Acta Numerica (2016), pp. 681–880
doi:10.1017/S0962492916000039

Partial differential equations and stochastic methods in molecular dynamics*

Tony Lelièvre and Gabriel Stoltz
Université Paris-Est, CERMICS (ENPC), INRIA,
F-77455 Marne-la-Vallée, France
E-mail: tony.lelievre@enpc.fr, gabriel.stoltz@enpc.fr

The objective of molecular dynamics computations is to infer macroscopic properties of matter from atomistic models via averages with respect to probability measures dictated by the principles of statistical physics. Obtaining accurate results requires efficient sampling of atomistic configurations, which are typically generated using very long trajectories of stochastic differential equations in high dimensions, such as Langevin dynamics and its overdamped limit. Depending on the quantities of interest at the macroscopic level, one may also be interested in dynamical properties computed from averages over paths of these dynamics.

This review describes how techniques from the analysis of partial differential equations can be used to devise good algorithms and to quantify their efficiency and accuracy. In particular, a crucial role is played by the study of the long-time behaviour of the solution to the Fokker–Planck equation associated with the stochastic dynamics.

CONTENTS

* Colour online for monochrome figures available at journals.cambridge.org/anu.

1. Introduction

It is now increasingly common to use computers to perform simulations to check various claims or to predict phenomena of interest. In some sense, numerical simulation has become the 'third way' of science, alongside theory and experiment: more and more scientists are checking the correctness, relevance or validity of their claims via numerical simulations (complementing or replacing experiments); in turn, many theories owe their origin to intriguing results obtained via numerical simulation. The fact that numerical simulations have essentially become numerical experiments is made clear in expressions such as '*in silico* experiments' or the 'computational microscope' (Lee *et al.* 2009).

Numerical experiments are particularly relevant to atomistic systems, which may be very difficult to understand with standard, macroscopic experimental devices in usual conditions of temperature, pressure, *etc.* For such atomistic systems, molecular simulation is now a recognized approach. For example, two Nobel Prizes in Chemistry have been awarded to pioneers of these methods: in 1998 the prize was awarded to Walter Kohn 'for his development of the density-functional theory' and John A. Pople 'for his development of computational methods in quantum chemistry', and in 2013 the prize was shared between Martin Karplus, Michael Levitt and Arieh Warshel 'for the development of multiscale models for complex chemical systems'.

In this introductory section, we first give a general presentation of molecular dynamics, which aims at simulating atomistic systems (see Section 1.1). In particular, we give a brief account of the fundamental concepts of statistical physics that are required in order to understand the models at hand. We next present in Section 1.2 the numerical challenges raised by molecular dynamics, with an emphasis on problems involving partial differential equations (PDEs). It may in fact be unclear to the reader why PDEs should come into the picture. We devote Section 1.3 to explaining why and where tools from the analysis of PDEs can be used to improve the numerical methods used for molecular simulation. The general outline of this article is given at the end of Section 1.3.

1.1. Molecular dynamics simulations

In statistical physics, matter is described at the atomic level in a classical or a quantum setting. There is a wide variety of models with various resolutions, ranging from quantum approaches to computing electronic structures, with methods such as density functional theory and its various implementations (Hohenberg and Kohn 1964, Kohn and Sham 1965, Lieb 1983, Cancès *et al.* 2003), to more coarse-grained descriptions, such as those considered in kinetic Monte Carlo (see for instance the review by Voter 2007) or

dissipative particle dynamics (Hoogerbrugge and Koelman 1992, Español and Warren 1995). Among these models, molecular dynamics based on empirical interaction potentials offers an interesting compromise between modelling at the atomistic level and computational cost. On the one hand, quantum approaches are indeed plagued by their high computational cost, although some promising improvements have been made towards the computation of interaction forces scaling linearly with the system size (mostly for insulators, the efficient simulation of metals remaining a challenge). On the other hand, more mesoscopic approaches, such as kinetic Monte Carlo or dissipative particle dynamics, are more difficult to parametrize, and typically provide only qualitative results in some specific physical regimes.

Within molecular dynamics, many computational models also co-exist. Some researchers prefer to use purely deterministic dynamics such as Hamiltonian dynamics or Nosé–Hoover-like dynamics (Nosé 1984*b*, Nosé 1984*a*, Hoover 1985, Bond and Leimkuhler 2007), while others resort to Monte Carlo approaches such as the Metropolis algorithm and its variants (Metropolis *et al.* 1953, Hastings 1970). In this review we focus on yet another category of models, which are quite popular: continuous stochastic dynamics, the most famous examples of which are Langevin dynamics and its overdamped limit. The distinction between stochastic and deterministic dynamics is not so clear-cut in many cases. Indeed, deterministic dynamics are sometimes weakly perturbed by a stochastic term, as in the Langevin–Nosé–Hoover method (Leimkuhler, Noorizadeh and Theil 2009). Moreover, there are stochastic dynamics such as Metropolis-like algorithms which are based on deterministic proposals, as in hybrid Monte Carlo approaches (Duane, Kennedy, Pendleton and Roweth 1987).

1.1.1. Microscopic description of physical systems

To present the dynamics we consider in this review more precisely, we first give a microscopic description of physical systems. The state of a system composed of N particles (typically atoms) is completely determined by the positions $q = (q_1, \dots, q_N)$ and the momenta $p = (p_1, \dots, p_N)$ of all the particles. Recall that, in Cartesian coordinates, momenta are velocities multiplied by the masses of the particles under consideration.

One of the fundamental issues in statistical physics is that N should *a priori* be of the order of the Avogadro number to determine the macroscopic behaviour of matter, while, in practice, current simulations are mostly performed for systems of $N = 10^6 - 10^9$ particles, with heroic computations involving record numbers of particles of order $N = 10^{12}$. The systems computed are therefore much smaller, by several orders of magnitude, than the macroscopic systems they represent. Nonetheless, it turns out that molecular simulation is often able to give relevant predictions for the simulated

systems. For the simplest systems, such as noble gases, bulk properties can in fact be estimated with only a few thousand particles.

Momenta take values in $\mathbb{R}^{3N}$ for three-dimensional systems, whereas positions take values in a position space $\mathcal{D}$, which can be either unbounded (the typical case being $\mathcal{D} = \mathbb{R}^{3N}$ for isolated systems in three dimensions) or bounded (the typical case being simulations performed in a periodic cell $\mathcal{D} = (L\mathbb{T})^{3N}$, with $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denoting the one-dimensional torus). The latter situation is encountered, for example, in the simulation of condensed matter systems, for which periodic boundary conditions are used to mimic bulk conditions. In this paper we denote the dimension of the system by d ($d = 3N$ for particles in three-dimensional space), and the phase space by $\mathcal{E} = \mathcal{D} \times \mathbb{R}^d$. Since some dynamics only involve the positions while other dynamics involve both positions and momenta, we will often use the abstract notation $\mathcal{X}$ to denote a state space which can be either $\mathcal{D}$ or $\mathcal{E}$, depending on context.

The physical content of the model is essentially encoded in the potential energy function $V(q)$ that gives the interaction energy of the N particles with positions q. Interaction potentials should ideally be obtained from quantum mechanical computations based on an approximate numerical solution of the Schrödinger equation, for fixed nuclei positions (Cancès *et al.* 2003, Le Bris 2005). In practice, as already hinted at above, this is very expensive, so empirical potentials are often used for large-scale simulations. The construction and validation of appropriate empirical energy functions is a research domain in its own right. For some simple systems, satisfactory empirical potentials are available. This is the case for noble gases such as argon, for which the potential energy can be written as the sum of pairwise interactions of Lennard-Jones type. For molecules, two types of interaction should be distinguished: (i) bonded interactions to model covalent bonds, and which involve terms depending on bond lengths between two atoms, bending angles defined using three atoms, and dihedral angles computed using four atoms; (ii) non-bonded interactions, involving for instance Coulomb interactions between charged atoms, and Lennard-Jones interactions describing dipole–dipole interactions (see *e.g.* Schlick 2002). For solids, the current trend is towards variations and extensions of the embedded-atom model proposed in Baskes (1992), or bond-order potentials such as REBO (Tersoff 1989) or ReaxFF (van Duin *et al.* 2001). The latter potentials involve terms depending on the local environment of the atoms, which allows them to account for chemical reactions (*i.e.* bond breakings and bond formations). For biological applications, a lot of effort is devoted to obtaining good water models, because of the importance of water as a solvent. Potential energy functions are implemented in popular molecular dynamics packages such as AMBER, CHARMM, GROMACS, LAMMPS and NAMD.

The total energy of microscopic systems is given by the Hamiltonian $H(q,p)$, which is typically the sum of the kinetic energy and the potential energy,

$$H(q,p) = \frac{1}{2}p^T M^{-1} p + V(q), \tag{1.1}$$

where we introduce the mass matrix $M = \mathrm{diag}(m_1 \mathrm{Id}_3, \ldots, m_N \mathrm{Id}_3)$, assuming that the system is three-dimensional, and m_i denotes the mass of the ith particle. Let us also emphasize here that for simplicity we always consider the potential V to be C^∞ in this review.

Using the potential V, two physical models of the system can be obtained: a model of the system at equilibrium via thermodynamic ensembles, and a model for the evolution of the system using appropriate dynamics.

1.1.2. Macroscopic description via thermodynamic ensembles

The macroscopic state of a system at equilibrium is described, within the framework of statistical physics, by a probability measure on the configuration space, called the thermodynamic ensemble. The macroscopic features of the system are then computed as averages of an observable with respect to this measure.

We focus here on the canonical ensemble, in which the number of particles, the domain and the temperature are fixed. Mathematically, it is defined as the minimizer of Boltzmann's entropy function under the constraint that the average energy is fixed. The Lagrange multiplier associated with this constraint is the temperature. See Balian (2007) for more physical background, as well as Lelièvre, Rousset and Stoltz (2010, Section 1.2.3) for a short, self-contained presentation.

For systems described by their positions and momenta, the canonical measure reads

$$\mu(\mathrm{d}q\,\mathrm{d}p) = Z_\mu^{-1}\,\mathrm{e}^{-\beta H(q,p)}\,\mathrm{d}q\,\mathrm{d}p, \quad Z_\mu = \int_{\mathcal{E}} \mathrm{e}^{-\beta H}, \tag{1.2}$$

where H is defined by (1.1), $\beta = (k_B T)^{-1} > 0$ is the inverse temperature, and the normalization constant Z_μ is assumed to be finite. The probability measure μ is also known as the Boltzmann–Gibbs distribution associated with the function H. Note that the components of the momenta are independent under the canonical measure, and follow a Gaussian distribution.

The macroscopic properties are obtained as averages of functions (called observables) of the microstate (q,p) of the system with respect to the canonical measure. For a given observable $\varphi(q,p)$, macroscopic properties are therefore obtained by computing

$$\mathbb{E}_\mu(\varphi) = \int_{\mathcal{E}} \varphi(q,p)\,\mu(\mathrm{d}q\,\mathrm{d}p). \tag{1.3}$$

In many cases of interest, the integration with respect to the momenta can be performed analytically (see for instance (1.6) below), or the observable of interest only depends on the positions q (as is the case for the potential energy, for instance). The integral (1.3) then simplifies to

$$\mathbb{E}_\nu(\varphi) = \int_{\mathcal{D}} \varphi(q)\, \nu(\mathrm{d}q), \tag{1.4}$$

where ν is the marginal law of μ in q:

$$\nu(\mathrm{d}q) = Z_\nu^{-1}\, \mathrm{e}^{-\beta V(q)}\, \mathrm{d}q, \quad Z_\nu = \int_{\mathcal{D}} \mathrm{e}^{-\beta V}. \tag{1.5}$$

The probability measure ν is the Boltzmann–Gibbs distribution associated with the function V. Averages such as (1.3) and (1.4) are called *thermodynamic quantities*. As highlighted in the next example and further discussed in Section 1.2, the actual challenge to the numerical evaluation of macroscopic properties is generating positions distributed according to ν using an appropriate numerical method (see Sections 3 and 4).

For concreteness, let us give one example: in three-dimensional condensed matter systems equipped with periodic boundary conditions, the pressure is obtained by computing (1.3) with

$$\varphi(q,p) = \frac{1}{3|\mathcal{D}|} \sum_{i=1}^{N} \left(\frac{|p_i|^2}{m_i} - q_i \cdot \frac{\partial V}{\partial q_i}(q) \right).$$

The contribution of the momenta can be integrated out analytically (corresponding to the ideal gas contribution to the pressure), so that $\mathbb{E}_\mu(\varphi) = \mathbb{E}_\nu(\overline{\varphi})$ where

$$\begin{aligned} \overline{\varphi}(q) &= \frac{\int_{\mathbb{R}^d} \varphi(q,p)\, \mathrm{e}^{-\beta H(q,p)}\, \mathrm{d}p}{\int_{\mathbb{R}^d} \mathrm{e}^{-\beta H(q,p)}\, \mathrm{d}p} = \int_{\mathbb{R}^d} \varphi(q,p)\, \frac{\mathrm{e}^{-\beta p^T M^{-1} p/2}}{(2\pi\beta^{-1})^{d/2}\sqrt{\det(M)}}\, \mathrm{d}p \\ &= \frac{1}{3|\mathcal{D}|} \left(\frac{d}{\beta} - \sum_{i=1}^{N} q_i \cdot \frac{\partial V}{\partial q_i}(q) \right). \end{aligned} \tag{1.6}$$

1.1.3. Microscopic dynamics

Once a description of the system is provided in terms of degrees of freedom, boundary conditions and potential energy function, we can also discuss its time evolution. For isolated systems, namely in situations where it is possible to simulate the system and its entire environment, systems should evolve according to Hamiltonian dynamics. However, for systems evolving in an environment so large that it fixes the temperature, for example, it is not generally possible to simulate both the system of interest and the ambient environment, which acts as a thermal bath. In this case, a pragmatic approach is to replace the average effect of the environment by some

fluctuation/dissipation process which indeed makes the dynamics ergodic with respect to the canonical measure (ergodicity will be discussed in detail below). One theoretical underpinning for this approach is provided by the Mori–Zwanzig framework (Mori 1965, Zwanzig 1973, Givon, Kupferman and Stuart 2004). In such cases, the physical relevance of the dynamics considered in molecular dynamics is debatable if one is precisely interested in dynamical properties. We will not discuss this modelling problem further, and focus on two dynamics which are widely used in the community: the Langevin dynamics and its overdamped limit.

Two situations should be distinguished. In some cases, dynamics are used as *sampling devices*. The focus is then on invariant measures or stationary properties rather than on dynamical information. The aim is to construct dynamics which are ergodic, in the sense that average properties computed along a typical realization converge in the long-time limit to averages with respect to some given measure of interest such as the canonical measure (1.2) (see equation (1.10) below). However, it is also often the case that *dynamical properties* (which really depend on the dynamics at hand and not only on the measure for which the dynamics is ergodic) are computed. These properties take the general form

$$\mathbb{E}[\varphi((x_t)_{t\in I})], \tag{1.7}$$

where $(x_t)_{t\in I}$ is a random trajectory over a time interval I (which can be finite or infinite), and φ is a path functional. In fact, many current numerical challenges concern the computation of such dynamical properties.

The two paradigmatic dynamics we consider in this work are the following stochastic differential equations (SDEs).

(1) *Overdamped Langevin dynamics*:

$$\mathrm{d}q_t = -\nabla V(q_t)\,\mathrm{d}t + \sqrt{\frac{2}{\beta}}\,\mathrm{d}W_t, \tag{1.8}$$

where W_t is a standard d-dimensional Brownian motion.

(2) *Langevin dynamics*:

$$\begin{cases} \mathrm{d}q_t = M^{-1}p_t\,\mathrm{d}t, \\ \mathrm{d}p_t = -\nabla V(q_t)\,\mathrm{d}t - \gamma\, M^{-1}p_t\,\mathrm{d}t + \sqrt{\dfrac{2\gamma}{\beta}}\,\mathrm{d}W_t, \end{cases} \tag{1.9}$$

where $\gamma > 0$ is a positive parameter.

As discussed in Lelièvre *et al.* (2010, Section 2.2.4), overdamped Langevin dynamics can be obtained from Langevin dynamics by two limiting processes when the mass matrix is proportional to the identity matrix ($M = m\,\mathrm{Id}$): (i) a large friction limit $\gamma \to +\infty$ with time rescaled as γt; (ii) a small mass limit $m \to 0$.

There exist many variations and extensions to the Langevin dynamics, such as Langevin dynamics with position-dependent friction $\gamma(q_t)$ in (1.9), dissipative particle dynamics, where the friction depends on the relative velocities between the particles (Hoogerbrugge and Koelman 1992, Español and Warren 1995), generalized Langevin dynamics, with an extra variable which can be seen as an integral of the fluctuation/dissipation (Zwanzig 1973), and Langevin–Nosé–Hoover (Leimkuhler *et al.* 2009). We restrict ourselves to the paradigms (1.8) and (1.9) for simplicity. Many numerical techniques can be extended, at least formally, to other dynamics, although mathematical proofs of convergence or consistency may be much more complicated, since the basic ergodic properties of these dynamics may not be known. This is the case, for example, for the dissipative particle dynamics, for which ergodicity is proved only in one dimension (Shardlow and Yan 2006).

1.2. Numerical challenges in molecular dynamics

Molecular dynamics raises a number of computational challenges. In this article we mainly focus on *sampling problems*. Sampling consists in approximating an average by using empirical averages over a collection of samples, namely realizations of the underlying random variable distributed according to a target probability measure. We consider here either thermodynamic quantities, such as (1.3)–(1.4) (in which case the target probability is μ or ν), or dynamical ones, such as (1.7) (in which case the target probability is the law of the process $(x_t)_{t\in I}$ over the path space). For completeness, we also mention other challenges in molecular dynamics at the end of this section.

1.2.1. Static sampling problems

Thermodynamic averages such as (1.3) and (1.4) are in general very difficult to approximate by direct numerical integration since the number of particles is very large (and thus the dimension of the space over which the integral is posed is large). This is why stochastic approaches are typically used. The fundamental idea of these methods is to replace ensemble averages with time averages along trajectories of well-chosen dynamics, which are ergodic with respect to the target probability measure. More precisely, when the target measure is the canonical measure (1.2), these dynamics should be such that

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \varphi(q_s, p_s)\, \mathrm{d}s = \int_{\mathcal{E}} \varphi(q,p)\mu(\mathrm{d}q\, \mathrm{d}p). \tag{1.10}$$

Ergodicity means that time averages over trajectories almost surely converge to a well-defined limit (for μ-almost all initial conditions, for almost every realization of the noise used in the dynamics), this limit being an

average over the phase space with respect to a probability measure (the canonical measure in our context). Examples of such ergodic dynamics include overdamped Langevin dynamics (1.8) (which is ergodic with respect to ν) and Langevin dynamics (1.9) (which is ergodic with respect to μ). We will present examples of efficient techniques to compute thermodynamic quantities in Section 3, the particular case of the computation of free energy differences being covered in Section 4.

1.2.2. Sampling dynamical properties

Other quantities of interest actually depend on the dynamics. This includes, for example, time correlation functions (see Section 5) or transition times and paths between regions of the configuration space (see Section 6). As an example, let us consider an exit problem, which consists in sampling the random variables $(\tau, (q_\tau, p_\tau))$ where

$$\tau = \inf\{t > 0,\, q_t \notin A\}.$$

Here A is a given region of the configuration space (an open bounded simply connected subset of $\mathbb{R}^d$), and it is implicitly assumed that $q_0 \in A$. The random variables $(\tau, (q_\tau, p_\tau))$ fully characterize the exit event from A: they give the exit times and exit points from A. Sampling the random variables $(\tau, (q_\tau, p_\tau))$ requires us to sample the paths $(q_t, p_t)_{0 \leqslant t \leqslant \tau}$: it is therefore again a sampling problem, but over an infinite-dimensional space, namely the space of trajectories. Averages over the path space are called *dynamical quantities*. We present efficient techniques to sample exit events in Sections 6.2 and 6.4. We will also discuss numerical methods to approximate time correlation functions to compute transport coefficients in Section 5.3.

1.2.3. Challenges in sampling problems

The main practical difficulty when approximating thermodynamic and dynamical quantities in molecular dynamics arises from the variety of space and time scales involved in these computations. To obtain relevant results at the macroscopic level, one needs to integrate over a *very long time* the dynamics for a *very large number* of interacting particles. The typical time step used to discretize the equations of motion at the atomistic level is of the order of one femtosecond (10^{-15} s), in order to resolve the highest-frequency modes arising from bond vibrations. On the other hand, the macroscopic phenomena of interest typically occur on time scales of the order of microseconds (10^{-6} s) to hours (10^{3} s). For the former, we mention the diffusion of adatoms on a substrate (see the review article by Voter, Montalenti and Germann 2002 and references therein), and for the latter, the unbinding of some ligands from proteins (see Schlick 2002, Freddolino, Harrison, Liu and Schulten 2010). In addition, the number of particles one

has to consider should be sufficiently large that the macroscopic phenomena of interest emerge from the collective behaviour of the many particles at the microscopic level. The time horizon and the number of particles to consider depend on the system and the phenomenon of interest, but it should be very large in any case – typically larger than in other fields where sampling of probability measures is needed, such as Bayesian inference in computational statistics.

