000 Sur les Inégalités de Sobolev Logarithmiques (Paris: Société Mathématique de France)
[2] Arnold A, Markowich P, Toscani G and Unterreiter A 2001 On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations Commun. Partial Diff. Eqs 26 43–100
[3] Bichteler K 2002 Stochastic Integration with Jumps (Cambridge: Cambridge University Press)
2162 F Legoll and T Lelièvre

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

[5] Cancès E, Legoll F and Stoltz G 2007 Theoretical and numerical comparison of some sampling methods for molecular dynamics Math. Modelling Numer. Anal. (M2AN) 41 351–89

[6] Chipot C and Pohorille A (ed) 2007 Free Energy Calculations (Springer Series in Chemical Physics vol 86) (Berlin: Springer)

[7] Chorin A J, Hald O H and Kupferman R 2000 Optimal prediction and the Mori–Zwanzig representation of irreversible processes Proc. Natl Acad. Sci. 97 2968–73

[8] Chorin A J, Hald O H and Kupferman R 2002 Optimal prediction with memory Physica D 166 230–57

[9] Ciccotti G, Lelièvre T and Vanden-Eijnden E 2008 Projection of diffusions on submanifolds: application to mean force computation Commun. Pure Appl. Math. 61 371–408

[10] Darve E, Solomon J and Kia A 2009 Computing generalized Langevin equations and generalized Fokker–Planck equations Proc. Natl Acad. Sci. 106 10884–9

[11] Dellago C, Bolhuis P G and Chandler D 1999 On the calculation of reaction rate constants in the transition path ensemble J. Chem. Phys. 110 6617–25

[12] Dudzár D 2007 Towards an optimal rate of convergence in the hydrodynamic limit for Kawasaki dynamics PhD Thesis Bonn University

[13] Evans L C and Gariepy R F 1992 Measure Theory and Fine Properties of Functions (Boca Raton, FL: CRC Press)

[14] Evans L C and Gariepy R F 1992 Measure Theory and Fine Properties of Functions (Boca Raton, FL: CRC Press)

[15] Friesecke G, Junge O and Koltai P 2009 Mean field approximation in conformation dynamics SIAM Multiscale Modelling Simul. 8 254–88

[16] Givon D, Kupferman R and Stuart A M 2004 Extracting macroscopic dynamics: model problems and algorithms Nonlinearity 17 55–127

[17] Grunewald N, Otto F, Villani C and Westdickenberg M G 2009 A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit Ann. Inst. Henri Poincaré Probab. Statist. 45 302–51

[18] Gyöngy I 1986 Mimicking the one-dimensional marginal distributions of processes having an Itô differential Probab. Theory Relat. Fields 71 501–16

[19] Hanggi P, Talkner P and Borkovec M 1990 Reaction-rate theory: fifty years after Kramers Rev. Mod. Phys. 62 251–342

[20] Hartmann C and Schütte C 2008 Balancing of partially-observed stochastic differential equations 47th IEEE Conf. on Decision and Control pp 4867–72

[21] Hartmann C, Vulcanov V-M and Schütte C 2007 Model reduction in classical molecular dynamics PhD Thesis Freie Universität Berlin http://www.diss.fu-berlin.de/2007/458

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

[23] Has’minskii R Z 1980 Stochastic Stability of Differential Equations (Alphen aan den Rijn: Sijthoff and Noordhoff)

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

[25] Iaguirre J A and Sweet C R 2009 Adaptive dimensionality reduction of stochastic differential equations for protein dynamics Proc. 2nd Int. Workshop on Model Reduction in Reacting Flows (Notre Dame, IN)

[26] Kliemann W 1987 Recurrence and invariant measures for degenerate diffusions Ann. Probab. 15 690–707

[27] Kramers H A 1940 Brownian motion in a field of force and the diffusion model of chemical reactions Physica 7 284–304

[28] Lelièvre T 2009 A general two-scale criteria for logarithmic Sobolev inequalities J. Funct. Anal. 256 2211–21

[29] Maragliano L, Fischer A, Vanden-Eijnden E and Ciccotti G 2006 String method in collective variables: minimum free energy paths and isocommittor surfaces J. Chem. Phys. 125 024106

[30] Otto F and Reznikoff M G 2007 A new criterion for the logarithmic Sobolev inequality and two applications J. Funct. Anal. 243 121–57

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

[32] Pavliotis G A and Stuart A M 2007 Multiscale Methods: Averaging and Homogenization (Berlin: Springer)

[33] Pekary Y, Stuart A M and Wiberg P 2009 Parameter estimation for partially observed hypoelliptic diffusions J. R. Statistical Soc. Ser. B 71 49–73

[34] Reich S 2000 Smoothed Langevin dynamics of highly oscillatory systems Physica D 138 210–24
[35] Ryckaert J P and Bellemans A 1978 Molecular dynamics of liquid alkanes Faraday Discuss. 66 95–106
[36] Schütte C, Fischer A, Huisinga W and Dede P 1999 A direct approach to conformational dynamics based on Hybrid Monte-Carlo J. Comput. Phys. 151 146–68
[37] Schütte C and Huisinga W 2003 Biomolecular conformational states can be identified as metastable sets of molecular dynamics Handbook of Numerical Analysis (Special volume on computational chemistry) vol X, ed P G Ciarlet and C Le Bris (Amsterdam: Elsevier) pp 699–744
[38] Straub J E, Borkovec M and Berne B J 1988 Molecular-dynamics study of an isomerizing diatomic in a Lennard-Jones fluid J. Chem. Phys. 89 4833–47
[39] Sweet C R, Petrone P, Pande V S and Izaguirre J A 2008 Normal mode partitioning of Langevin dynamics for biomolecules J. Chem. Phys. 128 145101
[40] Villani C 2003 Topics in Optimal Transportation (Graduate Studies in Mathematics vol 58) (Providence, RI: American Mathematical Society)
[41] Yang S, Onuchic J N and Levine H 2006 Effective stochastic dynamics on a protein folding energy landscape J. Chem. Phys. 125 054910