In terms of time scales, this discrepancy between the microscopic and the macroscopic scales is related to the fact that the stochastic processes that govern the evolutions of the positions of the particles are *metastable*: the system remains trapped for a long time in some region of the configurational space (called a metastable region) before hopping to another metastable region. Metastability implies that the convergence of averages over trajectories is slow, and that transitions between metastable regions are rare. As a simple illustration of metastability, let us present a two-dimensional example using overdamped Langevin dynamics (1.8), which we recall for convenience:

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

In Figure 1.1 the two origins of metastability are illustrated. First, as shown in Figure 1.1(a, b), the stochastic process $(q_t)_{t\geqslant 0}$ can be metastable because of energetic barriers: to migrate from one metastable region to another, the process has to cross level sets of the energy function V and to overcome a first-order saddle point. Second, metastability can be observed because of some steric constraints in the system. In the example in Figure 1.1(c, d), the potential function is zero inside the hourglass-shaped region and strongly repulsive outside this region. The stochastic process $(q_t)_{t\geqslant 0}$ is then a simple random walk in the region of zero potential. In this case, metastability is observed because it takes a long time to find the narrow corridor between the two metastable regions. Such potentials exhibit so-called entropic barriers, since the accessible space for the system to leave a metastable state is very small. These two types of metastable behaviour are different in nature, and as will be explained below, this distinction is important when discussing sampling techniques, since some of them are limited to energetic barriers (see in particular Section 6). The difference between these two types of barrier is best seen by modifying the temperature. When the temperature decreases, the typical time scale to leave a metastable region grows exponentially fast with respect to β for energetic barriers, in accordance with results from large deviation theory (Freidlin and Wentzell 1984, Dembo and Zeitouni 1998). For an entropic barrier, such as that of Figure 1.1(c, d), a change in temperature is (asymptotically) equivalent to a linear rescaling in time. These different behaviours are illustrated in Figure 1.2. In terms of numerical approximations, metastability implies that trajectory averages

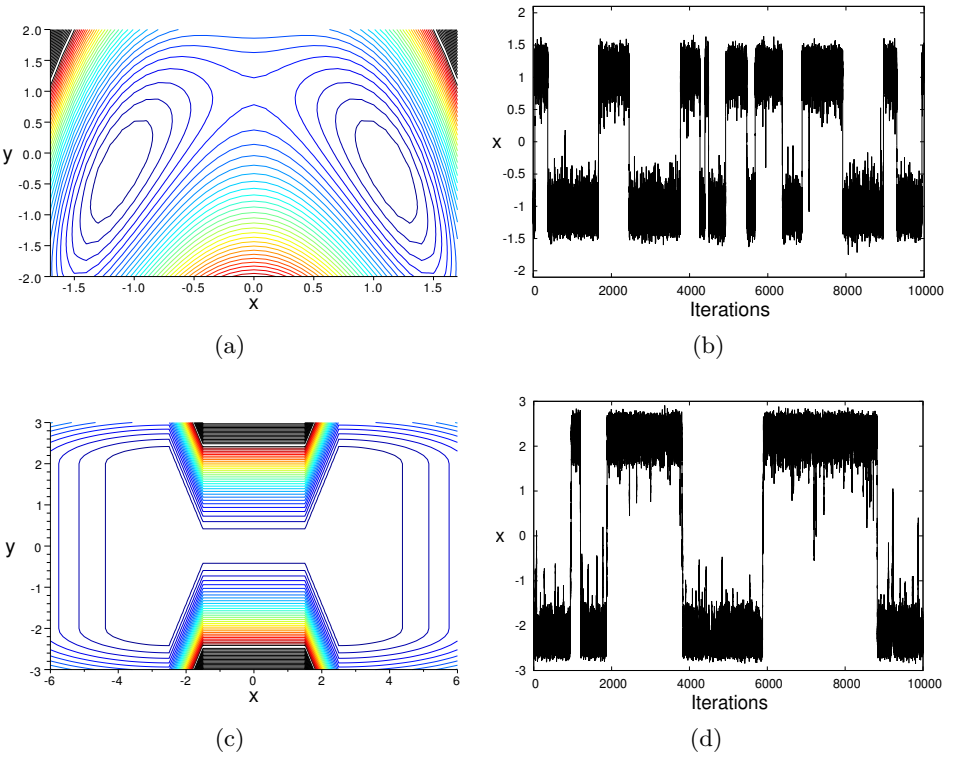
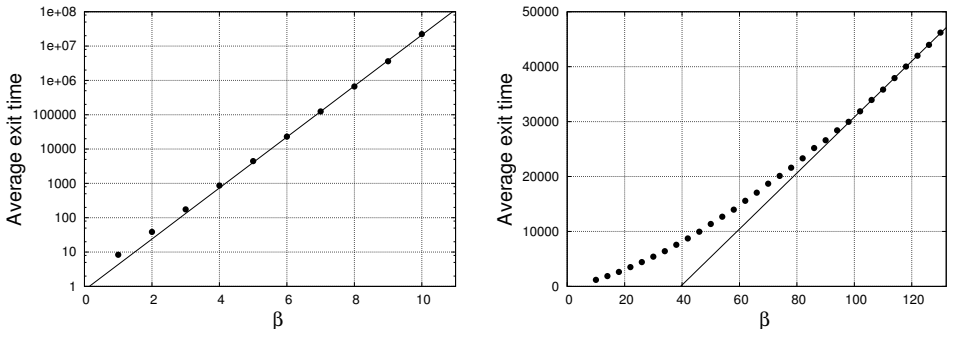


Figure 1.1. (a, c) Level sets of the two-dimensional potentials. (b, d) Time evolution of the x-coordinate of the stochastic process solution to overdamped Langevin dynamics (1.8) in these potentials. (a, b) Energetic barrier ($\beta = 4$), (c, d) entropic barrier ($\beta = 10$).

converge very slowly to their ergodic limits (the convergence $t \to \infty$ in (1.10) is difficult to reach numerically), because one has to wait for the process to visit all the metastable states sufficiently often. This explains the difficulties encountered when evaluating thermodynamic quantities accurately. In addition, metastability also implies that transitions between metastable states are rare events, but these events are typically the phenomena of interest when considering dynamical quantities. Metastability is thus at the heart of the numerical challenges observed in molecular dynamics. Therefore, the mathematical analysis of the efficiency of sampling algorithms requires us to quantify the metastability of the sampling dynamics (Lelièvre 2013).

As explained above, the metastability of the dynamics has some physical origin (discrepancy of time scales between the microscopic and the macroscopic worlds). From a purely numerical viewpoint, it is related to the multimodality of the probability measure sampled by these dynamics. A probability measure is multimodal if likely regions (modes) are separated



(a) (b)

Figure 1.2. Average exit times $\mathbb{E}(\tau)$ as a function of β, where $\tau = \inf\{t \geqslant 0 \mid x_t \geqslant x_*\}$. (a) Energetic barrier (potential in Figure 1.1(a)) with $x_* = 1.3$ and initial conditions $x_0 = (-1, 0)$. The average exit times increase exponentially with β (note the logarithmic scale on the ordinate axis). (b) Entropic barrier (potential in Figure 1.1(c)) with $x_* = 1.5$ and initial conditions $x_0 = (-3, 0)$. The average exit times increase linearly with β for sufficiently large values of β.

by areas of small probabilities. Unless deep knowledge of the system under study is available, two successive points sampled by ergodic dynamics are very close to each other, because large uninformed moves are very likely to fall in regions of small probability. This is why ergodic dynamics with respect to a multimodal measure are in general metastable: it is difficult to leave a mode of the target probability measure. This is closely related to the difficulties associated with the global minimization of a function in $\mathbb{R}^d$ (in the limit $\beta \to \infty$, the measure ν is indeed a Dirac mass on the global minimum of V): a global optimization method should be able to find the global minimum among all the local minima of the function (these local minima correspond to different modes of the associated Boltzmann distribution (1.5)).

1.2.4. Other computational challenges

As explained above, the focus of this review is on sampling methods, and on PDE techniques to assess the quality of these methods. Before presenting how PDEs naturally appear in such a context, we would like to briefly discuss other computational challenges raised by molecular dynamics computations, though we will not discuss this further.

The integration of the deterministic Hamiltonian dynamics raises many interesting numerical questions. The original motivation was to understand the excellent energy conservation properties of the Störmer–Verlet method (Verlet 1967). With only one force evaluation at each time step, this numerical method indeed exhibits excellent energy preservation over very large

time horizons. This stability property can be proved for the Störmer–Verlet method (and, more generally, symplectic integrators) using backward error analysis. This has been extended in various directions: for example, replacing symplecticity by symmetry, considering systems with constraints, or extending these results to Hamiltonian partial differential equations (see *e.g.* Faou 2012 for Schrödinger equations). We refer to Marsden and West (2001), Hairer, Lubich and Wanner (2003), Leimkuhler and Reich (2004), Hairer, Lubich and Wanner (2006) and Bond and Leimkuhler (2007) for an introduction to this branch of numerical analysis, called geometric numerical integration.

In order to reduce the dimensionality of the problem under consideration, so-called coarse-graining techniques have been devised. Hopefully, while reducing the number of degrees of freedom, one also expects to increase the stability threshold on the time step, and thus to be able to reach larger time horizons. Many ideas have been proposed along these lines. A typical idea, the basis of dissipative particle dynamics (Hoogerbrugge and Koelman 1992, Español and Warren 1995), is first to replace a molecule, or a group of molecules, by some effective particle, then to determine the interactions between such particles as some effective interaction obtained by averaging over the actual interactions of the fully detailed system it represents, and finally to add some stochastic fluctuations accounting for the microscopic variability that disappeared in the coarse-graining process. See also Kevrekidis and Samaey (2009) and Rohrdanz, Zheng, Maggioni and Clementi (2011), for example, for a few selected works on coarse-graining in molecular simulation. Mathematical analysis is needed in order to (i) derive good coarse-graining strategies with certified error bounds, and (ii) propose appropriate schemes to use these coarse-graining models. Coarse-grained models can be used as predictors, the results of which are then refined in a corrector step by resorting to more accurate models. Alternatively, it is possible to construct coupling strategies in space in order to use coarse-grained models in some places while using finer models in others – as in quasicontinuum methods, which couple atomistic simulations with finite element discretizations (Luskin and Ortner 2013). In some sense, the free energy introduced in Section 4 or the accelerated dynamics presented in Section 6.4 are also examples of coarse-graining strategies. In all instances of coarse-graining, the main difficulty consists in selecting the relevant degrees of freedom over which the coarse-graining procedure is performed (for instance, anticipating our later discussion, the reaction coordinate ξ for free energy computation, or the set S to exit from for metastable dynamics). Thus statistical methods (such as clustering, manifold learning or hidden Markov chain approaches) should play an increasingly important role in future, for example to extract the metastable states from a long trajectory; see *e.g.* Fischer *et al.* (2007).

To conclude this section, let us mention that the modelling issues raised by the quality of the empirical force fields become ever more crucial as longer time scales are reached. Notwithstanding the numerous examples where molecular dynamics computations have provided insights into molecular systems in biology or material sciences, many numerical experiments also show that better force fields need to be developed in order to reach the desirable level of accuracy in many applications. This includes improvement of the modelling of polarizability (see for instance Lipparini *et al.* 2015), the construction of dedicated force fields for water molecules (see Wallqvist and Mountain 1999 and Jorgensen and Tirado-Rives 2005 for early review articles), and the development of reactive potentials able to describe chemical reactions (with variants and improvements of the REBO and ReaxFF potentials mentioned in Section 1.1). Appropriate functional forms of the potential functions as well as parameters should be dictated by electronic structure calculations, and validated using experiments. In this context, mathematics should help to (i) solve the electronic structure calculation efficiently, and (ii) develop coupling approaches between quantum calculations and molecular dynamics computations. Advanced optimization procedures should be useful for constructing suitable force fields from many data (quantum computations, experiments). In particular, machine learning techniques play an increasingly crucial role in the field (Rupp 2015).

1.3. Improving numerical methods in molecular dynamics: how PDEs can help

In order to show how PDE techniques are useful for understanding and improving numerical techniques in molecular dynamics, we first recall the link between SDEs and PDEs. We then present the structure of the remainder of the article.

1.3.1. Link between SDEs and PDEs

Consider the general time-homogeneous stochastic differential equation on $\mathbb{R}^d$:

$$\mathrm{d}x_t = b(x_t)\,\mathrm{d}t + \sigma(x_t)\,\mathrm{d}W_t, \tag{1.11}$$

for a given initial condition $x_0 \in \mathbb{R}^d$ and standard Brownian motion $W_t \in \mathbb{R}^m$, and where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are assumed to be such that there exists a unique (strong) solution to (1.11). For example, the classical setting is to assume that b and σ are locally Lipschitz, which yields local-in-time existence up to a (stochastic) explosion time, and then to prove that the existence is actually global in time with appropriate Lyapunov functions; see for example Khasminskii (2012), Mao (2008) and Rey-Bellet (2006).

The stochastic differential equation (1.11) is naturally associated with a differential operator:

$$\mathcal{L} = b \cdot \nabla + \frac{1}{2}\sigma\sigma^T : \nabla^2, \tag{1.12}$$

which is the *infinitesimal generator* of the Markov process $(x_t)_{t\geqslant 0}$ solution to (1.11). Here, the symbol : denotes the Frobenius inner product. More explicitly, for a given C^∞ test function $\varphi : \mathbb{R}^d \to \mathbb{R}$, the operator $\mathcal{L}$ acts via

$$\mathcal{L}\varphi = \sum_{i=1}^{d} b_i \partial_{x_i}\varphi + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{m} \sigma_{i,k}\sigma_{j,k}\partial_{x_i,x_j}\varphi.$$

For example, the generator of overdamped Langevin dynamics (1.8) is

$$\mathcal{L} = -\nabla V \cdot \nabla + \beta^{-1}\Delta = \beta^{-1}\exp(\beta V)\mathrm{div}[(\exp(-\beta V)\nabla\cdot], \tag{1.13}$$

while, for Langevin dynamics (1.9), we have

$$\mathcal{L} = M^{-1}p \cdot \nabla_q - \nabla V \cdot \nabla_p + \gamma(-M^{-1}p \cdot \nabla_p + \beta^{-1}\Delta_p). \tag{1.14}$$

Here and elsewhere, we will always consider generators to be defined on appropriate dense subsets of the domain of the generator, such as C^∞ functions with compact support or whose derivatives grow at most polynomially.

The link between (1.11) and (1.12) is made explicit in the following lemma.

Lemma 1.1. Let φ be a compactly supported C^∞ function. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathbb{E}^x(\varphi(x_t))]\Big|_{t=0} = \mathcal{L}\varphi(x), \tag{1.15}$$

where the superscript x indicates that the initial condition of (1.11) is $x_0 = x$, for a fixed $x \in \mathcal{X}$, and the expectation is taken over the Brownian paths.

The proof is a direct consequence of the Itô formula,

$$\mathrm{d}\varphi(x_t) = \mathcal{L}\varphi(x_t)\,\mathrm{d}t + (\nabla\varphi(x_t))^T\sigma(x_t)\,\mathrm{d}W_t,$$

and the fact that φ is C^∞ and compactly supported, so that

$$\mathbb{E}^x\left(\int_0^t (\nabla\varphi(x_s))^T\sigma(x_s)\,\mathrm{d}W_s\right) = 0$$

because the argument of the expectation is a square-integrable martingale (Revuz and Yor 1999). Therefore

$$\mathbb{E}^x[\varphi(x_t)] - \varphi(x) = \int_0^t \mathbb{E}^x[(\mathcal{L}\varphi)(x_s)]\,\mathrm{d}s, \tag{1.16}$$

and the following limit is well defined:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbb{E}^x[\varphi(x_t)])\Big|_{t=0} = \lim_{t\to 0}\frac{\mathbb{E}^x[\varphi(x_t)] - \varphi(x)}{t} = \mathcal{L}\varphi(x).$$

The infinitesimal generator $\mathcal{L}$ associated with the stochastic dynamics plays a crucial role in the following since it is the natural differential operator that appears in three partial differential equations related to the stochastic dynamics: (i) the Fokker–Planck dynamics that governs the evolution of the law of x_t (see Section 2.1), (ii) the Poisson equation for the asymptotic variance of estimators based on time averages (see Section 3.1), and (iii) the Feynman–Kac formula (see Section 6.1).

Let us conclude this section by giving a very important characterization of the invariance of a probability measure by a stochastic evolution in terms of the generator: π is an invariant probability measure for (1.11) if and only if for any C^∞ and compactly supported test function φ,

$$\int_{\mathcal{X}} \mathcal{L}\varphi \, \mathrm{d}\pi = 0. \tag{1.17}$$

This characterization directly follows from (1.15) by integrating both sides with respect to π, and using that $\mathbb{E}(\varphi(x_t)) = \mathbb{E}(\varphi(x_0))$ when $x_0 \sim \pi$ since, by definition of the invariance of the measure π, we have $x_t \sim \pi$. (Here and in the following, the notation $x \sim \pi$ means that the random variable x is distributed according to the probability measure π.)

1.3.2. Assessing the efficiency and the accuracy of sampling algorithms
For the estimation of thermodynamic quantities, PDE techniques can be used in the following contexts.

- Determining (upper bounds to) the rate of convergence to equilibrium of stochastic dynamics. These rates of convergence can indeed be used to quantify the sampling efficiency of the underlying dynamics. The techniques used are typically functional inequalities allowing us to prove the existence of a spectral gap for the generator $\mathcal{L}$ (or equivalently the exponential convergence to 0 of the associated semigroup $\mathrm{e}^{t\mathcal{L}}$). Section 2 is devoted to these aspects for simple equilibrium dynamics. Extensions to more complicated dynamics are possible: in Section 4.3 we study the long-time behaviour of the solution to a non-linear PDE by an appropriate modification of one of the approaches used in the linear case (entropy techniques).

- Quantifying the errors in the computation of ensemble averages via ergodic averages over a single long trajectory (see Section 3). These errors are of two types: systematic errors arising from the bias on the invariant measure due to the time discretization of the continuous dynamics, and statistical errors governed by a central limit theorem. For systematic errors due to time discretization, the first PDE tool used for Langevin dynamics is a Trotter or Strang splitting of the semigroup to suggest new numerical schemes, or to analyse their accuracy.

Second, a PDE is also involved when considering the solution of a Poisson equation of the form $\mathcal{L}u = f$ to characterize the leading error term on the invariant measure. Poisson equations are also used to obtain the expression of the asymptotic variance dictating the magnitude of the statistical error.

For the estimation of dynamical quantities, PDEs are useful in the following contexts.

- Defining transport coefficients, expressed either as integrated correlation functions or given by the linear response of non-equilibrium dynamics in the limit of small non-equilibrium forcings (see Section 5). The PDE technique used is perturbation theory of linear operators.
- Understanding the properties of algorithms for the sampling of paths (see Section 6.3). The results rely on the properties of solutions to the Hamilton–Jacobi–Bellman equation for variance reduction techniques based on importance sampling, while the analysis of the exit problem for accelerated dynamics starting in a metastable state relies on quasi-stationary distributions, which are eigenfunctions of Fokker–Planck equations with Dirichlet boundary conditions.

One topic that is also relevant to the analysis of metastable processes, and that we will discuss only briefly in this review (see Section 6.2), is the determination and computation of the rate functional in large deviation results (Freidlin and Wentzell 1984, Olivieri and Vares 2005). Let us only mention that the rate functional is determined for Markov processes by the Donsker–Varadhan formula (obtained in a series of works starting with Donsker and Varadhan 1975), which involves the generator of the dynamics. In addition, the variance of time averages computed along a single realization of the dynamics can be computed from the second derivative of this rate functional (den Hollander 2000). This opens the way for variance reduction techniques based on large deviation results, relying on PDE techniques, as recently presented by Rey-Bellet and Spiliopoulos (2015).

1.3.3. Outline of this review

Section 2 gives the theoretical foundations needed in the next four sections. Sections 3 and 4 focus on the computation of thermodynamic quantities such as (1.3) and (1.4), while the estimation of dynamical quantities such as (1.7) is the subject of Sections 5 and 6. More precisely, we start in Section 2 by reviewing results on the long-time convergence of the laws of the stochastic processes under consideration. We then discuss in Section 3 the various errors arising in the computation of average properties along a single, long realization of some ergodic dynamics. Section 4 is devoted to

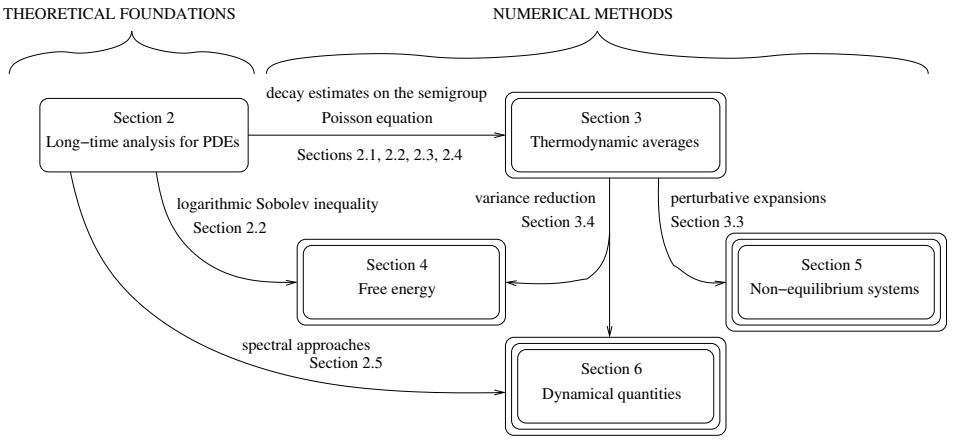


Figure 1.3. Interdependence of the sections. Double (respectively triple) boxes are for sections devoted to the computation of thermodynamic quantities (respectively dynamical quantities). In particular, Section 3 mainly uses decay estimates from Section 2 on the semigroup of the form $\|\mathrm{e}^{t\mathcal{L}}\|_{\mathcal{B}(E)} \leqslant C\,\mathrm{e}^{-\alpha t}$ for some positive constants C and α, which in turn implies the well-posedness of the Poisson equation $\mathcal{L}u = f$ in a certain Banach space E (see Proposition 2.1 and Corollaries 2.4, 2.16, 2.21 and 2.27).

the computation of free energy differences, for which dedicated techniques have to be used. The first dynamical quantities considered are transport coefficients in Section 5, numerically obtained either from dynamical information on equilibrium trajectories or from the linear response of steady-state non-equilibrium systems. Finally, in Section 6, we turn to the sampling of trajectories between metastable states.

In Figure 1.3 we provide an overview of the interdependence of the sections, with an emphasis on the theoretical tools from Section 2 needed in the next four sections. The reader who is mostly interested in numerical methods can directly attack any of the four Sections 3, 4, 5 or 6, which are essentially independent, coming back to Section 2 when some theoretical background is needed.

Before getting to the heart of the matter, let us emphasize again that the focus of this review is on sampling methods used in molecular dynamics and which can be analysed using partial differential equations. Other sampling methods will therefore be mentioned only briefly, even though they may be equally important in practice. Moreover, for the sake of brevity, we sometimes present only the main arguments or ideas without the full mathematical details. In both cases, we give extensive references for further reading.

2. Long-time analysis for partial differential equations: some standard approaches

As explained above, one approach to studying the efficiency of a sampling method is to study the rate of convergence to equilibrium of the law at time t of the process. The analysis techniques that have been developed to study the long-time behaviour of solutions to partial differential equations are very useful in this context. The law at time t of a stochastic process $(x_t)_{t\geqslant 0}$, following the general stochastic differential equation (1.11), indeed satisfies a partial differential equation called the Fokker–Planck equation (see Section 2.1). In this section we review standard methods to prove the convergence of the law at time t of stochastic processes to some invariant probability measure (in our context, typically the canonical measure).

These convergence results also provide bounds on the inverse of the generator, which turn out to be useful in various contexts, for example to define asymptotic variances of ergodic averages (Section 3.1), to make precise biases on invariant measures of discretizations of continuous dynamics (Section 3.3), or to state linear response results for the invariant measure of perturbed dynamics (Section 5.2). Some of the convergence results can also be transferred to discrete-time dynamics, which allows us to ensure the existence and uniqueness of invariant probability measures for discretizations of SDEs, or to compute the variance of the corresponding ergodic averages.

Some techniques to analyse the long-time behaviour of the stochastic process are limited to specific structural assumptions on the dynamics: Poincaré inequalities and logarithmic Sobolev inequalities are useful for randomly perturbed gradient dynamics such as overdamped Langevin dynamics (see Section 2.2), while hypocoercivity estimates are well suited to Langevin dynamics (see Section 2.3). On the other hand, some processes, such as those encountered by non-reversible perturbations of Langevin and overdamped Langevin dynamics, may not have the appropriate structural properties for the previously mentioned techniques. In this case, Lyapunov-type approaches turn out to be versatile methods (see Section 2.4). Finally, in Section 2.5, we describe how results on the spectrum of the infinitesimal generator of a stochastic process in the asymptotic regime $\beta \to \infty$ can be useful for studying a metastable process.

2.1. The Kolmogorov forward equation

Let us consider a stochastic process $(x_t)_{t\geqslant 0}$ in $\mathbb{R}^d$, following the dynamics (1.11). Let us assume for simplicity that the law of x_t at time t has a density ψ with respect to Lebesgue measure:[1] x_t is distributed according

[1] The fact that x_t admits a density with respect to Lebesgue measure for $t > 0$ can be proved in our context for Langevin and overdamped Langevin dynamics using the

to $\psi(t,x)\,\mathrm{d}x$. We also let $\psi(0,x)=\psi_0(x)$ denote the density of the initial distribution. Then ψ satisfies the Kolmogorov forward equation, also known as the Fokker–Planck equation:

$$\partial_t \psi = \mathcal{L}^\dagger \psi, \quad \psi(0)=\psi_0, \tag{2.1}$$

where $\mathcal{L}^\dagger$ denotes the L^2 adjoint of the operator $\mathcal{L}$:

$$\mathcal{L}^\dagger = -\mathrm{div}(b\,\cdot) + \frac{1}{2}\nabla^2 : (\sigma\sigma^T\cdot).$$

Explicitly, for a given C^∞ test function $\varphi : \mathbb{R}^d \to \mathbb{R}$,

$$\mathcal{L}^\dagger \psi = -\sum_{i=1}^d \partial_{x_i}(b_i\psi) + \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^m \partial_{x_i,x_j}(\sigma_{i,k}\sigma_{j,k}\psi).$$

In particular, for overdamped Langevin dynamics (1.8), the density $\psi(t,q)$ satisfies (2.1) with

$$\mathcal{L}^\dagger \psi = \mathrm{div}(\nabla V \psi) + \frac{1}{\beta}\Delta\psi = \frac{1}{\beta}\mathrm{div}[\mathrm{e}^{-\beta V}\nabla(\mathrm{e}^{\beta V}\psi)]. \tag{2.2}$$

For Langevin dynamics (1.9), the density $\psi(t,q,p)$ satisfies (2.1) with

$$\mathcal{L}^\dagger \psi = -p^T M^{-1}\nabla_q \psi + \nabla V^T \nabla_p \psi + \gamma\,\mathrm{div}_p\left(M^{-1}p\,\psi + \frac{1}{\beta}\nabla_p\psi\right). \tag{2.3}$$

The Fokker–Planck equation (2.1) is a consequence of (1.16). Indeed, for $h>0$,

$$\frac{\mathbb{E}^x[\varphi(x_{t+h})] - \mathbb{E}^x[\varphi(x_t)]}{h} = \frac{1}{h}\int_t^{t+h} \mathbb{E}^x[(\mathcal{L}\varphi)(x_s)]\,\mathrm{d}s,$$

which, after integration with respect to ψ_0, gives

$$\begin{aligned}\frac{1}{h}\left(\int_{\mathbb{R}^d}\varphi(x)\psi(t+h,x)\,\mathrm{d}x - \int_{\mathbb{R}^d}\varphi(x)\psi(t,x)\,\mathrm{d}x\right)\\ = \frac{1}{h}\int_t^{t+h}\int_{\mathbb{R}^d}(\mathcal{L}\varphi)(x)\psi(s,x)\,\mathrm{d}x\,\mathrm{d}s.\end{aligned}$$

This leads, in the limit $h\to 0$, to the following equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathbb{R}^d}\varphi(x)\psi(t,x)\,\mathrm{d}x\right) = \int_{\mathbb{R}^d}(\mathcal{L}\varphi)(x)\psi(t,x)\,\mathrm{d}x, \tag{2.4}$$

which is satisfied for all test functions φ (say, C^∞ and compactly supported).

hypoellipticity and the ellipticity of the infinitesimal generator (for hypoellipticity, see Section 2.3 below), respectively. If x_t does not admit a density, the Fokker–Planck equation still holds in a weak sense: see (2.4).

This is indeed a weak formulation of (2.1). For rigorous results concerning the existence and uniqueness of a solution to (2.1) and the link with stochastic differential equations, we refer for example to Friedman (1975).

Let us mention that the invariance of the canonical measure ν (resp. μ) for overdamped Langevin dynamics (resp. Langevin dynamics) is easily checked. A simple computation indeed shows that the canonical measures[2] are stationary solutions to (2.1), that is,

$$\operatorname{div}(\nabla V \nu) + \beta^{-1}\Delta\nu = 0, \tag{2.5}$$

and

$$-M^{-1}p \cdot \nabla_q \mu + \nabla V \cdot \nabla_p \mu + \gamma \operatorname{div}(M^{-1}p\mu + \beta^{-1}\nabla_p \mu) = 0, \tag{2.6}$$

respectively.

When a reference probability measure π is given (typically, the canonical measure for Langevin or overdamped Langevin dynamics), it is sometimes useful to consider the adjoint of the generator $\mathcal{L}$ in the weighted L^2-space

$$L^2(\pi) = \left\{ \varphi : \mathcal{X} \to \mathbb{R} \text{ measurable} \;\middle|\; \int_{\mathcal{X}} |\varphi|^2 \, \mathrm{d}\pi < \infty \right\}.$$

We let $\mathcal{L}^*$ denote the adjoint of $\mathcal{L}$ on $L^2(\pi)$. For any C^∞ and compactly supported functions φ, ϕ,

$$\int_{\mathcal{X}} (\mathcal{L}\varphi)\phi \, \mathrm{d}\pi = \int_{\mathcal{X}} \varphi(\mathcal{L}^*\phi) \, \mathrm{d}\pi. \tag{2.7}$$

In this framework, the invariance of π for the dynamics is formulated as $\mathcal{L}^*\mathbf{1} = 0$, where $\mathbf{1}$ is the constant function equal to 1. This result is obtained by setting $\phi = \mathbf{1}$ and using (1.17). For the overdamped Langevin dynamics (1.8) with invariant measure ν and generator $\mathcal{L}$ defined by (1.13), one can check that $\mathcal{L}^* = \mathcal{L}$. For the Langevin dynamics (1.9) with invariant measure μ and generator $\mathcal{L}$ defined by (1.14), one can check that

$$\mathcal{L}^* = -M^{-1}p \cdot \nabla_q + \nabla V \cdot \nabla_p + \gamma(-M^{-1}p \cdot \nabla_p + \beta^{-1}\Delta_p).$$

In the following sections we analyse the convergence to the canonical measure of the solution ψ to the Fokker–Planck equation (2.1), for both overdamped Langevin and Langevin dynamics. Using a semigroup representation, the question can be rephrased as studying the convergence of $\mathrm{e}^{t\mathcal{L}^\dagger}\psi_0$ to the invariant measure π (which is ν or μ depending on the context).

[2] In the following equations, we denote, with some abuse of notation, the density of the canonical measures with respect to Lebesgue measure by the same symbol as the measures themselves.

2.1.1. A dual viewpoint

It is in fact sometimes useful to adopt a dual viewpoint, and to study the convergence of observables rather than probability densities. This amounts to analysing the convergence of $\mathrm{e}^{t\mathcal{L}}\varphi$ to $\int_{\mathcal{X}} \varphi \, \mathrm{d}\pi$. To make this remark precise, let us first note that[3] $\mathrm{e}^{t\mathcal{L}}$ indeed encodes the time evolution of observables since

$$(\mathrm{e}^{t\mathcal{L}}\varphi)(x) = \mathbb{E}^x(\varphi(x_t)). \tag{2.8}$$

To prove (2.8) we start from Itô's formula: for a fixed time $t > 0$, for a C^∞ function φ, and $s \in [0,t]$, we have

$$\mathrm{d}(\mathrm{e}^{(t-s)\mathcal{L}}\varphi(x_s)) = (-\partial_t + \mathcal{L})(\mathrm{e}^{(t-s)\mathcal{L}}\varphi)(x_s)\,\mathrm{d}s + \nabla(\mathrm{e}^{(t-s)\mathcal{L}}\varphi)\cdot\sigma(x_s)\,\mathrm{d}W_s,$$

so that

$$\mathbb{E}^x(\varphi(x_t)) - \mathrm{e}^{t\mathcal{L}}\varphi(x) = \mathbb{E}^x\left(\int_0^t \nabla(\mathrm{e}^{(t-s)\mathcal{L}}\varphi)\cdot\sigma(x_s)\,\mathrm{d}W_s\right) = 0.$$

For a general stochastic differential equation, the exponential decay to equilibrium of $\mathrm{e}^{t\mathcal{L}}$, seen as a bounded operator on a reflexive Banach space $E \subset L^2$, is equivalent to the exponential decay to equilibrium of $\mathrm{e}^{t\mathcal{L}^\dagger}$ seen as a bounded operator on E', the dual of E. This equivalence is a consequence of the following equality: for C^∞ compactly supported test functions φ and ψ_0,

$$\mathbb{E}(\varphi(x_t)) = \int_{\mathcal{X}} \mathrm{e}^{t\mathcal{L}}\varphi(x_0)\psi_0(x_0)\,\mathrm{d}x_0 = \int_{\mathcal{X}} \varphi(x)\,\mathrm{e}^{t\mathcal{L}^\dagger}\psi_0(x)\,\mathrm{d}x,$$

where $(x_t)_{t\geqslant 0}$ satisfies (1.11) and the law of x_0 is $\psi_0(x)\,\mathrm{d}x$. In the particular case of overdamped Langevin and Langevin dynamics, the long-time behaviours of $\mathrm{e}^{t\mathcal{L}}$ and $\mathrm{e}^{t\mathcal{L}^\dagger}$ are in fact closely related thanks to simple algebraic identities: see respectively (2.15) and (2.36) below.

Note also that, upon replacing the observable φ with $\varphi - \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi$, it suffices to understand the convergence to 0 of $\mathrm{e}^{t\mathcal{L}}\varphi$ for functions φ with average 0 with respect to π.

2.1.2. Exponential convergence of semigroups and definition of inverses

We introduce the operator norm on the vector space $\mathcal{B}(E)$ of bounded linear operators on a Banach space E, defined by

$$\|\mathcal{A}\|_{\mathcal{B}(E)} = \sup_{g\neq 0} \frac{\|\mathcal{A}g\|_E}{\|g\|_E}.$$

Recall that $\mathcal{B}(E)$ equipped with this norm is a Banach space. As discussed

[3] This equality is a particular case of the Feynman–Kac formula, which will be presented in Section 6.1.

at the beginning of this section, it is important in various contexts to obtain bounds on the inverse of the generator $\mathcal{L}$ on E. The Banach space E is typically a subspace of functions with average 0 with respect to the invariant measure, otherwise $\mathrm{e}^{t\mathcal{L}}\varphi$ does not converge to 0 as $t \to +\infty$.

The invertibility of the generator, and bounds on this inverse in $\mathcal{B}(E)$, can be deduced from decay estimates on the semigroup, as is made precise in the following result.

Proposition 2.1. Assume that

$$\|\mathrm{e}^{t\mathcal{L}}\|_{\mathcal{B}(E)} \leqslant C\,\mathrm{e}^{-\lambda t}. \tag{2.9}$$

Then the operator $\mathcal{L}$ is invertible on E, and

$$\|\mathcal{L}^{-1}\|_{\mathcal{B}(E)} \leqslant \frac{C}{\lambda}. \tag{2.10}$$

Proof. Introducing

$$U = -\int_0^{+\infty} \mathrm{e}^{t\mathcal{L}}\,\mathrm{d}t,$$

a simple computation shows that, for $\varphi \in D(\mathcal{L}) = \{\phi \in E \mid \mathcal{L}\phi \in E\}$,

$$\mathcal{L}U\varphi = U\mathcal{L}\varphi = -\int_0^{+\infty} \frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{e}^{t\mathcal{L}}\varphi] = \varphi.$$

These equalities show that $\mathcal{L}$ is invertible on E, with $U = \mathcal{L}^{-1}$. From the definition of U, it follows that

$$\|U\|_{\mathcal{B}(E)} \leqslant \int_0^{+\infty} \|\mathrm{e}^{t\mathcal{L}}\|_{\mathcal{B}(E)}\,\mathrm{d}t \leqslant C\int_0^{+\infty} \mathrm{e}^{-\lambda t}\,\mathrm{d}t = \frac{C}{\lambda},$$

which proves the resolvent bound (2.10). □

As a final remark, let us emphasize that the manipulations performed here for continuous-in-time dynamics can also be appropriately modified for discrete-in-time dynamics, such as encountered in the discretization of SDEs; see Corollary 2.26 below.

2.2. Poincaré and logarithmic Sobolev inequalities

We start by considering overdamped Langevin dynamics (1.8), the generator of which reads

$$\mathcal{L}\phi = -\nabla V \cdot \nabla\phi + \frac{1}{\beta}\Delta\phi = \frac{\mathrm{e}^{\beta V}}{\beta}\mathrm{div}(\mathrm{e}^{-\beta V}\nabla\phi). \tag{2.11}$$

The associated Fokker–Planck equation is the parabolic PDE

$$\partial_t\psi = \mathcal{L}^\dagger\psi, \quad \psi(0) = \psi_0, \tag{2.12}$$

where $\mathcal{L}^\dagger$ is introduced in (2.2), and the initial condition[4] $\psi_0 \geqslant 0$ satisfies the normalization condition

$$\int_{\mathcal{D}} \psi_0 = 1.$$

We checked in (2.5) that the probability measure

$$\nu(\mathrm{d}q) = Z_\nu^{-1} \exp(-\beta V(q))\, \mathrm{d}q \tag{2.13}$$

is a stationary solution of the Fokker–Planck equation. It is therefore relevant to ask whether ψ converges to ν and, if so, at what rate.

Before presenting several answers to this question, let us first give two reformulations of the Fokker–Planck equation (2.12). A simple computation shows that the function $g = \psi\, \mathrm{e}^{\beta V/2}$ satisfies the Schrödinger-type equation

$$\partial_t g = \frac{1}{\beta}\Delta g + Wg, \quad W = \frac{1}{2}\left(\Delta V - \frac{\beta}{2}|\nabla V|^2\right), \tag{2.14}$$

together with the normalization condition

$$\int_{\mathcal{D}} g(t)\, \mathrm{e}^{-\beta V/2} = 1, \quad \text{for all } t \geqslant 0.$$

In addition, the function $f = \psi\, \mathrm{e}^{\beta V}$ satisfies

$$\partial_t f = \mathcal{L} f, \quad \int_{\mathcal{D}} f\, \mathrm{e}^{-\beta V} = 1. \tag{2.15}$$

In this last setting, the functional space to consider is the Hilbert space

$$L^2(\nu) = \left\{ \varphi \text{ measurable} \,\middle|\, \int_{\mathcal{D}} |\varphi|^2 \, \mathrm{d}\nu < +\infty \right\},$$

equipped with the scalar product[5]

$$\langle \varphi_1, \varphi_2 \rangle_{L^2(\nu)} = \int_{\mathcal{D}} \varphi_1\, \varphi_2 \, \mathrm{d}\nu.$$

A simple computation shows that $\mathcal{L}$ is symmetric on $L^2(\nu)$. More precisely, for test functions φ_1, φ_2 (C^∞ and compactly supported),

$$\langle \mathcal{L}\varphi_1, \varphi_2 \rangle_{L^2(\nu)} = \langle \varphi_1, \mathcal{L}\varphi_2 \rangle_{L^2(\nu)} = -\frac{1}{\beta} \int_{\mathcal{D}} \nabla\varphi_1 \cdot \nabla\varphi_2 \, \mathrm{d}\nu. \tag{2.16}$$

In fact, it can be proved that $\mathcal{L}$ is self-adjoint on $L^2(\nu)$: see *e.g.* Bakry, Gentil and Ledoux (2014).

[4] For simplicity, we consider initial conditions with densities with respect to Lebesgue measure. By parabolic regularization, any initial positive measure admits a density for positive times.

[5] To keep the notation simple, we restrict ourselves to real-valued scalar products, although, for spectral theory, complex-valued scalar products should be considered. However, the extension to the latter case is straightforward.

2.2.1. Poincaré inequalities and convergence in $L^2(\nu)$

We start by considering the Fokker–Planck equation reformulated as (2.15), for an initial condition $f_0 = \psi_0/\nu$ belonging to $L^2(\nu)$. Recall that

$$f_0 \geqslant 0, \quad \int_{\mathcal{D}} f_0 \, \mathrm{d}\nu = 1.$$

The simplest setting to consider is when a so-called Poincaré inequality holds; see for instance Bakry *et al.* (2014, Section 4) for a very nice introduction to these inequalities. In this case, $f(t) = \mathrm{e}^{t\mathcal{L}} f_0$ converges exponentially fast in $L^2(\nu)$ to the constant function $\mathbf{1}$. This can be rephrased as the exponential convergence of $\mathrm{e}^{t\mathcal{L}}(f_0 - \mathbf{1})$ to 0 in $L^2(\nu)$.

In order to state the convergence result, we introduce the following definition.

Definition 2.2 (Poincaré inequality). Consider the functional spaces

$$L^2_0(\nu) = \left\{ \varphi \in L^2(\nu) \,\middle|\, \int_{\mathcal{D}} \varphi \, \mathrm{d}\nu = 0 \right\},$$

and

$$H^1(\nu) = \left\{ \varphi \in L^2(\nu) \,\middle|\, \nabla\varphi \in (L^2(\nu))^d \right\}.$$

The measure ν is said to satisfy a Poincaré inequality with constant $R > 0$ when

$$\|\varphi\|^2_{L^2(\nu)} \leqslant \frac{1}{2R} \|\nabla\varphi\|^2_{L^2(\nu)}, \quad \text{for all } \varphi \in H^1(\nu) \cap L^2_0(\nu). \tag{2.17}$$

The constant $R > 0$ depends on the potential V, the inverse temperature β and the domain $\mathcal{D}$. We discuss below various sufficient conditions for ν to satisfy a Poincaré inequality.

The inequality (2.17) implies (and in fact is equivalent to) the exponential convergence to 0 of the semigroup $\mathrm{e}^{t\mathcal{L}}$ considered as an operator on $L^2_0(\nu)$.

Proposition 2.3. The measure ν satisfies a Poincaré inequality with constant $R > 0$ if and only if

$$\|\mathrm{e}^{t\mathcal{L}}\|_{\mathcal{B}(L^2_0(\nu))} \leqslant \mathrm{e}^{-2Rt/\beta}. \tag{2.18}$$

Proof. Let us first assume that the measure ν satisfies a Poincaré inequality with constant $R > 0$. In view of (2.16),

$$-\langle \mathcal{L}\varphi, \varphi \rangle_{L^2(\nu)} = \frac{1}{\beta} \|\nabla\varphi\|^2_{L^2(\nu)} \geqslant \frac{2R}{\beta} \|\varphi\|^2_{L^2(\nu)}, \quad \text{for all } \varphi \in L^2_0(\nu) \cap H^1(\nu). \tag{2.19}$$

Since 0 is an eigenvalue of the operator $\mathcal{L}$ (whose associated eigenvectors are constant functions), this inequality shows that the spectral gap of the self-adjoint operator $-\mathcal{L}$ on $L^2(\nu)$ is larger than or equal to $2R/\beta$ (see (2.82)

and the discussion following this equality). The inequality (2.19) also gives the exponential decrease of the semigroup on $L^2_0(\nu)$ since

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\|\mathrm{e}^{t\mathcal{L}}\varphi\|^2_{L^2(\nu)}\right) = \langle \mathrm{e}^{t\mathcal{L}}\varphi, \mathcal{L}\,\mathrm{e}^{t\mathcal{L}}\varphi\rangle_{L^2(\nu)} \leqslant -\frac{2R}{\beta}\|\mathrm{e}^{t\mathcal{L}}\varphi\|^2_{L^2(\nu)}. \tag{2.20}$$

By a Gronwall inequality, it follows that

$$\|\mathrm{e}^{t\mathcal{L}}\varphi\|_{L^2_0(\nu)} \leqslant \mathrm{e}^{-2Rt/\beta}\|\varphi\|_{L^2_0(\nu)}, \quad \text{for any } \varphi \in L^2_0(\nu).$$

Assume now that $\|\mathrm{e}^{t\mathcal{L}}\|_{\mathcal{B}(L^2_0(\nu))} \leqslant \mathrm{e}^{-2Rt/\beta}$. Then, for a given $\varphi \in L^2_0(\nu)$ and any $t > 0$,

$$\frac{\|\mathrm{e}^{t\mathcal{L}}\varphi\|^2_{L^2_0(\nu)} - \|\varphi\|^2_{L^2_0(\nu)}}{t} \leqslant \|\varphi\|^2_{L^2_0(\nu)}\,\frac{\mathrm{e}^{-4Rt/\beta} - 1}{t}.$$

We next pass to the limit $t \to 0$, using the equalities in (2.19) and (2.20):

$$-\frac{2}{\beta}\|\nabla\varphi\|^2_{L^2(\nu)} \leqslant -\frac{4R}{\beta}\|\varphi\|^2_{L^2_0(\nu)}.$$

Since $\varphi \in L^2_0(\nu)$ is arbitrary, the Poincaré inequality (2.17) follows. □

In view of Proposition 2.1, a useful corollary of the decay estimates on the semigroup is the following result.

Corollary 2.4. Assume that ν satisfies a Poincaré inequality with constant R. Then the operator $\mathcal{L}$ is invertible on $L^2_0(\nu)$, and the following equality holds in $\mathcal{B}(L^2_0(\nu))$:

$$\mathcal{L}^{-1} = -\int_0^{+\infty} \mathrm{e}^{t\mathcal{L}}\,\mathrm{d}t.$$

Moreover,

$$\|\mathcal{L}^{-1}\|_{\mathcal{B}(L^2_0(\nu))} \leqslant \frac{\beta}{2R}. \tag{2.21}$$

Let us now give some sufficient conditions to obtain Poincaré inequalities. First, such inequalities are easily obtained for probability measures equivalent to the Lebesgue measure, on connected domains, as a consequence of the standard Poincaré–Wirtinger inequality. This is the case when $\mathcal{D} = (L\mathbb{T})^d$, for example. Indeed, consider a measure with density $\rho(q)$ with respect to Lebesgue measure, and such that there exist two positive constants $\rho_{\min}$ and $\rho_{\max}$ with $0 < \rho_{\min} \leqslant \rho(q) \leqslant \rho_{\max}$ for all $q \in \mathcal{D}$. First, note that

$$\min_{c\in\mathbb{R}} \int_{\mathcal{D}} (\varphi - c)^2\,\mathrm{d}\rho = \int_{\mathcal{D}} (\varphi - \mathbb{E}_\rho(\varphi))^2\,\mathrm{d}\rho, \quad \text{for all } \varphi \in L^2(\rho).$$

Therefore,

$$\min_{c\in\mathbb{R}} \int_{\mathcal{D}} (\varphi - c)^2 \,\mathrm{d}\rho \leqslant \rho_{\max} \min_{c\in\mathbb{R}} \int_{\mathcal{D}} (\varphi - c)^2 \leqslant \frac{1}{2R_{\mathrm{PW}}} \rho_{\max} \int_{\mathcal{D}} |\nabla\varphi|^2 \leqslant \frac{1}{2R_{\mathrm{PW}}} \frac{\rho_{\max}}{\rho_{\min}} \int_{\mathcal{D}} |\nabla\varphi|^2 \,\mathrm{d}\rho,$$

where R_{PW} is the constant of the standard Poincaré–Wirtinger inequality on the bounded, connected domain $\mathcal{D}$. This shows that the canonical measure ν defined in (2.13) satisfies a Poincaré inequality when the domain $\mathcal{D}$ is bounded, with the following upper bound for the Poincaré constant R_ν:

$$R_\nu \leqslant R_{\mathrm{PW}}\, \mathrm{e}^{-\beta(V_{\max}-V_{\min})},$$

where

$$V_{\max} = \max_{q\in\mathcal{D}} V(q), \quad V_{\min} = \min_{q\in\mathcal{D}} V(q).$$

In fact, this argument can be made more general to show that Poincaré inequalities satisfy some stability property under bounded perturbations of the measure, similar to the result obtained in Theorem 2.11 below for logarithmic Sobolev inequalities.

On unbounded domains, some conditions on the potential V are required to have a Poincaré inequality. For example, such an inequality holds when V is uniformly convex (*i.e.*, $\nabla^2 V \geqslant R\,\mathrm{Id}_d$ with $R > 0$), similar to the result stated in Theorem 2.10 below for logarithmic Sobolev inequalities. Another result is based on the growth at infinity of the potential: see Bakry, Barthe, Cattiaux and Guillin (2008, Corollary 1.6).

Theorem 2.5. Consider $V \in C^2(\mathbb{R}^d)$. If there exists $a \in (0,1)$, $c > 0$ and $R \geqslant 0$ such that

$$a\beta|\nabla V(q)|^2 - \Delta V(q) \geqslant c, \quad \text{for all } q \in \mathbb{R}^d \text{ such that } |q| \geqslant R, \tag{2.22}$$

then ν satisfies a Poincaré inequality.

The above condition is satisfied, for instance, for potentials of the form

$$V(q) = K|q|^n + \widetilde{V}(q),$$

where $n \geqslant 1$ and $\widetilde{V}$ is C^∞ and such that

$$\lim_{|q|\to+\infty} \frac{|\widetilde{V}(q)|}{|q|^n} = 0 \quad \text{and} \quad \lim_{|q|\to+\infty} \frac{|\nabla\widetilde{V}|^2 + |\Delta\widetilde{V}|}{|q|^{2(n-1)}} = 0.$$

The first condition implies that $\mathrm{e}^{-\beta V}$ is integrable, while the second one ensures that (2.22) holds.

A useful result that we shall exploit below is that it is possible to deduce Poincaré inequalities by tensorization.

Proposition 2.6. If a probability measure is a product of d probability measures satisfying Poincaré inequalities with constants R_i, then the product measure satisfies a Poincaré inequality with constant

$$\min(R_1, \dots, R_d).$$

Proof. To simplify the notation, we prove the result in the case $d = 2$, for a probability measure $\rho(\mathrm{d}q_1\ \mathrm{d}q_2) = \rho_1(\mathrm{d}q_1)\,\rho_2(\mathrm{d}q_2)$, defined on $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$. Consider a C^∞ function φ with compact support, and average 0 with respect to ρ. We first introduce the partial average

$$\overline{\varphi}(q_2) = \int_{\mathcal{D}_1} \varphi(q_1', q_2)\,\rho_1(\mathrm{d}q_1').$$

Since ρ_1 satisfies a Poincaré inequality with constant R_1, and using the fact that the function $q_1 \mapsto \varphi(q_1, q_2) - \overline{\varphi}(q_2)$ has average 0 with respect to $\rho_1(\mathrm{d}q_1)$ for any value of q_2, we obtain

$$\int_{\mathcal{D}_1} |\varphi(q_1, q_2) - \overline{\varphi}(q_2)|^2 \rho_1(\mathrm{d}q_1) \leqslant \frac{1}{2R_1} \int_{\mathcal{D}_1} |\nabla_{q_1}\varphi(q_1, q_2)|^2 \rho_1(\mathrm{d}q_1).$$

This inequality can be rewritten as

$$\int_{\mathcal{D}_1} |\varphi(q_1, q_2)|^2 \rho_1(\mathrm{d}q_1) \leqslant |\overline{\varphi}(q_2)|^2 + \frac{1}{2R_1} \int_{\mathcal{D}_1} |\nabla_{q_1}\varphi(q_1, q_2)|^2 \rho_1(\mathrm{d}q_1). \quad (2.23)$$

Now, since $\overline{\varphi}$ has average 0 with respect to ρ_2 because φ has average 0 with respect to ρ,

$$\begin{aligned}\int_{\mathcal{D}_2} |\overline{\varphi}(q_2)|^2 \rho_2(\mathrm{d}q_2) &\leqslant \frac{1}{2R_2} \int_{\mathcal{D}_2} |\nabla_{q_2}\overline{\varphi}(q_2)|^2 \rho_2(\mathrm{d}q_2) \\ &\leqslant \frac{1}{2R_2} \int_{\mathcal{D}} |\nabla_{q_2}\varphi(q_1, q_2)|^2 \rho_1(\mathrm{d}q_1)\,\rho_2(\mathrm{d}q_2),\end{aligned}$$

where we used a Cauchy–Schwarz inequality in the last step. Integration of (2.23) with respect to $\rho_2(\mathrm{d}q_2)$ finally leads to

$$\begin{aligned}\int_{\mathcal{D}} |\varphi(q_1, q_2)|^2 \rho(\mathrm{d}q) &\leqslant \frac{1}{2R_1} \int_{\mathcal{D}} |\nabla_{q_1}\varphi(q_1, q_2)|^2 \rho(\mathrm{d}q) \\ &\quad + \frac{1}{2R_2} \int_{\mathcal{D}} |\nabla_{q_2}\varphi(q_1, q_2)|^2 \rho(\mathrm{d}q) \\ &\leqslant \frac{1}{2\min(R_1, R_2)} \int_{\mathcal{D}} |\nabla_q \varphi|^2 \,\mathrm{d}\rho,\end{aligned}$$

from which we deduce that ρ satisfies a Poincaré inequality with constant $R = \min(R_1, R_2)$. □

Finally, anticipating later discussion, we will see in Proposition 2.12 that the logarithmic Sobolev inequality implies the Poincaré inequality.

2.2.2. Logarithmic Sobolev inequalities and convergence in $L^1(\nu)$

Here we give some background on logarithmic Sobolev inequalities, which can be used to show convergence to the equilibrium state. More material can be read in the review papers by Guionnet and Zegarlinski (2003), Ledoux (2001) and Arnold, Markowich, Toscani and Unterreiter (2001), and in Bakry *et al.* (2014, Section 5). The latter two works have a PDE approach which may be more appealing to readers accustomed to analytical frameworks. Other useful introductory references include Ané *et al.* (2000) and Villani (2003).

The relative entropy and the Fisher information between two measures are defined as follows.

Definition 2.7 (entropy and Fisher information). For two probability measures π_1 and π_2 defined on a space $\mathcal{X}$, and such that π_1 is absolutely continuous with respect to π_2 (denoted $\pi_1 \ll \pi_2$ in the following), the entropy of π_1 with respect to π_2 is

$$\mathcal{H}(\pi_1 \mid \pi_2) = \int_{\mathcal{X}} \ln\left(\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2}\right) \mathrm{d}\pi_1. \tag{2.24}$$

The Fisher information of π_1 with respect to π_2 is

$$\mathcal{I}(\pi_1 \mid \pi_2) = \int_{\mathcal{X}} \left|\nabla \ln\left(\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_2}\right)\right|^2 \mathrm{d}\pi_1. \tag{2.25}$$

Using strict convexity of $x \mapsto x \ln x$ and the fact that $x \ln x - x + 1 \geqslant 0$, it is easy to check that the entropy is non-negative. Moreover, it is zero if and only if the two probability measures are identical. This is also true for the Fisher information.

There are many possible definitions of the entropy besides the relative entropy $\mathcal{H}$ defined in (2.24). Some mathematical motivations for the use of the relative entropy $\mathcal{H}$ can be found in Markowich and Villani (2000); see also Chafaï (2015) for a historical perspective. This particular entropy may be of interest for the following extensivity property. The relative entropy of the distribution of N independent variables (or, in less probabilistic terms, the entropy of a tensorized measure) is the sum of the relative entropies of the distributions of each random variable. This suggests that the rate of convergence to equilibrium estimated with relative entropies for weakly dependent variables may remain stable when the number of variables becomes large (see for instance the paragraph on Kac's spectral problem in Villani 2002, Chapter 5). This extensivity is a consequence of the extensivity of the logarithm function involved in definition (2.24) (by which we mean that $\ln(f_1 f_2) = \ln(f_1) + \ln(f_2)$). This extensivity property makes the logarithmic Sobolev inequality and entropy approaches much more convenient for studying some nonlinear Fokker–Planck equations, as will become clear in Section 4.3.

Recall that the total variation between two measures, which reduces to the L^1-norm of the difference between the two densities when the two measures are absolutely continuous with respect to Lebesgue measure, can be bounded by the relative entropy. This is the Csiszár–Kullback inequality (see for example Ané *et al.* 2000, Royer 2007): introducing the total variation norm

$$\|\pi_1 - \pi_2\|_{\mathrm{TV}} = \sup_{\|\varphi\|_{L^\infty} \leqslant 1} \int_{\mathcal{X}} \varphi(\mathrm{d}\pi_1 - \mathrm{d}\pi_2), \tag{2.26}$$

it holds that

$$\|\pi_1 - \pi_2\|_{\mathrm{TV}} \leqslant \sqrt{2\mathcal{H}(\pi_1 \mid \pi_2)}. \tag{2.27}$$

In other words, an upper bound for the entropy between π_1 and π_2 yields an upper bound for a distance between π_1 and π_2 (even though the relative entropy is not a distance because it is not symmetric in its arguments).

We now present a way to obtain an estimate of the rate of convergence to zero of the entropy $\mathcal{H}(\psi(t,\cdot) \mid \nu)$ for the solutions of the Fokker–Planck equation (2.12). To this end we introduce the following functional inequality.

Definition 2.8 (logarithmic Sobolev inequality). A probability measure π_2 satisfies a logarithmic Sobolev inequality with constant $R > 0$ (or LSI(R) for short) if, for all probability measures π_1 such that $\pi_1 \ll \pi_2$,

$$\mathcal{H}(\pi_1 \mid \pi_2) \leqslant \frac{1}{2R}\mathcal{I}(\pi_1 \mid \pi_2). \tag{2.28}$$

The exponential decay of $\mathcal{H}(\psi(t,\cdot) \mid \nu)$ to 0 can then be shown provided ν satisfies a so-called logarithmic Sobolev inequality (LSI). In fact, there is some equivalence between the two notions, as is made precise in the following result, which is the equivalent of Proposition 2.3 where a similar result is obtained for Poincaré inequalities and exponential decay in some L^2-space.

Proposition 2.9. The measure ν satisfies LSI(R) if and only if, for any initial condition $\psi_0 \geqslant 0$ with integral 1 and finite relative entropy with respect to ν (*i.e.* $\mathcal{H}(\psi_0 \mid \nu) < +\infty$), the solution $\psi(t)$ of the Fokker–Planck equation (2.12) satisfies

$$\mathcal{H}(\psi(t) \mid \nu) \leqslant \mathcal{H}(\psi_0 \mid \nu) \exp(-2\beta^{-1}Rt), \quad \text{for all } t \geqslant 0. \tag{2.29}$$

From the Csiszár–Kullback inequality (2.27), it is then possible to deduce the exponential convergence to zero with rate $\beta^{-1}R$ of the norm $\|\psi(t,\cdot) - \nu\|_{L^1}$.

Proof. Note first that the Fokker–Planck equation (2.12) can be rewritten as

$$\partial_t \psi = \mathcal{L}^\dagger \psi = \frac{1}{\beta}\mathrm{div}\left(\nu \nabla\left(\frac{\psi}{\nu}\right)\right).$$

A straightforward computation shows that

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{H}(\psi(t,\cdot)\mid\nu)) = \int_{\mathcal{D}} \partial_t\psi\left(\ln\left(\frac{\psi}{\nu}\right)+1\right) = \int_{\mathcal{D}} \partial_t\psi\,\ln\left(\frac{\psi}{\nu}\right),$$

since

$$\int_{\mathcal{D}} \psi(t) = 1.$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{H}(\psi(t,\cdot)\mid\nu)) = \int_{\mathcal{D}} (\mathcal{L}^\dagger\psi)\,\ln\left(\frac{\psi}{\nu}\right) = -\frac{1}{\beta}\int_{\mathcal{D}} \nabla\left(\frac{\psi}{\nu}\right)\cdot\nabla\left[\ln\left(\frac{\psi}{\nu}\right)\right]\mathrm{d}\nu,$$

which can be summarized as

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{H}(\psi(t,\cdot)\mid\nu)) = -\frac{1}{\beta}\mathcal{I}(\psi(t,\cdot)\mid\nu). \tag{2.30}$$

Assume now that ν satisfies an LSI(R). The estimate (2.29) then directly follows from (2.30).

Assume conversely that (2.29) holds for any initial condition ψ_0 with finite relative entropy with respect to ν. Then,

$$\frac{\mathcal{H}(\psi(t)\mid\nu)-\mathcal{H}(\psi_0\mid\nu)}{t} \leqslant \mathcal{H}(\psi_0\mid\nu)\frac{\exp(-2\beta^{-1}Rt)-1}{t}.$$

In the limit $t\to 0$, one obtains, with (2.30),

$$-\frac{1}{\beta}\mathcal{I}(\psi_0\mid\nu) \leqslant -\frac{2R}{\beta}\,\mathcal{H}(\psi_0\mid\nu),$$

which indeed leads to (2.28). □

We now give several ways to obtain LSIs for measures of the form (2.13).

(i) When the potential V satisfies a strict convexity condition of the form $\mathrm{Hess}(V) \geqslant R\,\mathrm{Id}_d$ with $R>0$, then an LSI with constant βR holds for ν, as first shown in Bakry and Emery (1985) (see Theorem 2.10 below).

(ii) If $\psi_\infty = \prod_{i=1}^d \psi_\infty^i$ and each measure $\psi_\infty^i(q)\,\mathrm{d}q$ satisfies an LSI with constant R_i, then ψ_∞ satisfies an LSI with constant $R=\min\{R_1,\dots,R_d\}$ (see Gross 1975). This is the equivalent of Proposition 2.6 for LSI.

(iii) When an LSI with constant R is satisfied by $\nu(\mathrm{d}q) = Z_V^{-1}\,\mathrm{e}^{-V(q)}\,\mathrm{d}q$, then the modified measure

$$Z_{V+\widetilde{V}}^{-1}\,\mathrm{e}^{-(V(q)+\widetilde{V}(q))}\,\mathrm{d}q$$

(with $\widetilde{V}$ bounded) satisfies an LSI with constant $\widetilde{R} = R\,\mathrm{e}^{\inf\widetilde{V}-\sup\widetilde{V}}$. This property expresses some stability with respect to bounded perturbations (see Holley and Stroock 1987 and Theorem 2.11 below).

(iv) There are also results on an LSI for the measure when a marginal law and the corresponding conditional laws satisfy an LSI (see Blower and Bolley 2006), or when all the conditional laws satisfy an LSI under some weak coupling assumption (see Otto and Reznikoff 2007). Such results can be extended to the non-linear setting, that is, in the case when the marginal distribution is obtained for some variable $z = \xi(q) \in \mathbb{R}^m$ (with $m < d$): see Lelièvre (2009).

Let us now state more precisely, and prove, two fundamental results ensuring that an LSI holds for canonical measures of the form (1.5).

Theorem 2.10 (Bakry–Emery criterion). Assume that $\mathcal{D} = \mathbb{R}^d$ and $\nabla^2 V \geqslant R\,\mathrm{Id}_d$ for some constant $R > 0$. Then ν satisfies LSI(βR).

Proof. The idea is to differentiate the Fisher information with respect to time, in order to prove its exponential convergence to 0, and then to insert it in (2.30). Introduce $f(t) = \psi(t)/\nu$ for an arbitrary initial condition $f_0 = \psi_0/\nu$. In view of (2.15), this function evolves according to $\partial_t f = \mathcal{L} f$, so that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}[\mathcal{I}(\psi(t) \mid \nu)] &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathcal{D}} |\nabla \ln f(t)|^2 f(t)\,\mathrm{d}\nu\right) \\ &= 2\int_{\mathcal{D}} \nabla\left(\frac{\partial_t f(t)}{f(t)}\right)\cdot \nabla f(t)\,\mathrm{d}\nu + \int_{\mathcal{D}} |\nabla \ln f(t)|^2 \partial_t f(t)\,\mathrm{d}\nu \\ &= 2\int_{\mathcal{D}} \nabla\left(\frac{\mathcal{L} f(t)}{f(t)}\right)\cdot \nabla f(t)\,\mathrm{d}\nu + \int_{\mathcal{D}} |\nabla \ln f(t)|^2 \mathcal{L} f(t)\,\mathrm{d}\nu.\end{aligned}$$

Since

$$\mathcal{L}(\ln f) = \frac{\mathcal{L} f}{f} - \frac{|\nabla f|^2}{\beta f^2} = \frac{\mathcal{L} f}{f} - \frac{1}{\beta}|\nabla(\ln f)|^2,$$

we obtain, after some algebraic calculations,

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}[\mathcal{I}(\psi(t) \mid \nu)] = &\int_{\mathcal{D}} \nabla[\mathcal{L}(\ln f(t))]\cdot \nabla[\ln f(t)]\, f(t)\,\mathrm{d}\nu \\ &- \int_{\mathcal{D}} \mathcal{L}\left[\frac{1}{2}|\nabla(\ln f(t))|^2\right] f(t)\,\mathrm{d}\nu.\end{aligned}$$

Now, using

$$\nabla\left(\frac{1}{2}|\nabla h|^2\right) = \nabla^2 h \cdot \nabla h, \quad \Delta\left(\frac{1}{2}|\nabla h|^2\right) = \nabla^2 h : \nabla^2 h + \nabla(\Delta h)\cdot \nabla h,$$

it follows that

$$\begin{aligned}\mathcal{L}\left(\frac{1}{2}|\nabla h|^2\right) &= \frac{1}{\beta}(\nabla^2 h : \nabla^2 h + \nabla(\Delta h)\cdot\nabla h) - (\nabla V)^T(\nabla^2 h)\nabla h \\ &\geqslant \nabla(\mathcal{L}h)\cdot\nabla h + (\nabla h)^T(\nabla^2 V)\nabla h \\ &\geqslant \nabla(\mathcal{L}h)\cdot\nabla h + R|\nabla h|^2,\end{aligned}$$

where we have used $\nabla^2 h : \nabla^2 h \geqslant 0$ in the second line, and the assumption on $\nabla^2 V$ in the third one. We next replace h with $\ln f(t)$ and obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}[\mathcal{I}(\psi(t)\mid\nu)] \leqslant -R\int_{\mathcal{D}} |\nabla(\ln f(t))|^2 f(t)\,\mathrm{d}\nu = -R\mathcal{I}(\psi(t)\mid\nu).$$

This shows that the Fisher information converges exponentially fast to zero:

$$\mathcal{I}(\psi(t)\mid\nu) \leqslant \mathrm{e}^{-2Rt}\mathcal{I}(\psi(0)\mid\nu). \tag{2.31}$$

The next step is to prove that $\mathcal{H}(\psi(t,\cdot)\mid\nu)$ converges to 0 as $t\to+\infty$. To this end, we use the argument provided in Arnold *et al.* (2001). Since the operator $-\mathcal{L}$ is self-adjoint on $L^2(\nu)$ and positive by (2.19), its spectral measure P_λ is supported by $[0,+\infty)$; see Reed and Simon (1975*a*) and Dautray and Lions (1990) for the definition of the spectral measure. In addition, (2.19) also shows that 0 is a non-degenerate eigenvalue whose associated eigenvectors are constant functions (recall that since ν satisfies LSI(βR), ν also satisfies a Poincaré inequality with the same constant: see Proposition 2.12 below). Therefore,

$$f(t) = \mathrm{e}^{t\mathcal{L}}f_0 = P_0 f_0 + \int_{(0,+\infty)} \mathrm{e}^{-t\lambda}\,\mathrm{d}(P_\lambda f_0).$$

In fact, P_0 is the projector onto the eigenspace associated with the eigenvalue 0, so that $P_0 f_0 = \int_{\mathcal{D}} f_0\,\mathrm{d}\nu = 1$. Moreover,

$$\|f(t)-\mathbf{1}\|^2_{L^2(\nu)} = \int_{(0,+\infty)} \mathrm{e}^{-2t\lambda}\,m_{f_0}(\mathrm{d}\lambda), \quad m_{f_0}(\mathrm{d}\lambda) = \mathrm{d}\big(\|P_\lambda f_0\|^2_{L^2(\nu)}\big).$$

By dominated convergence, the integral on the right-hand side of the above equality converges to 0 as $t\to+\infty$.

The conclusion now follows by a time integration of (2.30) together with the decay estimate (2.31), which leads to

$$\mathcal{H}(\psi_0\mid\nu) - \mathcal{H}(\psi(t,\cdot)\mid\nu) \leqslant \frac{1}{\beta}\int_0^t \mathcal{I}(\psi(s)\mid\nu)\,\mathrm{d}s \leqslant \frac{1}{2\beta R}\mathcal{I}(\psi_0\mid\nu).$$

In the limit $t\to+\infty$, this inequality becomes

$$\mathcal{H}(\psi_0\mid\nu) \leqslant \frac{1}{2\beta R}\mathcal{I}(\psi_0\mid\nu),$$

which gives the claimed LSI since ψ_0 is arbitrary. □

Theorem 2.11 (Holley–Stroock). If ν satisfies LSI(R) and if $\widetilde{V} : \mathcal{D} \to \mathbb{R}$ is a bounded function, then $\widetilde{\nu} = \widetilde{Z}^{-1} \mathrm{e}^{-\widetilde{V}} \nu$ satisfies LSI($\widetilde{R}$) with $\widetilde{R} = R \, \mathrm{e}^{\inf \widetilde{V} - \sup \widetilde{V}}$.

Proof. Introduce $\phi(x) = x \ln x - 1 + x$. Consider any function $f \geqslant 0$ such that

$$\int_{\mathcal{D}} f \, \mathrm{d}\widetilde{\nu} = 1, \tag{2.32}$$

and let $\overline{f}$ denote the average of f with respect to ν:

$$\overline{f} = \int_{\mathcal{D}} f \, \mathrm{d}\nu. \tag{2.33}$$

Then

$$\begin{aligned} \mathcal{H}(f\widetilde{\nu} \mid \widetilde{\nu}) = \int_{\mathcal{D}} f \ln f \, \mathrm{d}\widetilde{\nu} &= \int_{\mathcal{D}} \phi(f) \, \mathrm{d}\widetilde{\nu} \\ &\leqslant \int_{\mathcal{D}} \left[\phi(f) - (\phi(\overline{f}) + \phi'(\overline{f})(1 - \overline{f})) \right] \mathrm{d}\widetilde{\nu} \\ &\leqslant \int_{\mathcal{D}} \left[\phi(f) - \phi(\overline{f}) + \phi'(\overline{f})(\overline{f} - f) \right] \mathrm{d}\widetilde{\nu}, \end{aligned}$$

since the convexity of ϕ implies that $\phi(1) = 0 \geqslant \phi(\overline{f}) + \phi'(\overline{f})(1 - \overline{f})$, and where we have used (2.32) to obtain the last line. Note that $\phi(f(q)) - \phi(\overline{f}) + \phi'(\overline{f})(\overline{f} - f(q)) \geqslant 0$, still by the convexity of ϕ. Since the integrand is non-negative, we can therefore reintroduce ν as follows:

$$\begin{aligned} \mathcal{H}(f\widetilde{\nu} \mid \widetilde{\nu}) &\leqslant \frac{1}{\widetilde{Z}} \mathrm{e}^{-\inf \widetilde{V}} \int_{\mathcal{D}} \left[\phi(f) - \phi(\overline{f}) + \phi'(\overline{f})(\overline{f} - f) \right] \mathrm{d}\nu \\ &= \frac{1}{\widetilde{Z}} \mathrm{e}^{-\inf \widetilde{V}} \left(\int_{\mathcal{D}} \phi(f) \, \mathrm{d}\nu - \phi(\overline{f}) \right), \end{aligned}$$

in view of (2.33). On the other hand,

$$\int_{\mathcal{D}} \phi(f) \, \mathrm{d}\nu - \phi(\overline{f}) = \int_{\mathcal{D}} f \ln f \, \mathrm{d}\nu - \overline{f} \ln \overline{f} = \overline{f} \int_{\mathcal{D}} \frac{f}{\overline{f}} \ln \frac{f}{\overline{f}} \, \mathrm{d}\nu = \overline{f} \int_{\mathcal{D}} \phi\left(\frac{f}{\overline{f}} \right) \mathrm{d}\nu.$$

Since ν satisfies LSI(R), and $f/\overline{f}$ has integral 1 with respect to ν,

$$\int_{\mathcal{D}} \phi\left(\frac{f}{\overline{f}} \right) \mathrm{d}\nu \leqslant \frac{1}{2R} \int_{\mathcal{D}} \left| \nabla \ln \left(\frac{f}{\overline{f}} \right) \right|^2 \left(\frac{f}{\overline{f}} \right) \mathrm{d}\nu = \frac{1}{2R\overline{f}} \int_{\mathcal{D}} |\nabla (\ln f)|^2 f \, \mathrm{d}\nu.$$

In conclusion,

$$\begin{aligned} \mathcal{H}(f\widetilde{\nu} \mid \widetilde{\nu}) &\leqslant \frac{1}{2R\widetilde{Z}} \mathrm{e}^{-\inf \widetilde{V}} \int_{\mathcal{D}} |\nabla (\ln f)|^2 f \, \mathrm{d}\nu \\ &\leqslant \frac{1}{2R} \mathrm{e}^{\sup \widetilde{V} - \inf \widetilde{V}} \int_{\mathcal{D}} |\nabla (\ln f)|^2 f \, \mathrm{d}\widetilde{\nu}, \end{aligned}$$

which gives the claimed statement. □

The previous result gives lower bounds on the LSI constants. It suggests that the LSI constant, and hence the rate of convergence to equilibrium of the overdamped Langevin dynamics, decreases exponentially as the temperature decreases and/or energetic barriers increase. Consider for instance the low-temperature regime $\beta \to +\infty$, for a reference convex potential $V_0(q) = a|q|^2/2$ with $a > 0$, perturbed by a bounded potential $\widetilde{V}$, for instance $\widetilde{V}(q) = b\exp(-|q|^2)$ with $b > 0$. In this case, Theorem 2.10 shows that the canonical measure associated with V, at inverse temperature β, satisfies LSI(βa), while the canonical measure associated with $V+\widetilde{V}$, at inverse temperature β, satisfies LSI($\beta a\,\mathrm{e}^{-\beta b}$) by Theorem 2.11. The LSI constants obtained by the Holley–Stroock perturbative argument are of course lower bounds, but it is indeed expected that the optimal constant decreases exponentially fast as the temperature decreases when the potential exhibits energetic barriers. This can be rigorously proved in one-dimensional systems for example; see Section 2.4 in Menz and Schlichting (2014).

2.2.3. Relationship between logarithmic Sobolev inequalities and Poincaré inequalities

Initial conditions ψ_0 belonging to $L^2(\nu)$ also have a finite relative entropy with respect to ν. Indeed, using the inequality

$$x\ln x - x + 1 \leqslant (x-1)^2, \quad \text{for all } x \geqslant 0,$$

it holds that

$$\mathcal{H}(\psi_0 \,|\, \nu) \leqslant \left\| \frac{\psi_0}{\nu} - \mathbf{1} \right\|^2_{L^2(\nu)}.$$

This shows that LSIs allow us to cover a larger set of initial conditions than Poincaré inequalities. Unsurprisingly, the following result shows that LSIs are stronger than Poincaré inequalities.

Proposition 2.12 (LSI implies Poincaré inequality). If a measure ν satisfies an LSI with constant $R > 0$, then it satisfies a Poincaré inequality with the same constant.

Proof. The idea of the proof is that the Poincaré inequality can be seen as a linearization of the logarithmic Sobolev inequality. Start from the LSI

$$\int_{\mathcal{D}} \phi \ln \phi \,\mathrm{d}\nu \leqslant \frac{1}{2R} \int_{\mathcal{D}} \frac{|\nabla \phi|^2}{\phi} \,\mathrm{d}\nu,$$

where $\phi \geqslant 0$ has average 1 with respect to ν. The function ϕ can be chosen

of the form $\phi = 1 + \varepsilon\varphi$ where $\varphi \in L^\infty(\mathcal{D})$ has average 0 with respect to ν. For $0 < \varepsilon \leqslant 1/(2\|\varphi\|_{L^\infty})$,

$$\int_{\mathcal{D}} (1+\varepsilon\varphi)\ln(1+\varepsilon\varphi)\,\mathrm{d}\nu \leqslant \frac{\varepsilon^2}{2R}\int_{\mathcal{D}} \frac{|\nabla\varphi|^2}{1+\varepsilon\varphi}\,\mathrm{d}\nu = \frac{\varepsilon^2}{2R}\left(\int_{\mathcal{D}} |\nabla\varphi|^2\,\mathrm{d}\nu + \mathrm{O}(\varepsilon)\right).$$

Since $(1+\varepsilon\varphi)\ln(1+\varepsilon\varphi) = \varepsilon\varphi + (\varepsilon\varphi)^2 + \mathrm{O}(\varepsilon^3)$, and using the fact that φ has average 0 with respect to ν,

$$\int_{\mathcal{D}} (1+\varepsilon\varphi)\ln(1+\varepsilon\varphi)\,\mathrm{d}\nu = \varepsilon^2\left(\int_{\mathcal{D}} \varphi^2\,\mathrm{d}\nu + \mathrm{O}(\varepsilon)\right).$$

The conclusion follows for bounded functions φ by taking the limit $\varepsilon \to 0$. The general case can finally be deduced by a density argument. □

2.3. Hypocoercivity

The results from the previous section can be used only for dynamics of overdamped Langevin type, the generators of which are elliptic operators. In this section we provide some methods for the study of convergence to equilibrium of Langevin dynamics (1.9), for which the noise is degenerate since it acts only on the momenta. The generator of Langevin dynamics reads

$$\mathcal{L} = p^T M^{-1}\nabla_q - \nabla V^T \nabla_p + \gamma\left(-p^T M^{-1}\nabla_p + \frac{1}{\beta}\Delta_p\right), \tag{2.34}$$

while its adjoint considered on $L^2(\mathcal{E})$ is

$$\mathcal{L}^\dagger = -p^T M^{-1}\nabla_q + \nabla V^T \nabla_p + \gamma\,\mathrm{div}_p\left(M^{-1}p\cdot + \frac{1}{\beta}\nabla_p\cdot\right).$$

Recall that (2.6) shows that the canonical measure

$$\mu(\mathrm{d}q\,\mathrm{d}p) = Z_\mu^{-1}\exp(-\beta H(q,p))\,\mathrm{d}q\,\mathrm{d}p, \tag{2.35}$$

where H is defined in (1.1), is invariant.

2.3.1. Hypoellipticity

Although the generator of the Langevin dynamics and its adjoint are not elliptic, they can be shown to be hypoelliptic, following the work of Hörmander (1967, 1985). To this end we introduce the commutator of two operators A, B:

$$[A,B] = AB - BA.$$

We can then state Hörmander's theorem.

Theorem 2.13. Consider C^∞ vector fields on the D-dimensional space $\mathcal{Y}$,

$$A_j = \sum_{i=1}^{D} A_{j,i}(y)\partial_{y_i},$$

and introduce the operator

$$\mathcal{A} = A_0 + \sum_{j=1}^{J} A_j^\dagger A_j,$$

where $A_j^\dagger$ is the (formal) adjoint of A_j on $L^2(\mathcal{Y})$. Assume that the Lie algebra spanned by

$$\{A_j\}_{j=0,\dots,J}, \quad \{[A_j, A_k]\}_{j,k=0,\dots,J}, \quad \{[[A_j, A_k], A_l]\}_{j,k,l=0,\dots,J}, \quad \dots$$

has maximal rank D at every point $y \in \mathcal{Y}$. Then $\mathcal{A}$ is hypoelliptic, namely there exists $\varepsilon > 0$ such that $\mathcal{A}f \in H^s_{\mathrm{loc}}$ implies $f \in H^{s+\varepsilon}_{\mathrm{loc}}$.

In particular, solutions f of the equation $\mathcal{A}f = 0$ are C^∞.

In order to use this result for Langevin dynamics, we consider $\mathcal{Y} = \mathcal{E}$, $D = 2d$, $J = d$, and rewrite the generator $\mathcal{L}$ as

$$\mathcal{L} = A_0 + \sum_{i=1}^{d} A_i^\dagger A_i,$$

with

$$A_0 = \mathcal{L}_{\mathrm{ham}} - \gamma p^T M^{-1} \nabla_p, \quad A_i = \sqrt{\frac{\gamma}{\beta}}\partial_{p_i} \quad \text{for } 1 \leqslant i \leqslant d,$$

where

$$\mathcal{L}_{\mathrm{ham}} = p^T M^{-1} \nabla_q - \nabla V^T \nabla_p$$

is the generator of the Hamiltonian part of the dynamics. Assuming for simplicity that M is diagonal with entries m_i (although the computations can be extended to account for general positive definite mass matrices), we obtain

$$[A_i, A_0]\varphi = \sqrt{\frac{\gamma}{\beta}} \frac{1}{m_i} (\partial_{q_i} - \gamma \partial_{p_i})\varphi \quad \text{for all } i \in \{1, \dots, d\}.$$

Therefore, the Lie algebra generated by $\{A_i\}_{i=1,\dots,d}$ and $\{[A_i, A_0]\}_{i=1,\dots,d}$ has maximal rank $2d$ at every configuration $(q, p) \in \mathcal{E}$. This proves that $\mathcal{L}$ is hypoelliptic. Similar computations show that $\mathcal{L}^\dagger$ is also hypoelliptic.

Note also that Theorem 2.13 can be used to prove that $\partial_t - \mathcal{L}$ and $\partial_t - \mathcal{L}^\dagger$ are hypoelliptic on $\mathcal{Y} = \mathbb{R} \times \mathcal{E}$. This ensures that the transition kernel associated with the Langevin dynamics admits a C^∞ density (see Section 2.4).

2.3.2. Lack of coercivity on $L^2_0(\mu)$

We recall that we let $\mathcal{A}^*$ denote the adjoint of a closed operator $\mathcal{A}$ with respect to the scalar product on $L^2(\mu)$ (see (2.7)). It turns out to be convenient to rewrite the generator (2.34) as

$$\mathcal{L} = \mathcal{L}_{\rm ham} - \frac{\gamma}{\beta} \sum_{i=1}^{d} \partial_{p_i}^* \partial_{p_i},$$

where $\partial_{p_i}^* = \beta (M^{-1} p)_i - \partial_{p_i}$. This decomposition highlights an important structural property of Langevin generators: they are the sum of a Hamiltonian part, and a symmetric dissipation operator which is degenerate, however, since it acts only on the momenta. A simple computation shows that the Hamiltonian part is antisymmetric on $L^2(\mu)$, so that the adjoint of $\mathcal{L}$ on $L^2(\mu)$ reads

$$\mathcal{L}^* = -\mathcal{L}_{\rm ham} - \frac{\gamma}{\beta} \sum_{i=1}^{d} \partial_{p_i}^* \partial_{p_i}.$$

The Fokker–Planck equation (2.1)–(2.3) can be reformulated in terms of this operator, upon writing $\psi(t) = f(t)\,\mu$, as

$$\partial_t f = \mathcal{L}^* f, \tag{2.36}$$

where the initial condition $f(0) = f_0 = \psi_0/\mu$ satisfies

$$f_0 \geqslant 0, \quad \int_{\mathcal{E}} f_0 \,\mathrm{d}\mu = 1.$$

Note that there is a slight inconsistency in this rewriting: the initial condition f_0 should be considered in $L^1(\mu)$, whereas the functional framework considered here requires the stronger integrability condition $f_0 \in L^2(\mu)$. This situation can be improved by using relative entropies: see Proposition 2.22 below.

In any case, $\mathrm{e}^{t\mathcal{L}^*} = (\mathrm{e}^{t\mathcal{L}})^*$, so decay estimates obtained for the bounded operator $\mathrm{e}^{t\mathcal{L}}$ immediately transfer to its adjoint. For simplicity of notation, the convergence results in Theorems 2.15, 2.18 and Proposition 2.20 are stated for the semigroup $\mathrm{e}^{t\mathcal{L}}$, but the reader should bear in mind that the results equally apply to the semigroup $\mathrm{e}^{t\mathcal{L}^*}$, which then yields results on the long-time behaviour of the Fokker–Planck equation by (2.36).

Note also that solutions to the Fokker–Planck equation (2.36) are expected to converge to the constant function $\mathbf{1}$. Upon subtracting this constant function from the initial condition f_0, the convergence of the law amounts to the convergence to 0 of $\mathrm{e}^{t\mathcal{L}^*}(f_0 - \mathbf{1})$. This motivates the introduction of the following functional space:

$$L^2_0(\mu) = \left\{ \varphi \in L^2(\mu) \,\middle|\, \int_{\mathcal{E}} \varphi \,\mathrm{d}\mu = 0 \right\}.$$

The Langevin generator $\mathcal{L}$ defined in (2.34), however, fails to be coercive on $L^2_0(\mu)$ since second derivatives in q are missing. In fact, for C^∞ and compactly supported test functions φ,

$$-\langle \mathcal{L}\varphi, \varphi\rangle_{L^2(\mu)} = \frac{1}{\beta}\|\nabla_p \varphi\|^2_{L^2(\mu)}, \tag{2.37}$$

which should be compared to (2.19) for overdamped Langevin dynamics. The key idea of hypocoercivity is to introduce some mixed derivatives in q and p in a modified scalar product in order to retrieve some dissipation in q through some commutator identities. This motivates the name for the technique, in view of the analogy with hypoellipticity, which allows us, via certain commutator identities, to state results for non-elliptic operators that are similar to regularity results obtained for elliptic operators.

More precisely, coercivity is obtained in the Hilbert space $H^1(\mu) \cap L^2_0(\mu)$, where

$$H^1(\mu) = \left\{\varphi \in L^2(\mu) \mid \nabla_p \varphi, \nabla_q \varphi \in (L^2(\mu))^d\right\}$$

is endowed with a scalar product different from (but equivalent to) the canonical one:

$$\langle \varphi_1, \varphi_2\rangle_{H^1(\mu)} = \langle \varphi_1, \varphi_2\rangle_{L^2(\mu)} + \langle \nabla_q \varphi_1, \nabla_q \varphi_2\rangle_{L^2(\mu)} + \langle \nabla_p \varphi_1, \nabla_p \varphi_2\rangle_{L^2(\mu)}.$$

It is then possible to resort to some Gronwall estimates and deduce the longtime convergence of $f(t)$ when a Poincaré inequality holds for μ (as in the proof of Proposition 2.3).

The idea of using mixed derivatives was already present in the computations performed in Talay (2002, Section 3), and was later generalized in Villani (2009). A careful application of the general hypocoercivity framework to Langevin dynamics can be read in Hairer and Pavliotis (2008), where some emphasis is laid on the limiting regime $\gamma \to 0$; see also Leimkuhler, Matthews and Stoltz (2015) for the regime $\gamma \to +\infty$.

2.3.3. Estimates in $H^1(\mu)$

Although hypocoercivity eventually provides decay estimates in $H^1(\mu)$, it turns out to be convenient, for the proof, to work with a specific scalar product equivalent to the canonical scalar product on $H^1(\mu)$. We introduce the following scalar product:

$$\langle\!\langle u, v\rangle\!\rangle = \langle u, v\rangle + a\langle \nabla_p u, \nabla_p v\rangle - b\langle \nabla_p u, \nabla_q v\rangle - b\langle \nabla_q u, \nabla_p v\rangle + c\langle \nabla_q u, \nabla_q v\rangle, \tag{2.38}$$

where, for simplicity of notation, we let $\langle \cdot, \cdot\rangle$ denote the standard scalar product on $L^2(\mu)$. In order for the above scalar product to be equivalent to the canonical scalar product on $H^1(\mu)$, we assume in the remainder of this section that

$$a, c > 0 \quad \text{and} \quad ac - b^2 > 0. \tag{2.39}$$

Lemma 2.14. Assume that (2.39) holds. Then the bilinear form $(u,v) \mapsto \langle\!\langle u, v \rangle\!\rangle$ induces a scalar product equivalent to the canonical scalar product on $H^1(\mu)$.

Proof. Of course,

$$|\langle\!\langle u, u \rangle\!\rangle| \leqslant \max(1, a + |b|, c + |b|) \|u\|^2_{H^1(\mu)}.$$

It therefore remains to prove that $\|u\|^2_{H^1(\mu)}$ can be controlled by $\langle\!\langle u, u \rangle\!\rangle$. Note first that, by a Cauchy–Schwarz inequality,

$$\begin{aligned} a\langle \nabla_p u, \nabla_p u \rangle - b\langle \nabla_p u, \nabla_q u \rangle - b\langle \nabla_q u, \nabla_p u \rangle + c\langle \nabla_q u, \nabla_q u \rangle \\ \geqslant \begin{pmatrix} \|\nabla_p u\|_{L^2(\mu)} \\ \|\nabla_q u\|_{L^2(\mu)} \end{pmatrix}^T \begin{pmatrix} a & -|b| \\ -|b| & c \end{pmatrix} \begin{pmatrix} \|\nabla_p u\|_{L^2(\mu)} \\ \|\nabla_q u\|_{L^2(\mu)} \end{pmatrix} \\ \geqslant \alpha \big(\|\nabla_p u\|^2_{L^2(\mu)} + \|\nabla_q u\|^2_{L^2(\mu)} \big), \end{aligned}$$

where

$$\alpha = \frac{1}{2}\big(a + c - \sqrt{(a-c)^2 + 4b^2}\big) = \frac{2(ac - b^2)}{a + c + \sqrt{(a-c)^2 + 4b^2}} > 0. \quad (2.40)$$

This shows that

$$\langle\!\langle u, u \rangle\!\rangle \geqslant \min(1, \alpha) \|u\|^2_{H^1(\mu)},$$

which allows us to conclude the proof. □

As mentioned above, for notational simplicity, we study the convergence to 0 of $\mathrm{e}^{t\mathcal{L}}\varphi$ for $\varphi \in L^2_0(\mu)$ rather than the convergence to 0 of $\mathrm{e}^{t\mathcal{L}^*}(f_0 - 1)$ for $f_0 \in L^2(\mu)$. However, the results of this section can be straightforwardly extended to the latter case by changing the sign of the antisymmetric part of the operator (which is handled by changing the sign of b in the definition of the scalar product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$). The following result can then be stated.

Theorem 2.15 (hypocoercivity). Fix $\gamma > 0$, and assume either that the domain $\mathcal{D}$ is bounded, or that there exists $\rho > 0$ such that

$$|\nabla^2 V(q)| \leqslant \rho(1 + |\nabla V(q)|), \quad \text{for all } q \in \mathcal{D}, \quad (2.41)$$

when $\mathcal{D}$ is not bounded. Then there exist $a, b, c \in \mathbb{R}$ satisfying (2.39) and $K > 0$ such that, for any $\varphi \in H^1(\mu) \cap L^2_0(\mu)$ and any $t \geqslant 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\big[\langle\!\langle \mathrm{e}^{t\mathcal{L}}\varphi, \mathrm{e}^{t\mathcal{L}}\varphi \rangle\!\rangle\big] \leqslant -K\big(\|\nabla_p \mathrm{e}^{t\mathcal{L}}\varphi\|^2_{L^2(\mu)} + \|\nabla_q \mathrm{e}^{t\mathcal{L}}\varphi\|^2_{L^2(\mu)}\big). \quad (2.42)$$

Further, if in addition a Poincaré inequality holds for the measure $\nu(\mathrm{d}q) = Z_\nu^{-1}\,\mathrm{e}^{-\beta V(q)}\,\mathrm{d}q$, then there exists $\kappa > 0$ such that, for any $t \geqslant 0$,

$$\langle\!\langle \mathrm{e}^{t\mathcal{L}}\varphi, \mathrm{e}^{t\mathcal{L}}\varphi \rangle\!\rangle \leqslant \mathrm{e}^{-2\kappa t}\langle\!\langle \varphi, \varphi \rangle\!\rangle, \quad \text{for all } \varphi \in H^1(\mu) \cap L^2_0(\mu). \quad (2.43)$$

As a consequence, there exists $C \geqslant 1$ such that

$$\|\mathrm{e}^{t\mathcal{L}}\|_{\mathcal{B}(H^1(\mu)\cap L^2_0(\mu))} \leqslant C\,\mathrm{e}^{-\kappa t}, \quad \text{for all } t \geqslant 0. \tag{2.44}$$

In the last inequality, $H^1(\mu)$ is endowed with the canonical scalar product. The operator bound (2.44) is obtained from (2.43) by taking the supremum over functions $\varphi \in H^1(\mu) \cap L^2_0(\mu)$ and using the equivalence of norms provided by Lemma 2.14. Note that this implies that $C \geqslant 1$.

An immediate consequence of the above convergence result is the following corollary (obtained from Proposition 2.1 with $E = H^1(\mu) \cap L^2_0(\mu)$).

Corollary 2.16. Under the same assumptions as in Theorem 2.15, the operator $\mathcal{L}$ is invertible on $H^1(\mu) \cap L^2_0(\mu)$, and the following equality holds in $\mathcal{B}(H^1(\mu) \cap L^2_0(\mu))$:

$$\mathcal{L}^{-1} = -\int_0^{+\infty} \mathrm{e}^{t\mathcal{L}}\,\mathrm{d}t.$$

Moreover,

$$\|\mathcal{L}^{-1}\|_{\mathcal{B}(H^1(\mu)\cap L^2_0(\mu))} \leqslant \frac{C}{\kappa}, \tag{2.45}$$

where C and κ are the same constants as in (2.44).

Let us now present the proof of Theorem 2.15.

Proof. Fix a C^∞ function $\varphi \in H^1(\mu) \cap L^2_0(\mu)$ (the general case being obtained by a density argument). Note that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\langle\!\langle \mathrm{e}^{t\mathcal{L}}\varphi,\, \mathrm{e}^{t\mathcal{L}}\varphi\rangle\!\rangle\right) = \langle\!\langle \mathrm{e}^{t\mathcal{L}}\varphi, \mathcal{L}\,\mathrm{e}^{t\mathcal{L}}\varphi\rangle\!\rangle.$$

Our aim is to find non-negative constants a, b, c such that (2.39) holds and, for all C^∞ functions φ with compact support,

$$\langle\!\langle \varphi, \mathcal{L}\varphi\rangle\!\rangle \leqslant -\kappa\langle\!\langle \varphi, \varphi\rangle\!\rangle.$$

This allows us to obtain (2.42) and thus the desired exponential decrease (2.43), using the Poincaré inequality and the Gronwall lemma. The idea of hypocoercivity is thus to find a scalar product equivalent to the canonical $H^1(\mu)$ scalar product and such that $-\mathcal{L}$ is coercive with respect to this scalar product (while it fails to be coercive with respect to the canonical $L^2(\mu)$ scalar product (see (2.37)), as well as with respect to the canonical $H^1(\mu)$ scalar product).

Let us first compute the various terms in $\langle\!\langle \varphi, \mathcal{L}\varphi\rangle\!\rangle$. It is useful to first establish some commutator identities:

$$[\partial_{p_i}, \mathcal{L}]\varphi = \frac{1}{m_i}(\partial_{q_i} - \gamma\partial_{p_i})\varphi,$$
$$[\partial_{q_i}, \mathcal{L}]\varphi = -\nabla(\partial_{q_i}V)^T\nabla_p\varphi.$$

Some straightforward computations then show that

$$\begin{aligned}\langle\langle \varphi, \mathcal{L}\varphi\rangle\rangle &= -\frac{\gamma}{\beta}\|\nabla_p\varphi\|^2_{L^2(\mu)} \\ &+ a\sum_{i=1}^d\left(-\frac{\gamma}{\beta}\|\nabla_p(\partial_{p_i}\varphi)\|^2_{L^2(\mu)} - \frac{\gamma}{m_i}\|\partial_{p_i}\varphi\|^2_{L^2(\mu)} + \frac{1}{m_i}\langle\partial_{p_i}\varphi, \partial_{q_i}\varphi\rangle\right) \\ &+ c\sum_{i=1}^d\left(-\frac{\gamma}{\beta}\|\nabla_p(\partial_{q_i}\varphi)\|^2_{L^2(\mu)} - \langle\partial_{q_i}\varphi, \nabla(\partial_{q_i}V)^T\nabla_p\varphi\rangle\right) \\ &- b\sum_{i=1}^d\left(-\frac{\gamma}{\beta}\langle\nabla_p(\partial_{p_i}\varphi), \nabla_p(\partial_{q_i}\varphi)\rangle - \langle\partial_{p_i}\varphi, \nabla(\partial_{q_i}V)^T\nabla_p\varphi\rangle\right) \\ &- b\sum_{i=1}^d\left(-\frac{\gamma}{\beta}\langle\nabla_p(\partial_{q_i}\varphi), \nabla_p(\partial_{p_i}\varphi)\rangle + \frac{1}{m_i}\|\partial_{q_i}\varphi\|^2_{L^2(\mu)} - \frac{\gamma}{m_i}\langle\partial_{q_i}\varphi, \partial_{p_i}\varphi\rangle\right).\end{aligned}$$

Note in particular that dissipation terms $\|\partial_{q_i}\varphi\|^2_{L^2(\mu)}$ appear in the last line. This motivates choosing the parameter b positive.[6] The next step is to bound the right-hand side of the previous equality from above using Cauchy–Schwarz inequalities, by (recall that we assume $b \geqslant 0$)

$$\begin{aligned}\langle\langle \varphi, \mathcal{L}\varphi\rangle\rangle \leqslant &-\gamma\left(\frac{1}{\beta} + \frac{a}{m_+}\right)\|\nabla_p\varphi\|^2_{L^2(\mu)} - \frac{b}{m_+}\|\nabla_q\varphi\|^2_{L^2(\mu)} \\ &+ \frac{a + b\gamma}{m_-}\|\nabla_p\varphi\|_{L^2(\mu)}\|\nabla_q\varphi\|_{L^2(\mu)} \\ &- \frac{a\gamma}{\beta}\sum_{i=1}^d\|\nabla_p(\partial_{p_i}\varphi)\|^2_{L^2(\mu)} - \frac{c\gamma}{\beta}\sum_{i=1}^d\|\nabla_p(\partial_{q_i}\varphi)\|^2_{L^2(\mu)} \\ &+ \frac{2b\gamma}{\beta}\sum_{i=1}^d\|\nabla_p(\partial_{p_i}\varphi)\|_{L^2(\mu)}\|\nabla_p(\partial_{q_i}\varphi)\|_{L^2(\mu)} \\ &+ \langle(b\nabla_p - c\nabla_q)\varphi, (\nabla^2 V)\nabla_p\varphi\rangle,\end{aligned}$$

where

$$m_+ = \max(m_1, \ldots, m_d), \quad m_- = \min(m_1, \ldots, m_d). \tag{2.46}$$

Condition (2.39) shows that (by a computation similar to that performed in the proof of Lemma 2.14)

$$\begin{aligned}&a\|\nabla_p(\partial_{p_i}\varphi)\|^2_{L^2(\mu)} + c\|\nabla_p(\partial_{q_i}\varphi)\|^2_{L^2(\mu)} - 2b\|\nabla_p(\partial_{p_i}\varphi)\|_{L^2(\mu)}\|\nabla_p(\partial_{q_i}\varphi)\|_{L^2(\mu)} \\ &\quad\geqslant \alpha(\|\nabla_p(\partial_{p_i}\varphi)\|^2_{L^2(\mu)} + \|\nabla_p(\partial_{q_i}\varphi)\|^2_{L^2(\mu)}),\end{aligned} \tag{2.47}$$

[6] When working with $\mathcal{L}^*$ instead of $\mathcal{L}$, the parameter b should be chosen negative.

with $\alpha > 0$ defined in (2.40). Therefore, the upper bound on $\langle\langle \varphi, \mathcal{L}\varphi \rangle\rangle$ simplifies as

$$\begin{aligned}\langle\langle \varphi, \mathcal{L}\varphi \rangle\rangle \leqslant & -\gamma\left(\frac{1}{\beta} + \frac{a}{m_+}\right)\|\nabla_p \varphi\|^2_{L^2(\mu)} - \frac{b}{m_+}\|\nabla_q \varphi\|^2_{L^2(\mu)} \\ & + \frac{a + b\gamma}{m_-}\|\nabla_p \varphi\|_{L^2(\mu)}\|\nabla_q \varphi\|_{L^2(\mu)} \\ & - \frac{\alpha\gamma}{\beta}\sum_{i=1}^{d}(\|\nabla_p(\partial_{p_i}\varphi)\|^2_{L^2(\mu)} + \|\nabla_p(\partial_{q_i}\varphi)\|^2_{L^2(\mu)}) \qquad (2.48) \\ & + \langle (b\nabla_p - c\nabla_q)\varphi, (\nabla^2 V)\nabla_p \varphi\rangle.\end{aligned}$$

In order to control the term $\langle (b\nabla_p - c\nabla_q)\varphi, (\nabla^2 V)\nabla_p\varphi\rangle$, two cases have to be distinguished.

(i) *The position space $\mathcal{D}$ is compact.* Here $\nabla^2 V$ is uniformly bounded, and there exists a constant $C_V > 0$ such that

$$\begin{aligned}|\langle (b\nabla_p - c\nabla_q)\varphi, (\nabla^2 V)\nabla_p\varphi\rangle| &\leqslant C_V\|\nabla_p\varphi\|_{L^2(\mu)}\|(b\nabla_p - c\nabla_q)\varphi\|_{L^2(\mu)} \\ &\leqslant bC_V\|\nabla_p\varphi\|^2_{L^2(\mu)} + cC_V\|\nabla_p\varphi\|_{L^2(\mu)}\|\nabla_q\varphi\|_{L^2(\mu)}.\end{aligned}$$

(ii) *The position space is not compact.* Conditions must be imposed on the potential energy function in order to control the growth of the Hessian at infinity. One possible condition is that $\nabla^2 V$ (considered as a multiplication operator acting on vectors) is relatively bounded by ∇_q on $L^2(\mu)$, that is, there exist $A_V, B_V \geqslant 0$ such that

$$\forall \phi \in H^1(\mu), \quad \|\phi\nabla^2 V\|_{L^2(\mu)} \leqslant A_V\|\phi\|_{L^2(\mu)} + B_V\|\nabla_q\phi\|_{L^2(\mu)}.$$

This condition is satisfied when (2.41) holds (see Lemma A.24 of Villani 2009). In this case,

$$\begin{aligned}&|\langle (b\nabla_p - c\nabla_q)\varphi, (\nabla^2 V)\nabla_p\varphi\rangle| \\ &\quad\leqslant (A_V\|\nabla_p\varphi\|_{L^2(\mu)} + B_V\|\nabla_p\nabla_q\varphi\|_{L^2(\mu)})\|(b\nabla_p - c\nabla_q)\varphi\|_{L^2(\mu)}.\end{aligned}$$

The term involving derivatives in both q and p can be controlled by the dissipative terms $-\frac{\alpha\gamma}{\beta}\|\partial_{p_j}\partial_{q_i}\varphi\|^2_{L^2(\mu)}$ in (2.48).

For simplicity, we consider the case when $\mathcal{D}$ is compact. Then,

$$\begin{aligned}\langle\langle \varphi, \mathcal{L}\varphi \rangle\rangle \leqslant & \left[bC_V - \gamma\left(\frac{1}{\beta} + \frac{a}{m_+}\right)\right]\|\nabla_p\varphi\|^2_{L^2(\mu)} - \frac{b}{m_+}\|\nabla_q\varphi\|^2_{L^2(\mu)} \\ & + \left(\frac{a + b\gamma}{m_-} + cC_V\right)\|\nabla_p\varphi\|_{L^2(\mu)}\|\nabla_q\varphi\|_{L^2(\mu)} \\ = & -X^T\mathcal{M}X,\end{aligned}$$

where

$$X = \begin{pmatrix} \|\nabla_p \varphi\|_{L^2(\mu)} \\ \|\nabla_q \varphi\|_{L^2(\mu)} \end{pmatrix}, \quad \mathcal{M} = \begin{pmatrix} A & C/2 \\ C/2 & B \end{pmatrix},$$

with

$$A = \gamma\left(\frac{1}{\beta} + \frac{a}{m_+}\right) - bC_V, \quad B = \frac{b}{m_+}, \quad C = \frac{a + b\gamma}{m_-} + cC_V.$$

In addition to (2.39), the values of a, b, c should be such that

$$4AB > C^2, \tag{2.49}$$

in order for $\mathcal{M}$ to be positive definite. If this is the case, (2.42) follows. The condition (2.49) is satisfied for instance when

$$a = c = \varepsilon, \quad b = \varepsilon^{1+\delta}, \quad \delta \in (0, 1), \tag{2.50}$$

for $\varepsilon > 0$ sufficiently small. Note also that this choice is compatible with (2.39) when $\varepsilon < 1$.

Let us now explain how to get (2.43) from (2.42), assuming that ν satisfies a Poincaré inequality. This requires us to retrieve some control on the $L^2(\mu)$-norm of φ from the norms of the gradient. We use to this end the fact that, by the tensorization argument stated in Proposition 2.6, the canonical measure μ satisfies a Poincaré inequality, whose constant we denote by R. Therefore, $X^T S X \geqslant \langle\!\langle \varphi, \varphi \rangle\!\rangle$ with

$$S = \begin{pmatrix} a + 1/(2R) & -b \\ -b & c + 1/(2R) \end{pmatrix}.$$

We finally define κ as the largest positive constant such that $X^T \mathcal{M} X \geqslant \kappa X^T S X$. In fact, κ is the smallest eigenvalue of $S^{-1/2} \mathcal{M} S^{-1/2}$ (or equivalently of $\mathcal{M} S^{-1}$). Since S is symmetric positive definite, this shows that $\kappa > 0$ since $\mathcal{M}$ is also positive definite thanks to (2.50). This concludes the proof of (2.43). □

One interest of the hypocoercive approach is that the constants κ and C in (2.44) can be made quite explicit in terms of the various factors (related to the potential such as the bound on the Hessian C_V and the Poincaré constant R, or to the masses). Let us illustrate this for various asymptotic cases.

(i) $\beta \to +\infty$. The constant R typically decreases exponentially with the temperature, with a lower bound scaling as $\mathrm{e}^{\beta(\inf \widetilde{V} - \sup \widetilde{V})}$ when $V = V_{\text{convex}} + \widetilde{V}$. In this decomposition, the potential is separated into a strongly convex part V_{convex} whose Hessian is uniformly lower bounded by a positive constant while $\widetilde{V}$ is some bounded perturbation. On the other hand, upon rescaling the values a, b, c in the definition of the

scalar product by a factor $1/\beta$ (which amounts to considering $a = \overline{a}/\beta$, *etc.*), it can be shown that the smallest eigenvalue α of $\mathcal{M}$ is of order $1/\beta$. Therefore, the smallest eigenvalue κ of $S^{-1/2}\mathcal{M}S^{-1/2}$ admits a lower bound which decreases exponentially with β.

(ii) *Hamiltonian limit* $\gamma \to 0$ *or overdamped limit* $\gamma \to +\infty$. The results stated in Theorem 2.15 are obtained for a given value of γ. On the other hand, the Langevin dynamics becomes singular in the limit $\gamma \to 0$, where it reduces to the Hamiltonian dynamics (which is not ergodic with respect to the canonical measure), and in the limit $\gamma \to +\infty$, where it converges to the overdamped Langevin dynamics (see the discussion after (1.9)). It is therefore expected that the convergence rate to equilibrium of the Langevin dynamics degrades as $\gamma \to 0$ or $\gamma \to +\infty$.

Hypocoercivity allows us to quantify the degradation of the convergence rate. Let κ_γ denote the exponential decay rate given by Theorem 2.15 for a given value of γ. Upon choosing $(a, b, c) = \gamma(\overline{a}, \overline{b}, \overline{c})$ when $\gamma \to 0$ and $(a, b, c) = (\overline{a}, \overline{b}, \overline{c})/\gamma$ when $\gamma \to +\infty$, a careful analysis of the scaling of κ_γ shows that there exists $\overline{\kappa} > 0$ such that

$$\kappa_\gamma \geqslant \min\left(\gamma, \frac{1}{\gamma}\right)\overline{\kappa}. \tag{2.51}$$

The decay rate therefore becomes singular both in the Hamiltonian and overdamped limits. In both cases, the decay is apparent only at long time scales, of order t/γ as $\gamma \to 0$ (the fluctuation/dissipation is so small that energy diffusion is only observed at long times: see Hairer and Pavliotis 2008 for a precise statement) and γt as $\gamma \to +\infty$ (the fluctuation/dissipation is so large that the momenta are continuously randomized, which leads to some effective Brownian motion on the positions over long times). Note however that, for the limits considered here, the constant C in (2.44) becomes singular. A more convenient formulation of the scaling limits $\gamma \to 0$ or $\gamma \to +\infty$ is provided by Proposition 2.20 below.

Remark 2.17 (degenerate scalar product). The standard hypocoercive approach relies on estimates in $H^1(\mu)$, obtained under the non-degeneracy condition (2.39). However, in the situation when $\nabla^2 V$ is bounded, it is possible with a slight modification of the above argument[7] to state an exponential convergence in the degenerate case $a = b = c$, for which the associated squared norm is

$$\|f\|^2_{L^2(\mu)} + a\|(\nabla_p - \nabla_q)f\|^2_{L^2(\mu)}.$$

[7] We thank Stefano Olla for pointing out this possibility.

2.3.4. Estimates in $L^2(\mu)$

The passage from bounds in $H^1(\mu)$ to bounds in $L^2(\mu)$ follows from hypoelliptic regularization results. Such results are presented in Villani (2009, Theorem A.8) or Hairer and Pavliotis (2008, Section 6.1). We follow the latter approach, which is more straightforward, although the so-obtained results are not as strong as the results presented in Villani (2009).

Theorem 2.18 (hypoelliptic regularization). Assume $\nabla^2 V \in L^\infty(\mathcal{D})$ or that (2.41) holds. Then there exists $K > 0$ such that, for any $\varphi \in L^2(\mu)$,

$$\|\nabla_p \mathrm{e}^{t\mathcal{L}}\varphi\|_{L^2(\mu)} + \|\nabla_q \mathrm{e}^{t\mathcal{L}}\varphi\|_{L^2(\mu)} \leqslant \frac{K}{t^{3/2}}\|\varphi\|_{L^2(\mu)}, \quad \text{for all } 0 < t \leqslant 1.$$

Combining this inequality with $t = 1$ and Theorem 2.15, we can conclude that, for $t \geqslant 1$ and $\varphi \in L^2_0(\mu)$,

$$\begin{aligned}\|\mathrm{e}^{t\mathcal{L}}\varphi\|^2_{L^2(\mu)} &\leqslant \langle\!\langle \mathrm{e}^{t\mathcal{L}}\varphi,\, \mathrm{e}^{t\mathcal{L}}\varphi\rangle\!\rangle \\ &\leqslant \mathrm{e}^{-2\kappa(t-1)}\langle\!\langle \mathrm{e}^{\mathcal{L}}\varphi,\, \mathrm{e}^{\mathcal{L}}\varphi\rangle\!\rangle \\ &\leqslant \widetilde{C}\,\mathrm{e}^{-2\kappa t}\|\varphi\|^2_{L^2(\mu)},\end{aligned}$$

which gives an exponential decay in $L^2(\mu)$. For completeness, let us recall the proof of Theorem 2.18, as presented in Hairer and Pavliotis (2008).

Proof. As at the end of the proof of Theorem 2.15, we consider for simplicity the case when $\nabla^2 V$ is bounded. We denote $\mathrm{e}^{t\mathcal{L}}\varphi$ by $\varphi(t)$ in this proof. Define

$$\begin{aligned}N_\varphi(t) = \frac{1}{2}\big[&\|\varphi(t)\|^2_{L^2(\mu)} + c_1 t\|\nabla_p\varphi(t)\|^2_{L^2(\mu)} \\ &- c_2 t^2\langle\nabla_q\varphi(t), \nabla_p\varphi(t)\rangle_{L^2(\mu)} + c_3 t^3\|\nabla_q\varphi(t)\|^2_{L^2(\mu)}\big],\end{aligned} \tag{2.52}$$

for some positive constants c_1, c_2, c_3 to be determined later on. Note that $2N_\varphi(t)$ corresponds to the norm induced by the scalar product (2.38) for time-dependent coefficients $a(t) = c_1 t$, $2b(t) = c_2 t^2$ and $c(t) = c_3 t^3$. This choice is motivated in Remark 2.19 below. The time derivative of this quantity reads

$$\begin{aligned}\frac{\mathrm{d}N_\varphi(t)}{\mathrm{d}t} &= \langle\varphi(t), \mathcal{L}\varphi(t)\rangle + \frac{c_1}{2}\|\nabla_p\varphi(t)\|^2_{L^2(\mu)} + c_1 t\sum_{i=1}^d \langle\partial_{p_i}\varphi(t), \partial_{p_i}\mathcal{L}\varphi(t)\rangle \\ &\quad - c_2 t\langle\nabla_q\varphi(t), \nabla_p\varphi(t)\rangle \\ &\quad - \frac{c_2 t^2}{2}\sum_{i=1}^d[\langle\partial_{q_i}\varphi(t), \partial_{p_i}\mathcal{L}\varphi(t)\rangle + \langle\partial_{p_i}\varphi(t), \partial_{q_i}\mathcal{L}\varphi(t)\rangle] \\ &\quad + \frac{3c_3 t^2}{2}\|\nabla_q\varphi(t)\|^2_{L^2(\mu)} + c_3 t^3\sum_{i=1}^d\langle\partial_{q_i}\varphi(t), \partial_{q_i}\mathcal{L}\varphi(t)\rangle.\end{aligned}$$

Using computations similar to those in the proof of Theorem 2.15, we obtain

$$\begin{aligned}\frac{\mathrm{d}N_\varphi(t)}{\mathrm{d}t} &= -\left(\frac{\gamma}{\beta}-\frac{c_1}{2}\right)\|\nabla_p\varphi(t)\|^2_{L^2(\mu)} \\ &\quad - c_1 t\sum_{i=1}^d\left[\frac{\gamma}{\beta}\|\nabla_p(\partial_{p_i}\varphi(t))\|^2_{L^2(\mu)}+\frac{\gamma}{m_i}\|\partial_{p_i}\varphi(t)\|^2_{L^2(\mu)}\right] \\ &\quad + \sum_{i=1}^d\left[t\left(\frac{c_1}{m_i}-c_2\right)+\frac{\gamma c_2 t^2}{2m_i}\right]\langle\partial_{p_i}\varphi(t),\partial_{q_i}\varphi(t)\rangle \\ &\quad + \frac{c_2t^2}{2}\sum_{i=1}^d\frac{2\gamma}{\beta}\langle\nabla_p(\partial_{p_i}\varphi(t)),\nabla_p(\partial_{q_i}\varphi(t))\rangle+\langle\partial_{p_i}\varphi(t),\nabla(\partial_{q_i}V)^T\nabla_p\varphi(t)\rangle \\ &\quad + t^2\sum_{i=1}^d\left(\frac{3c_3}{2}-\frac{c_2}{m_i}\right)\|\partial_{q_i}\varphi(t)\|^2_{L^2(\mu)} \\ &\quad - c_3t^3\sum_{i=1}^d\left[\frac{\gamma}{\beta}\|\nabla_p(\partial_{q_i}\varphi(t))\|^2_{L^2(\mu)}+\langle\partial_{q_i}\varphi(t),\nabla(\partial_{q_i}V)^T\nabla_p\varphi(t)\rangle\right].\end{aligned} \tag{2.53}$$

We now choose c_1, c_2, c_3 such that $\mathrm{d}N_\varphi/\mathrm{d}t \leqslant 0$. First, we restrict ourselves to coefficients such that

$$c_2t^2xy \leqslant c_1tx^2 + c_3t^3y^2, \quad \text{for all } (x,y)\in\mathbb{R}^2,$$

which is satisfied once

$$c_1, c_3 \geqslant 0, \quad 4c_1c_3 \geqslant c_2^2. \tag{2.54}$$

Then,

$$\begin{aligned}-c_1t\sum_{i=1}^d\frac{\gamma}{\beta}\|\nabla_p(\partial_{p_i}\varphi(t))\|^2_{L^2(\mu)}+\frac{c_2t^2}{2}\sum_{i=1}^d\frac{2\gamma}{\beta}\langle\nabla_p(\partial_{p_i}\varphi(t)),\nabla_p(\partial_{q_i}\varphi(t))\rangle \\ -c_3t^3\sum_{i=1}^d\frac{\gamma}{\beta}\|\nabla_p(\partial_{q_i}\varphi(t))\|^2_{L^2(\mu)}\leqslant 0,\end{aligned}$$

so that, with $C_V = \|\nabla^2 V\|_{L^\infty}$,

$$\begin{aligned}\frac{\mathrm{d}N_\varphi(t)}{\mathrm{d}t} &\leqslant -\left(\frac{\gamma}{\beta}+\frac{\gamma c_1 t}{m_+}-\frac{c_1}{2}-C_V\frac{c_2t^2}{2}\right)\|\nabla_p\varphi(t)\|^2_{L^2(\mu)} \\ &\quad - t^2\left(\frac{c_2}{m_+}-\frac{3c_3}{2}\right)\|\nabla_q\varphi(t)\|^2_{L^2(\mu)} \\ &\quad + \left[t\left(\frac{c_1}{m_-}-c_2\right)+\frac{\gamma c_2t^2}{2m_-}+C_Vc_3t^3\right]\|\nabla_p\varphi(t)\|_{L^2(\mu)}\|\nabla_q\varphi(t)\|_{L^2(\mu)}.\end{aligned}$$

On the time interval $[0,1]$, we finally obtain the following upper bound:

$$\begin{aligned}\frac{\mathrm{d}N_\varphi(t)}{\mathrm{d}t} \leqslant & -\left(\frac{\gamma}{\beta}-\frac{c_1}{2}-C_V\frac{c_2}{2}\right)\|\nabla_p\varphi(t)\|^2_{L^2(\mu)} \\ & -t^2\left(\frac{c_2}{m_+}-\frac{3c_3}{2}\right)\|\nabla_q\varphi(t)\|^2_{L^2(\mu)} \\ & +t\left[\frac{c_1}{m_-}+\frac{\gamma c_2}{2m_-}+C_V c_3\right]\|\nabla_p\varphi(t)\|_{L^2(\mu)}\|\nabla_q\varphi(t)\|_{L^2(\mu)}.\end{aligned}$$

We next consider coefficients $c_3 \ll c_2 \ll c_1 \ll 1$ satisfying (2.54) and $c_1^2 \ll c_2$. For instance, $c_1=\varepsilon$, $c_2=\varepsilon^{1+\delta}$ and $c_3=\varepsilon^{1+\delta+\eta}$, with $0<\eta<\delta<1$ and for $\varepsilon>0$ sufficiently small. We can then deduce that $\mathrm{d}N_\varphi(t)/\mathrm{d}t \leqslant 0$, which implies $N_\varphi(t) \leqslant N_\varphi(0) = \|\varphi\|^2_{L^2(\mu)}/2$. The desired conclusion immediately follows. □

Remark 2.19. Let us now motivate definition (2.52) more precisely. As discussed after this equation, the choice (2.52) corresponds to the scalar product (2.38) with time-dependent coefficients $a(t), b(t), c(t)$. First, in order to only have the $L^2(\mu)$-norm of φ at time $t=0$ in $N_\varphi(0)$, it is necessary that $a(0)=b(0)=c(0)=0$. Second, when generalizing the computations leading to (2.53), the prefactors of the third and fifth terms on the right-hand side of (2.53) now read $a(t)/m_i - b'(t) + \gamma b(t)/m_i$ instead of $t(c_1/m_i - c_2) + \gamma c_2 t^2/(2m_i)$ and $c'(t)/2 - 2b(t)/m_i$ instead of $t^2(3c_3/2 - c_2/m_i)$. This suggests that $a(t)$ and $b'(t)$ should be of the same order of magnitude, as well as $b(t)$ and $c'(t)$. When $a(t)$ is linear in t, this implies that $b(t)$ is quadratic and $c(t)$ cubic.

By tracking the dependency of the constant K on γ, Theorem 2.18 combined with Theorem 2.15 and the estimate (2.51) leads to the following result.

Proposition 2.20. Assume that $\nabla^2 V \in L^\infty(\mathcal{D})$ and that a Poincaré inequality holds for ν. Then there exist $C, \overline{\kappa} > 0$ such that, for all $\gamma > 0$, for all $\varphi \in L^2_0(\mu)$,

$$\|\mathrm{e}^{t\mathcal{L}}\varphi\|_{L^2(\mu)} \leqslant C\,\mathrm{e}^{-\overline{\kappa}\min(\gamma,\gamma^{-1})t}\|\varphi\|_{L^2(\mu)}. \tag{2.55}$$

This result again emphasizes the fact that, in the limiting regimes of very low or very large frictions, the system should be considered over long times $\max(\gamma, 1/\gamma)t$ in order to display a non-trivial behaviour.

Similarly to Corollary 2.16, the following result can be stated.

Corollary 2.21. Under the same assumptions as in Proposition 2.20, the operator $\mathcal{L}$ is invertible on $L^2_0(\mu)$, and the equality

$$\mathcal{L}^{-1} = -\int_0^{+\infty} \mathrm{e}^{t\mathcal{L}}\,\mathrm{d}t$$

holds in $\mathcal{B}(L^2_0(\mu))$. Moreover,

$$\|\mathcal{L}^{-1}\|_{\mathcal{B}(L^2_0(\mu))} \leqslant \frac{C}{\overline{\kappa}\min(\gamma,\gamma^{-1})}, \tag{2.56}$$

where C and $\overline{\kappa}$ are the same constants as in (2.55).

2.3.5. Hypocoercivity in the entropic sense

Here we show how the computations performed in $H^1(\mu)$ can be modified to prove the decay of some logarithmic entropy, similar to the one used for LSIs. The interest of this approach is that initial conditions in larger functional spaces can be considered. The general strategy is presented in Villani (2009, Section 6). These computations can be simplified for Langevin-type dynamics: see Olla and Letizia (2015, Appendix D).

We state here convergence results for the function $f(t)$, which is the Radon–Nikodym derivative of the law $\psi(t)$ of the process at time t with respect to the density of the invariant measure μ. Starting from $f_0 \geqslant 0$ with $\int_{\mathcal{E}} f_0 \,\mathrm{d}\mu = 1$, the function $f(t)$ therefore evolves according to the Fokker–Planck equation (2.36), that is, $f(t) = \mathrm{e}^{t\mathcal{L}^*} f_0$. It is expected that $f(t)$ converges to the constant function $\mathbf{1}$.

To quantify the convergence rate, we introduce the entropy functional

$$\mathscr{E}(f) = \int_{\mathcal{E}} f \,\ln f \,\mathrm{d}\mu + \int_{\mathcal{E}} \frac{\nabla f^T S \nabla f}{f}\,\mathrm{d}\mu,$$

where $S \in \mathbb{R}^{2d\times 2d}$ is a non-negative symmetric matrix. Note that the functional $\mathscr{E}$ combines the relative entropy and a generalization of the Fisher information introduced in Definition 2.7. It is sufficient for our purposes to restrict ourselves to the case when S is constant (but see Villani 2009, Remark 29 for a context where a dependence on the position q may be useful). Finally, note that $\mathscr{E}(\mathbf{1}) = 0$, so that the entropy is expected to converge to 0 as $t \to +\infty$.

Proposition 2.22. Assume $\nabla^2 V \in L^\infty(\mathcal{D})$. Then there exist a non-negative symmetric matrix $S \in \mathbb{R}^{2d\times 2d}$ and $\alpha > 0$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathscr{E}(f(t))] \leqslant -\alpha \int_{\mathcal{E}} \frac{|\nabla f(t)|^2}{f(t)}\,\mathrm{d}\mu, \quad \text{for all } t \geqslant 0. \tag{2.57}$$

If in addition $\nu(\mathrm{d}q)$ satisfies a logarithmic Sobolev inequality, then there exists $\kappa > 0$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathscr{E}(f(t))] \leqslant -\kappa\,\mathscr{E}(f(t)), \quad \text{for all } t \geqslant 0. \tag{2.58}$$

In particular, $0 \leqslant \mathscr{E}(f(t)) \leqslant \mathscr{E}(f_0)\,\mathrm{e}^{-\kappa t}$ for any $t \geqslant 0$.

Notice that we do not require the matrix S to be definite for the result to hold.

Proof. One of the main ideas in the proof is to rewrite the time derivative of the second term in the expression of $\mathscr{E}$ as a sum of terms similar to the ones appearing in the proof of Theorem 2.15. To this end, consider $g(t) = \sqrt{f(t)}$, so that

$$\int_{\mathcal{E}} \frac{\nabla f(t)^T S \nabla f(t)}{f(t)} \, \mathrm{d}\mu = 4 \int_{\mathcal{E}} \nabla g(t)^T S \nabla g(t) \, \mathrm{d}\mu. \tag{2.59}$$

In order to determine the time evolution of this quantity, we first write the time evolution for $g(t)$. Simple computations show that

$$\begin{aligned} \partial_t g(t) &= \frac{\partial_t f(t)}{2\sqrt{f(t)}} = \frac{\mathcal{L}^* f(t)}{2g(t)}, \\ \Delta_p g(t) &= \frac{\Delta_p f(t)}{2\sqrt{f(t)}} - \frac{|\nabla_p f(t)|^2}{4f(t)^{3/2}} = \frac{\Delta_p f(t)}{2g(t)} - \frac{|\nabla_p g(t)|^2}{g(t)}, \end{aligned}$$

so that

$$\partial_t g(t) = \mathcal{L}^* g(t) + \frac{\gamma}{\beta} \frac{|\nabla_p g(t)|^2}{g(t)}.$$

We next compute the time derivatives of the various terms in $\mathscr{E}(f(t))$, relying on the reformulation (2.59). This leads to expressions very similar to those encountered in the proof of Theorem 2.15. First,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{\mathcal{E}} f(t) \, \ln f(t) \, \mathrm{d}\mu \right) &= \int_{\mathcal{E}} (1 + \ln f(t)) \mathcal{L}^* f(t) \, \mathrm{d}\mu \\ &= \int_{\mathcal{E}} \mathcal{L}_{\mathrm{ham}} (1 + \ln f(t)) f(t) \, \mathrm{d}\mu - \frac{\gamma}{\beta} \int_{\mathcal{E}} \nabla_p f(t) \cdot \nabla_p (\ln f(t)) \, \mathrm{d}\mu \\ &= \int_{\mathcal{E}} \mathcal{L}_{\mathrm{ham}} f(t) \, \mathrm{d}\mu - \frac{\gamma}{\beta} \int_{\mathcal{E}} \frac{|\nabla_p f(t)|^2}{f(t)} \, \mathrm{d}\mu = -\frac{4\gamma}{\beta} \int_{\mathcal{E}} |\nabla_p g(t)|^2 \, \mathrm{d}\mu, \end{aligned}$$

in view of the invariance of μ by $\mathcal{L}_{\mathrm{ham}}$. Next,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{\mathcal{E}} |\partial_{p_i} g(t)|^2 \, \mathrm{d}\mu \right) &= 2 \int_{\mathcal{E}} \partial_{p_i} g(t) \, \partial_{p_i} (\mathcal{L}^* g(t)) \, \mathrm{d}\mu \\ &\quad + \frac{2\gamma}{\beta} \int_{\mathcal{E}} \partial_{p_i} g(t) \, \partial_{p_i} \left(\frac{|\nabla_p g(t)|^2}{g(t)} \right) \mathrm{d}\mu \\ &= 2 \int_{\mathcal{E}} \partial_{p_i} g(t) \, \mathcal{L}^* (\partial_{p_i} g(t)) \, \mathrm{d}\mu - \frac{2}{m_i} \int_{\mathcal{E}} \partial_{p_i} g(t) \, (\partial_{q_i} + \gamma \partial_{p_i}) g(t) \, \mathrm{d}\mu \\ &\quad + \frac{2\gamma}{\beta} \int_{\mathcal{E}} \partial_{p_i} g(t) \left(\frac{2 \nabla_p (\partial_{p_i} g(t)) \nabla_p g(t)}{g(t)} - \frac{|\nabla_p g(t)|^2}{g(t)^2} \partial_{p_i} g(t) \right) \mathrm{d}\mu \\ &= -\frac{2}{m_i} \int_{\mathcal{E}} \partial_{p_i} g(t) \, (\partial_{q_i} + \gamma \partial_{p_i}) g(t) \, \mathrm{d}\mu \end{aligned}$$

$$- \frac{2\gamma}{\beta} \int_{\mathcal{E}} |\nabla_p \partial_{p_i} g(t)|^2 - 2\nabla_p(\partial_{p_i} g(t)) \cdot \frac{(\partial_{p_i} g(t))\nabla_p g(t)}{g(t)} \, \mathrm{d}\mu$$
$$- \frac{2\gamma}{\beta} \int_{\mathcal{E}} \frac{|\nabla_p g(t)|^2 (\partial_{p_i} g(t))^2}{g(t)^2} \, \mathrm{d}\mu,$$

where we have used

$$\int_{\mathcal{E}} \partial_{p_i} g(t) \, \mathcal{L}^*(\partial_{p_i} g(t)) \, \mathrm{d}\mu = - \int_{\mathcal{E}} \partial_{p_i} g(t) \, \mathcal{L}_{\mathrm{ham}}(\partial_{p_i} g(t)) \, \mathrm{d}\mu$$
$$- \frac{2\gamma}{\beta} \sum_{j=1}^{d} \int_{\mathcal{E}} \partial_{p_i} g(t) \, \partial^*_{p_j} \partial_{p_j} (\partial_{p_i} g(t)) \, \mathrm{d}\mu,$$

and the first integral on the right-hand side vanishes since $\mathcal{L}_{\mathrm{ham}}$ is antisymmetric. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{\mathcal{E}} |\partial_{p_i} g(t)|^2 \, \mathrm{d}\mu \right) = -\frac{2}{m_i} \int_{\mathcal{E}} \partial_{p_i} g(t) \, (\partial_{q_i} + \gamma \partial_{p_i}) g(t) \, \mathrm{d}\mu$$
$$- \frac{2\gamma}{\beta} \int_{\mathcal{E}} \left| \nabla_p \partial_{p_i} g(t) - \frac{(\partial_{p_i} g(t)) \nabla_p g(t)}{g(t)} \right|^2 \mathrm{d}\mu.$$

Similar computations give

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{\mathcal{E}} |\partial_{q_i} g(t)|^2 \, \mathrm{d}\mu \right) = 2 \int_{\mathcal{E}} \partial_{q_i} g(t) \, (\nabla \partial_{q_i} V) \cdot \nabla_p g(t) \, \mathrm{d}\mu$$
$$- \frac{2\gamma}{\beta} \int_{\mathcal{E}} \left| \nabla_p \partial_{q_i} g(t) - \frac{(\partial_{q_i} g(t)) \nabla_p g(t)}{g(t)} \right|^2 \mathrm{d}\mu,$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{\mathcal{E}} \partial_{p_i} g(t) \, \partial_{q_i} g(t) \, \mathrm{d}\mu \right) = \int_{\mathcal{E}} \partial_{p_i} g(t) (\nabla \partial_{q_i} V) \cdot \nabla_p g(t) \, \mathrm{d}\mu$$
$$- \frac{1}{m_i} \int_{\mathcal{E}} \partial_{q_i} g(t) \, (\partial_{q_i} + \gamma \partial_{p_i}) g(t) \, \mathrm{d}\mu$$
$$- \frac{2\gamma}{\beta} \int_{\mathcal{E}} \left(\nabla_p \partial_{q_i} g(t) - \frac{(\partial_{q_i} g(t)) \nabla_p g(t)}{g(t)} \right) \cdot \left(\nabla_p \partial_{p_i} g(t) - \frac{(\partial_{p_i} g(t)) \nabla_p g(t)}{g(t)} \right) \mathrm{d}\mu.$$

Considering

$$S = \begin{pmatrix} a \, \mathrm{Id}_d & b \, \mathrm{Id}_d \\ b \, \mathrm{Id}_d & c \, \mathrm{Id}_d \end{pmatrix},$$

it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t} (\mathscr{E}(f(t))) = -\frac{4\gamma}{\beta} \int_{\mathcal{E}} |\nabla_p g(t)|^2 \, \mathrm{d}\mu + \sum_{i=1}^{d} \int_{\varepsilon} U_i(t) \, \mathrm{d}\mu,$$

where

$$
\begin{aligned}
U_i(t) &= 8a\partial_{q_i} g(t)\,(\nabla\partial_{q_i} V)\cdot\nabla_p g(t) \\
&- 8b\left(-\partial_{p_i} g(t)\,(\nabla\partial_{q_i} V)\cdot\nabla_p g(t) + \frac{1}{m_i}\partial_{q_i} g(t)\,(\partial_{q_i}+\gamma\partial_{p_i})g(t)\right) \\
&- \frac{8c}{m_i}\partial_{p_i} g(t)\,(\partial_{q_i}+\gamma\partial_{p_i})g(t) \\
&- \frac{8\gamma a}{\beta}\left|\nabla_p\partial_{q_i} g(t) - \frac{(\partial_{q_i} g(t))\nabla_p g(t)}{g(t)}\right|^2 \\
&- \frac{16\gamma b}{\beta}\left(\nabla_p\partial_{q_i} g(t) - \frac{(\partial_{q_i} g(t))\nabla_p g(t)}{g(t)}\right)\cdot\left(\nabla_p\partial_{p_i} g(t) - \frac{(\partial_{p_i} g(t))\nabla_p g(t)}{g(t)}\right) \\
&- \frac{8\gamma c}{\beta}\left|\nabla_p\partial_{p_i} g(t) - \frac{(\partial_{p_i} g(t))\nabla_p g(t)}{g(t)}\right|^2.
\end{aligned}
$$

Note that the sum of the last three terms in $U_i(t)$, integrated against μ, can be bounded from above by

$$
-\frac{8\gamma}{\beta}\sum_{i=1}^{d}(aX_i^2 - 2bX_iY_i + cY_i^2), \tag{2.60}
$$

with

$$
\begin{aligned}
X_i &= \left\|\nabla_p\partial_{q_i} g(t) - \frac{(\partial_{q_i} g(t))\nabla_p g(t)}{g(t)}\right\|_{L^2(\mu)}, \\
Y_i &= \left\|\nabla_p\partial_{p_i} g(t) - \frac{(\partial_{p_i} g(t))\nabla_p g(t)}{g(t)}\right\|_{L^2(\mu)}.
\end{aligned}
$$

We assume in the following

$$
ac - b^2 \geqslant 0,
$$

which ensures that the matrix S is non-negative and that (2.60) is non-positive. Then, recalling definition (2.46) for m_+ and m_-,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(\mathscr{E}(f(t))) \leqslant &-4\gamma\left(\frac{1}{\beta}+\frac{2c}{m_+}\right)\|\nabla_p g(t)\|_{L^2(\mu)}^2 - \frac{8b}{m_+}\|\nabla_q g(t)\|_{L^2(\mu)}^2 \\
&+ 8\sum_{i=1}^{d}\int_{\mathcal{E}}(a\partial_{q_i} g(t) + b\partial_{p_i} g(t))\,(\nabla\partial_{q_i} V)\cdot\nabla_p g(t)\,\mathrm{d}\mu \\
&- 8(\gamma b + c)\sum_{i=1}^{d}\frac{1}{m_i}\int_{\mathcal{E}}\partial_{q_i} g(t)\,\partial_{p_i} g(t)\,\mathrm{d}\mu.
\end{aligned}
$$

Since $\nabla^2 V$ is assumed to be bounded and

$$\|\nabla_p g(t)\|^2_{L^2(\mu)} + \|\nabla_q g(t)\|^2_{L^2(\mu)} = \frac{1}{4}\int_{\mathcal{E}} \frac{|\nabla f(t)|^2}{f(t)}\,\mathrm{d}\mu,$$

inequality (2.57) follows by choosing appropriate values for the coefficients a, b, c in S. For instance, $a = b = c = \varepsilon > 0$ with $\varepsilon > 0$ sufficiently small.

To deduce (2.58), we note that, by tensorization of LSI, there exists $R > 0$ such that, for any C^∞ function $h \geqslant 0$ satisfying

$$\int_{\mathcal{E}} h\,\mathrm{d}\mu = 1,$$

the following inequality holds:

$$\int_{\mathcal{E}} h \ln h\,\mathrm{d}\mu \leqslant \frac{1}{2R}\int_{\mathcal{E}} \frac{|\nabla_p h|^2 + |\nabla_q h|^2}{h}\,\mathrm{d}\mu.$$

In addition,

$$\int_{\mathcal{E}} \frac{(\nabla h)^T S \nabla h}{h}\,\mathrm{d}\mu \leqslant \lambda_S \int_{\mathcal{E}} \frac{|\nabla h|^2}{h}\,\mathrm{d}\mu,$$

where λ_S is the largest eigenvalue of S. Combining these two estimates, we obtain the existence of $\kappa > 0$ such that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\mathscr{E}(f(t))) &\leqslant -\kappa\left(\int_{\mathcal{E}} f(t)\ln f(t)\,\mathrm{d}\mu + \int_{\mathcal{E}} \frac{(\nabla f(t))^T S \nabla f(t)}{f(t)}\,\mathrm{d}\mu\right)\\ &= -\kappa\mathscr{E}(f(t)),\end{aligned}$$

which is indeed (2.58). The exponential decay follows by a Gronwall inequality. □

Here again, the final constant κ can be made quite explicit in terms of the various parameters in the model (*e.g.* LSI constant R, bounds on $\nabla^2 V$).

2.3.6. Subelliptic estimates

To conclude this section on Langevin dynamics, let us briefly mention another technique to prove decay estimates for $\mathrm{e}^{t\mathcal{L}}$ and obtain finer spectral properties of the generator $\mathcal{L}$ of Langevin dynamics or its adjoint, namely subelliptic estimates as provided in Eckmann and Hairer (2003) and Hérau and Nier (2004). The techniques for proving these estimates are based on generalizations of Hörmander's techniques. One output is that $\mathcal{L}$ is an operator with a compact resolvent on $L^2(\mu)$ (in the sense that $(\lambda - \mathcal{L})^{-1}$ is compact for any $\lambda \notin \sigma(\mathcal{L})$), so that its spectrum is discrete. Moreover, it can be proved that the spectrum is contained in some cusp-shaped region: there exist $C, \alpha > 0$ such that

$$\sigma(\mathcal{L}) \subset \{z \in \mathbb{C} \mid \mathrm{Re}(z) \geqslant 0,\ |\mathrm{Im}(z)| \leqslant C(1 + \mathrm{Re}(z))^\alpha\}.$$

2.4. Weighted L^∞ estimates

In this section we present a result on the exponential convergence of the evolution operator in some weighted L^∞-spaces, based on the elementary derivation provided in Hairer and Mattingly (2011). Similar results are also provided in Meyn and Tweedie (2009) and Rey-Bellet (2006). We first give the result in a general form on an abstract configuration space $\mathcal{X}$ since it will be used in many situations: discrete-in-time or continuous-in-time dynamics, in terms of positions or positions and momenta. We next apply it to study the convergence of Langevin and overdamped Langevin dynamics.

Let us emphasize that the convergence results provided here are for the semigroup $\mathrm{e}^{t\mathcal{L}}$, which describes how average properties converge to their stationary values. As explained in Section 2.1.1, similar convergence results can then be deduced on the Fokker–Planck semigroup $\mathrm{e}^{t\mathcal{L}^\dagger}$. This semigroup is defined in a space of probability measures integrating the inverse of the weight function used to define the weighted L^∞-spaces. Moreover, in the cases of the Langevin and overdamped Langevin dynamics, the adjoints $\mathcal{L}^*$ of the generator on L^2-spaces weighted by the canonical measure (see (2.7)) are similar to $\mathcal{L}$, which allows us to easily extend convergence results stated for $\mathrm{e}^{t\mathcal{L}}$ to $\mathrm{e}^{t\mathcal{L}^*}$ and therefore to $\mathrm{e}^{t\mathcal{L}^\dagger}$ by duality: see formulas (2.15) and (2.36).

2.4.1. General framework

Consider a stochastic evolution described by its evolution operator P, with associated transition kernel $P(x, \mathrm{d}x')$. This kernel induces an operator on bounded measurable functions as follows: for any $\varphi \in L^\infty(\mathcal{X})$,

$$P\varphi(x) = \int_{\mathcal{X}} \varphi(x') P(x, \mathrm{d}x').$$

For example, a natural way to define a transition kernel for the general SDE (1.11) is the following for a given time $t_0 > 0$: for any measurable set S,

$$P(x, S) = \mathbb{E}^x(\mathbf{1}_S(x_{t_0})). \tag{2.61}$$

Here $\mathbb{E}^x$ is the expectation over the realizations of (1.11) starting from $x_0 = x$, and $\mathbf{1}_S$ is the indicator function of the set S. In view of (2.8), the operator associated with the integral kernel (2.61) is $P = \mathrm{e}^{t_0\mathcal{L}}$.

We make the following assumptions.

Assumption 2.23 (Lyapunov condition). There exists a function $\mathcal{K} : \mathcal{X} \to [1, +\infty)$ and constants $R \geqslant 0$ and $\alpha \in (0, 1)$ such that

$$(P\mathcal{K})(x) \leqslant \alpha\mathcal{K}(x) + R, \quad \text{for all } x \in \mathcal{X}. \tag{2.62}$$

Let us first insist on the fact that we require $\mathcal{K} \geqslant 1$. The Lyapunov condition implies that the dynamics returns to the region of the configuration

space where the values of $\mathcal{K}$ are not too large. Note that it is trivially satisfied when the configuration space $\mathcal{X}$ is bounded (in which case $\mathcal{K} = \mathbf{1}$ is a possible choice). For unbounded spaces, one typically chooses functions $\mathcal{K}$ which go to infinity at infinity, so that the Lyapunov condition ensures that the dynamics returns to a compact region around the origin.

Assumption 2.24 (minorization condition). There exists a constant $\eta \in (0, 1)$ and a probability measure λ such that

$$\inf_{x \in \mathcal{C}} P(x, \mathrm{d}y) \geqslant \eta \, \lambda(\mathrm{d}y),$$

where $\mathcal{C} = \{x \in \mathcal{X} \mid \mathcal{K}(x) \leqslant \mathcal{K}_{\max}\}$ for some $\mathcal{K}_{\max} > 1 + 2R/(1-\alpha)$, where α, R are introduced in Assumption 2.23.

This condition ensures that there is a sufficiently strong coupling of the evolution in the region where the Lyapunov function is bounded by $\mathcal{K}_{\max}$. When the configuration space $\mathcal{X}$ is compact, or when the set $\mathcal{C}$ is compact, this condition is easy to prove for SDEs, since it is implied by a combination of the following two conditions.

(i) Some irreducibility property of the following form: for a given time $t_0 > 0$, and for any $x \in \mathcal{C}$ and Borel set $S \subset \mathcal{X}$ of positive Lebesgue measure, we have $\mathbb{P}^x(x_{t_0}, S) > 0$.

(ii) Some regularity results on the density of the transition kernel, as provided by hypoellipticity (see Section 2.3.1), for instance.

A typical choice for $\lambda(\mathrm{d}y)$ is then the Lebesgue measure restricted to $\mathcal{C}$, namely $\lambda(\mathrm{d}y) = \mathbf{1}_{\mathcal{C}}(y)\,\mathrm{d}y$. For further details, see the application of the general framework provided in Section 2.4.2 to Langevin and overdamped Langevin dynamics.

We next introduce the functional space used to measure convergence, which we denote by $L^\infty_{\mathcal{K}}(\mathcal{X})$. It is the Banach space of measurable functions φ such that

$$\|\varphi\|_{L^\infty_{\mathcal{K}}} = \left\| \frac{\varphi}{\mathcal{K}} \right\|_{L^\infty} < +\infty. \tag{2.63}$$

Theorem 2.25 (exponential convergence in weighted L^∞-spaces). Suppose that Assumptions 2.23 and 2.24 hold. Then P admits a unique invariant probability measure π. This measure is such that

$$\int_{\mathcal{X}} \mathcal{K} \, \mathrm{d}\pi < +\infty. \tag{2.64}$$

Moreover, there exist $C > 0$ and $r \in (0, 1)$ such that, for any $\varphi \in L^\infty_{\mathcal{K}}(\mathcal{X})$

and any $n \in \mathbb{N}$,

$$\left\| P^n \varphi - \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi \right\|_{L^\infty_{\mathcal{K}}} \leqslant C r^n \left\| \varphi - \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi \right\|_{L^\infty_{\mathcal{K}}}. \tag{2.65}$$

The proof of this result can be read in Hairer and Mattingly (2011). Let us summarize the main steps. The first idea is to introduce a family of distances d_a on $\mathcal{X}$ (with $a > 0$), and to consider the induced Lipschitz seminorms $\| \cdot \|_a$ on real-valued functions defined on $\mathcal{X}$. These seminorms turn out to be equivalent to the $L^\infty_{\mathcal{K}}(\mathcal{X})$-norm (up to the addition of constants to the functions under consideration). Moreover, it is possible to prove a contraction principle of the form

$$|P\varphi(x) - P\varphi(y)| \leqslant r d_a(x,y) \|\varphi\|_a \tag{2.66}$$

for some constant $r \in (0,1)$, which implies a contraction principle for the operator P considered on $L^\infty_{\mathcal{K}}(\mathcal{X})$ endowed with the seminorm $\|\cdot\|_a$. The existence of the invariant measure is finally obtained by a fixed-point theorem, and (2.65) is deduced from the contraction inequality upon starting from y distributed according to the invariant measure π. The prefactor $C \geqslant 1$ is related to the equivalence of the $\| \cdot \|_{L^\infty_{\mathcal{K}}}$ and $\| \cdot \|_a$ norms (as for the proof of hypocoercivity; see the comment after Theorem 2.15).

The key point in the proof is therefore to obtain the contraction inequality (2.66). This is done by distinguishing the cases $\mathcal{K}(x) + \mathcal{K}(y) \geqslant R$ (in this situation, the Lyapunov condition is used) or $\mathcal{K}(x) + \mathcal{K}(y) \leqslant R$ (the contraction then relies on the coupling ensured by the minorization condition).

Let us now present a more compact reformulation of (2.65) in terms of operators. Introduce

$$L^\infty_{\mathcal{K},0}(\mathcal{X}) = \left\{ \varphi \in L^\infty_{\mathcal{K}}(\mathcal{X}) \,\middle|\, \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi = 0 \right\}.$$

Then (2.65) can be equivalently written as an inequality for bounded operators on $L^\infty_{\mathcal{K},0}(\mathcal{X})$:

$$\|P^n\|_{\mathcal{B}(L^\infty_{\mathcal{K},0})} \leqslant C r^n. \tag{2.67}$$

The exponential convergence allows us to obtain bounds on the resolvent $\mathrm{Id} - P$, which will prove useful in various contexts (for error estimates on time discretizations or for the definition of the asymptotic variance of Markov chains). The following result is the version of Proposition 2.1 obtained in the context of this section.

Corollary 2.26 (invertibility of $\mathrm{Id} - P$). Let Assumptions 2.23 and 2.24 hold. Then the bounded operator $\mathrm{Id} - P$ is invertible on $L^\infty_{\mathcal{K},0}(\mathcal{X})$,

and

$$\|(\mathrm{Id}-P)^{-1}\|_{\mathcal{B}(L^\infty_{\mathcal{K},0})} \leqslant \frac{C}{1-r}.$$

Proof. Note that the sum $\sum_{n\geqslant 0} P^n$ is convergent in $\mathcal{B}(L^\infty_{\mathcal{K},0})$ in view of (2.67). In addition, a simple computation gives

$$(\mathrm{Id}-P)\sum_{n=0}^{+\infty} P^n = \left(\sum_{n=0}^{+\infty} P^n\right)(\mathrm{Id}-P) = \mathrm{Id}.$$

This shows that $\mathrm{Id}-P$ is invertible and

$$(\mathrm{Id}-P)^{-1} = \sum_{n=0}^{+\infty} P^n. \tag{2.68}$$

Moreover,

$$\|(\mathrm{Id}-P)^{-1}\|_{\mathcal{B}(L^\infty_{\mathcal{K},0})} \leqslant \sum_{n=0}^{+\infty} \|P^n\|_{\mathcal{B}(L^\infty_{\mathcal{K},0})} \leqslant C\sum_{n=0}^{+\infty} r^n = \frac{C}{1-r},$$

which gives the claimed upper bound. □

2.4.2. Application to Langevin and overdamped Langevin dynamics

To apply the convergence results to the stochastic evolutions typically encountered in molecular dynamics, we fix some time $t_0 > 0$ and consider the evolution operator

$$P = \mathrm{e}^{t_0\mathcal{L}},$$

where $\mathcal{L}$ is the generator of the dynamics under consideration. For simplicity, we consider the dynamics (1.8) and (1.9), for which the force $-\nabla V(q)$ is a gradient field. The results obtained below, however, can easily be extended to more general force fields. This is one of the main interests of weighted L^∞ estimates: they do not require structural assumptions on the type of dynamics considered.

We first discuss how to establish the Lyapunov condition (2.62) in Assumption 2.23 for Langevin and overdamped Langevin dynamics. Let us start by recalling that, as discussed after Assumption 2.23, it is trivial to satisfy a Lyapunov condition when the state space is compact, by choosing $\mathcal{K} = \mathbf{1}$. This situation is encountered for overdamped Langevin dynamics when $\mathcal{D}$ is compact. For overdamped Langevin dynamics in unbounded position spaces, or Langevin dynamics (for which the momentum space is always unbounded), the Lyapunov condition (2.62) can be established on the basis of a Lyapunov condition stated in terms of the generator of the dynamics. Indeed, assume that the following inequality is true for some

$a > 0$ and $b \geqslant 0$, and a given Lyapunov function $\mathcal{K} \geqslant 1$:

$$\mathcal{L}\mathcal{K} \leqslant -a\mathcal{K} + b. \tag{2.69}$$

This leads, by a time integration on the interval $[0, s]$, to

$$\mathrm{e}^{s\mathcal{L}}\mathcal{K} \leqslant \mathrm{e}^{-as}\mathcal{K} + \frac{b}{a}(1 - \mathrm{e}^{-as}) \leqslant \mathrm{e}^{-as}\mathcal{K} + \frac{b}{a}. \tag{2.70}$$

In particular, for $s = t_0$, we obtain

$$P\mathcal{K} \leqslant \mathrm{e}^{-at_0}\mathcal{K} + \frac{b}{a}, \tag{2.71}$$

which is (2.62) with $\alpha = \mathrm{e}^{-at_0} \in (0, 1)$ and $R = b/a \geqslant 0$. We therefore choose $\mathcal{K}$ in order for (2.69) to be satisfied. For overdamped Langevin dynamics in unbounded spaces, a typical choice is $\mathcal{K}_n(q) = |q|^n$ for $n \geqslant 2$. In this case, for all $q \in \mathcal{D}$,

$$\mathcal{L}\mathcal{K}_n(q) = n\left(-\nabla V(q) \cdot q + \frac{d + n - 2}{\beta}\right)|q|^{n-2}.$$

Some assumptions on the asymptotic behaviour of the drift are needed at this stage, such as

$$q \cdot \nabla V(q) \geqslant A|q|^2 - B, \quad \text{for all } q \in \mathcal{D}, \tag{2.72}$$

for some constant $A > 0$ and $B \in \mathbb{R}$. This condition is satisfied for potentials $V(q)$ behaving at infinity as $|q|^k$ with $k \geqslant 2$. With (2.72), the Lyapunov condition (2.69) indeed holds since, for all $q \in \mathcal{D}$,

$$\mathcal{L}\mathcal{K}_n(q) \leqslant -An\mathcal{K}_n(q) + \frac{n(d + n - 2)}{\beta}|q|^{n-2} + Bn,$$

so that

$$\limsup_{|q| \to +\infty} \frac{\mathcal{L}\mathcal{K}_n(q)}{\mathcal{K}_n(q)} \leqslant -An.$$

There exist therefore $b_n \geqslant 0$ such that

$$\mathcal{L}\mathcal{K}_n \leqslant -\frac{An}{2}\,\mathcal{K}_n + b_n.$$

For Langevin dynamics, the choice $\mathcal{K}_n(q, p) = 1 + |p|^n$ with $n \geqslant 2$ is appropriate when the position space is compact. Indeed, with m_+ defined in (2.46), for all $(q, p) \in \mathcal{E}$,

$$\begin{aligned}\mathcal{L}\mathcal{K}_n(q, p) &= -np^T\nabla V(q)|p|^{n-2} - \gamma n\left(p^T M^{-1}p - \frac{n + d - 2}{\beta}\right)|p|^{n-2} \\ &\leqslant -\frac{\gamma n}{m_+}|p|^n + n\|\nabla V\|_{L^\infty}|p|^{n-1} + \frac{\gamma n(n + d - 2)}{\beta}|p|^{n-2},\end{aligned}$$

so that there exists $b_n \geqslant 0$ such that

$$\mathcal{L}\mathcal{K}_n \leqslant -\frac{\gamma n}{2m_+}\,\mathcal{K}_n + b_n.$$

For Langevin dynamics in an unbounded position space $\mathcal{D}$, a possible choice is (see Mattingly, Stuart and Higham 2002)

$$\mathcal{K}_n(q,p) = \left(1+H(q,p)-V_-+\frac{\gamma}{2}p^TM^{-1}q+\frac{\gamma^2}{4}q^TM^{-1}q\right)^n, \quad \text{for all } (q,p)\in\mathcal{E}, \tag{2.73}$$

where we assume that the potential energy function $V(q)$ is bounded from below by $V_- > -\infty$ and that there exist $A, B > 0$ and $C \in \mathbb{R}$ such that

$$q^TM^{-1}\nabla V(q) \geqslant AV(q) + Bq^TM^{-1}q + C, \quad \text{for all } q \in \mathcal{D}. \tag{2.74}$$

This condition is satisfied for instance when $V(q)$ behaves at infinity as $|q|^k$ with $k \geqslant 2$. Then, for any $n \geqslant 1$, we have $\mathcal{K}_n(q,p) \geqslant 1$ and $\mathcal{K}_n(q,p) \to +\infty$ as $|(q,p)| \to +\infty$. In addition, there exist $a > 0$ and $b_n \in \mathbb{R}$ such that

$$\mathcal{L}\mathcal{K}_n \leqslant -na\mathcal{K}_n + b_n.$$

Note that the Lyapunov function has a component $p^TM^{-1}q$ which allows us to retrieve some dissipation in the q-direction from the dissipation in the momenta. This feature is reminiscent of the cross-derivative used in the scalar product introduced for the proof of hypocoercivity in Section 2.3 (see the terms proportional to b in (2.38)).

Let us now discuss how to obtain the minorization condition in Assumption 2.24. Note first that this condition can be rewritten as: for any Borel set $S \subset \mathcal{X}$,

$$P(x,S) \geqslant \eta\lambda(S).$$

A convenient way to obtain such a lower bound is to first establish that the transition kernel can be rewritten as

$$P(x,S) = \int_S p_{t_0}(x,y)\,\mathrm{d}y, \tag{2.75}$$

where the density $p_{t_0}(x,y)$ of the transition kernel is a C^∞ function of $(x,y) \in \mathcal{X}^2$. This is the case, appealing to parabolic regularity for overdamped Langevin dynamics and hypoellipticity for Langevin dynamics (see for instance the discussion in Section 8 of Rey-Bellet 2006). The second step is to prove that, for any open set $\mathcal{O} \subset \mathcal{X}$ and any $x \in \mathcal{X}$, we have

$$P(x,\mathcal{O}) > 0. \tag{2.76}$$

This is proved by controllability arguments, as reviewed in Section 7 of Rey-Bellet (2006), where the cases of Langevin and overdamped Langevin are explicitly treated; see also Mattingly, Stuart and Higham (2002) and Talay (2002) for Langevin dynamics. The idea is to choose some end configuration

$y \in \mathcal{O}$, and to construct a realization of the Brownian motion such that the dynamics started at x ends up at y at time t_0. The continuity of the solutions of SDEs with respect to the realization of the Brownian motion then leads to (2.76). In view of (2.75), the condition (2.76) in fact shows that, for any compact set $\mathcal{C} \subset \mathcal{X}$,

$$p_{t_0}(x,y) > 0, \quad \text{for all } (x,y) \in \mathcal{C}^2.$$

Since $(x,y) \mapsto p_{t_0}(x,y)$ is continuous, this proves the existence of $\delta_{\mathcal{C}} > 0$ such that $p_{t_0}(x,y) \geqslant \delta_{\mathcal{C}}$ for all $(x,y) \in \mathcal{C}^2$. The minorization condition then follows by considering the probability measure

$$\lambda(\mathrm{d}y) = \frac{1}{|\mathcal{C}|}\mathbf{1}_{\mathcal{C}}(y)\,\mathrm{d}y,$$

and setting $\eta = \delta_{\mathcal{C}}|\mathcal{C}|$.

At this stage, we have discussed how to obtain Assumptions 2.23 and 2.24. This already implies the decay estimates (2.65) by Theorem 2.25. These bounds correspond to evolutions observed at integer multiples of the reference time t_0. Let us now finally show how to deduce an exponential convergence result at all times $t \geqslant 0$. To this end, we decompose the time t as $t = kt_0 + \theta$ with $k \in \mathbb{N}$ and $\theta \in [0, t_0)$. Then, for any $\varphi \in L^\infty_{\mathcal{K},0}(\mathcal{X})$ we obtain

$$\|\mathrm{e}^{t\mathcal{L}}\varphi\|_{L^\infty_{\mathcal{K}}} \leqslant Cr^k \|\mathrm{e}^{\theta\mathcal{L}}\varphi\|_{L^\infty_{\mathcal{K}}} \leqslant Cr^k \sup_{0\leqslant s\leqslant t_0} \|\mathrm{e}^{s\mathcal{L}}\varphi\|_{L^\infty_{\mathcal{K}}}. \tag{2.77}$$

Now, the inequality

$$|\varphi(x)| \leqslant \|\varphi\|_{L^\infty_{\mathcal{K}}}\mathcal{K}(x), \quad \text{for all } x \in \mathcal{X},$$

leads to

$$\|\mathrm{e}^{s\mathcal{L}}\varphi\|_{L^\infty_{\mathcal{K}}} \leqslant \|\varphi\|_{L^\infty_{\mathcal{K}}} \left\| \frac{\mathrm{e}^{s\mathcal{L}}\mathcal{K}}{\mathcal{K}} \right\|_{L^\infty}.$$

In view of the inequality (2.70), for any $s \geqslant 0$ we obtain

$$\left\| \frac{\mathrm{e}^{s\mathcal{L}}\mathcal{K}}{\mathcal{K}} \right\|_{L^\infty} \leqslant \mathrm{e}^{-as} + \frac{b}{a}\left\| \frac{1}{\mathcal{K}} \right\|_{L^\infty} \leqslant 1 + \frac{b}{a}.$$

Upon introducing $\kappa = -\log(r)/t_0 > 0$, (2.77) implies

$$\|\mathrm{e}^{t\mathcal{L}}\varphi\|_{L^\infty_{\mathcal{K}}} \leqslant C\left(1 + \frac{b}{a}\right)\mathrm{e}^{-\kappa k t_0}\|\varphi\|_{L^\infty_{\mathcal{K}}} \leqslant \widetilde{C}\mathrm{e}^{-\kappa t}\|\varphi\|_{L^\infty_{\mathcal{K}}}, \tag{2.78}$$

with

$$\widetilde{C} = C\left(1 + \frac{b}{a}\right)\mathrm{e}^{\kappa t_0}.$$

The exponential convergence result (2.78) leads to the following bounds on the resolvent.

Corollary 2.27. Consider one of the following situations:

(1) For overdamped Langevin dynamics, set $\mathcal{X} = \mathcal{D}$ and $\pi = \nu$, and assume either that the position space is compact, in which case $\mathcal{K} = \mathbf{1}$, or that (2.72) holds, in which case $\mathcal{K}(q) = 1 + |q|^n$ for some $n \geqslant 2$.

(2) For Langevin dynamics, set $\mathcal{X} = \mathcal{E}$ and $\pi = \mu$, and assume either that the position space is compact, in which case $\mathcal{K}(q,p) = 1 + |p|^n$ for some $n \geqslant 2$, or that (2.74) holds, in which case $\mathcal{K}$ is given by (2.73) for some $n \geqslant 1$.

Then the operator $\mathcal{L}$ is invertible on $L^\infty_{\mathcal{K},0}(\mathcal{X})$, and the following equality holds in $\mathcal{B}(L^\infty_{\mathcal{K},0})$:

$$\mathcal{L}^{-1} = -\int_0^{+\infty} \mathrm{e}^{t\mathcal{L}}\,\mathrm{d}t.$$

Moreover,

$$\|\mathcal{L}^{-1}\|_{\mathcal{B}(L^\infty_{\mathcal{K},0})} \leqslant \frac{\widetilde{C}}{\kappa}, \tag{2.79}$$

where $\widetilde{C}$ and κ are the same constants as in (2.78).

Although it is possible to keep track of the values of the various constants until (2.78), the final rate of convergence in weighted L^∞-spaces is often not very sharp. Generally, the difficulty is in controlling the constant η in the minorization condition of Assumption 2.24.

Remark 2.28 (sub-exponential convergence rates). It is possible to weaken the Lyapunov condition (2.69) for instance as $\mathcal{L}\mathcal{K} \leqslant -\phi(\mathcal{K}) + b$, where ϕ is a non-negative, non-decreasing, concave function such that $\phi(x)/x \to 0$ as $x \to +\infty$. In this case, algebraic rates of decay are typically obtained instead of exponential ones; see Douc, Fort and Guillin (2009) or Cattiaux (2014) for further details.

2.5. Spectral approaches

One interest of the approaches mentioned in the previous sections is that they can also be used for dynamics other than Langevin or overdamped Langevin dynamics. This is illustrated in Section 4, for example, where entropy techniques and logarithmic Sobolev inequalities play a crucial role in analysing non-linear partial parabolic equations related to adaptive importance sampling methods, and in Section 5, where Lyapunov techniques are useful for the study of non-reversible perturbations of Langevin or overdamped Langevin dynamics. However, when discussing the long-time behaviour of linear parabolic partial differential equations and the link with metastable stochastic processes, we could not avoid mentioning spectral approaches and semi-classical analysis.

Let us consider the overdamped Langevin process. Recall the definitions of the infinitesimal generator $\mathcal{L}$ (see (1.13)) and its adjoint $\mathcal{L}^\dagger$:

$$\mathcal{L} = -\nabla V \cdot \nabla + \beta^{-1}\Delta,$$

which is symmetric on the Hilbert space $L^2(\nu)$, and

$$\mathcal{L}^\dagger = \operatorname{div}(\nabla V .) + \beta^{-1}\Delta,$$

which is symmetric on the Hilbert space $L^2(1/\nu)$. As explained in Section 2.2, the exponential convergence to equilibrium of the Kolmogorov forward dynamics

$$\partial_t \psi = \mathcal{L}^\dagger \psi \tag{2.80}$$

in the weighted $L^2(1/\nu)$-space is related to the fact that ν satisfies a Poincaré inequality (2.17). The two following assertions are indeed equivalent by Proposition 2.3 and the fact that ψ satisfies (2.80) if and only if $\partial_t(\psi/\nu) = \mathcal{L}(\psi/\nu)$:

- For all ψ_0 such that $\psi_0/\nu \in L^2(\nu)$, and for all $t \geqslant 0$,

$$\left\| \frac{\psi(t,\cdot)}{\nu} - 1 \right\|^2_{L^2(\nu)} \leqslant \left\| \frac{\psi_0}{\nu} - 1 \right\|^2_{L^2(\nu)} \exp\left(-\frac{4Rt}{\beta} \right),$$

  where ψ satisfies (2.80) with initial condition $\psi(0,\cdot) = \psi_0$, and $R > 0$ is some constant.

- The measure ν satisfies the following Poincaré inequality: for all $\varphi \in H^1(\nu)$,

$$\int_{\mathcal{D}} \left(\varphi - \int_{\mathcal{D}} \varphi \, \mathrm{d}\nu \right)^2 \mathrm{d}\nu \leqslant \frac{1}{2R} \int_{\mathcal{D}} |\nabla \varphi|^2 \, \mathrm{d}\nu.$$

By noting that

$$\int_{\mathcal{D}} |\nabla \varphi|^2 \, \mathrm{d}\nu = -\beta \int_{\mathcal{D}} \varphi \, (\mathcal{L}\varphi) \, \mathrm{d}\nu, \tag{2.81}$$

we observe that the optimal constant R is related to the first non-zero eigenvalue of the operator $-\mathcal{L}$ on $L^2(\nu)$:

$$\frac{2R}{\beta} = \lambda_1 = \inf_{\substack{\varphi \in H^1(\nu) \\ \int_{\mathcal{D}} \varphi \, \mathrm{d}\nu = 0}} \frac{-\int_{\mathcal{D}} \varphi \, (\mathcal{L}\varphi) \, \mathrm{d}\nu}{\int_{\mathcal{D}} \varphi^2 \, \mathrm{d}\nu}. \tag{2.82}$$

Recall that the first eigenvalue of $\mathcal{L}$ on $L^2(\nu)$ is $\lambda_0 = 0$, with associated eigenfunctions proportional to the constant function $\mathbf{1}$. The corresponding eigenspace is non-degenerate since $\mathcal{L}\varphi = 0$ implies that φ is constant by (2.81). Therefore, the second eigenvalue λ_1 obtained by the Rayleigh–Ritz principle (2.82) is indeed the first non-zero eigenvalue of $-\mathcal{L}$. The difference $\lambda_1 - \lambda_0 = \lambda_1$ is called the spectral gap of the operator. The above

discussion already shows the importance of the spectrum of the operator $\mathcal{L}$ in understanding the long-time behaviour of (2.80).

In fact, in order to understand and quantify the metastability of overdamped Langevin dynamics, it is interesting to consider not only the first eigenvalue of $\mathcal{L}$ but an ensemble of small eigenvalues. In order to explain this, we need to introduce the notion of the Witten Laplacian. The Witten Laplacian is an operator which is unitarily equivalent to $\mathcal{L}$ and $\mathcal{L}^\dagger$, and which acts on the non-weighted (flat) L^2-space. Let us first recall that the spectrum of $\mathcal{L}$ is the same as the spectrum of $\mathcal{L}^\dagger$ in view of the unitary equivalence

$$\mathcal{L}^\dagger = U\mathcal{L}U^{-1},$$

where U is the unitary transformation defined by

$$U : \left\{ \begin{array}{rcl} L^2(\nu) & \to & L^2(1/\nu), \\ \varphi & \mapsto & \varphi\nu. \end{array} \right.$$

Now, the spectrum of these operators is also related to the spectrum of a Witten Laplacian. For a function $f : \mathbb{R}^d \to \mathbb{R}$ and a positive parameter h, the Witten Laplacian[8] is defined by

$$\Delta_{f,h} = (-h\nabla + \nabla f) \cdot (h\nabla + \nabla f) = -h^2\Delta + (|\nabla f|^2 - h\Delta f),$$

considered on the flat Hilbert space L^2. Note that this operator is of Schrödinger type, which is convenient since the spectral properties of these operators are quite well understood; see Reed and Simon (1975*b*). The spectrum of $\mathcal{L}$ is $-\beta$ times the spectrum of $\Delta_{V/2,\beta^{-1}}$ since $\mathcal{L}$ and $-\beta\Delta_{V/2,\beta^{-1}}$ are unitarily equivalent (see (2.14)), that is,

$$-\beta\Delta_{V/2,\beta^{-1}} = \widetilde{U}\mathcal{L}\widetilde{U}^{-1},$$

where $\widetilde{U}$ is the unitary transformation

$$\widetilde{U} : \left\{ \begin{array}{rcl} L^2(\nu) & \to & L^2, \\ \varphi & \mapsto & \varphi\sqrt{\nu}. \end{array} \right.$$

Many outstanding mathematicians have contributed to the analysis of the small eigenvalues of Witten Laplacians, and the associated eigenvectors, in the so-called semi-classical limit, namely when $h \to 0$. In our context this corresponds to the low-temperature regime $\beta^{-1} \to 0$. We refer the reader to Witten (1982), Helffer and Sjöstrand (1985), Cycon, Froese, Kirsch and Simon (1987), Burghelea (1997) and Zhang (2001) for an introduction to the semi-classical analysis of Witten Laplacians and its famous application to Morse inequalities. It is not our aim here to present the details of this theory.

[8] Notice that by convention, the Witten Laplacian $\Delta_{f,h}$ is a non-negative operator. This explains the sign convention.

Let us simply mention the main results that are of interest in the context of understanding the metastability of overdamped Langevin dynamics. We assume that the potential $V : \mathbb{R}^d \to \mathbb{R}$ is a Morse function (*i.e.*, its critical points are non-degenerate) and that

$$\lim_{|x|\to\infty} \frac{|\nabla V|^2}{2} - \beta^{-1}\Delta V = +\infty$$

for sufficiently small β^{-1} (so the spectrum of $\Delta_{V/2,\beta^{-1}}$ is discrete. Then, we have the following asymptotic results, in the limit $\beta \to \infty$ (see Helffer, Klein and Nier 2004).

- The operator $\Delta_{V/2,\beta^{-1}}$ admits exactly m_0 eigenvalues (including 0) which are smaller than $\exp(-c\beta)$ for a constant $c > 0$. All the other eigenvalues are bounded from below by a constant. The integer m_0 is the number of local minima of V.

- The associated eigenfunctions are of the form $\chi_i \exp(-\beta V/2)$, where the functions χ_i are locally constant over the basins of attraction of the local minima of V for the gradient dynamics $\dot{x} = -\nabla V(x)$.

- The smallest eigenvalue is $\lambda_0 = 0$, associated with the eigenvector $\exp(-\beta V/2)$. The other positive small eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_{m_0-1}$ satisfy the Eyring–Kramers law: for $i = 1, \ldots, m_0 - 1$,

$$\lambda_i = \frac{|\hat{\lambda}(s_{j(i)})|}{2\beta\pi} \sqrt{\frac{\det(\nabla^2 V(x_i))}{|\det(\nabla^2 V(s_{j(i)}))|}}\, \mathrm{e}^{-\beta[V(s_{j(i)})-V(x_i)]}(1 + \mathrm{O}(\beta^{-1})), \tag{2.83}$$

  where $\{x_0, x_1, \ldots, x_{m_0-1}\}$ is the set of local minima of V and, for $i \in \{1, \ldots, m_0-1\}$, $s_{j(i)}$ is a saddle point attached to the local minimum x_i, $\hat{\lambda}(s_{j(i)})$ being the negative eigenvalue of $\nabla^2 V(s_{j(i)})$. Let us now explain how a saddle point $s_{j(i)}$ is attached to each local minimum x_i. From a local minimum x_i, one considers the communication height between x_i and the union of small balls around local minima which are lower in energy than $V(x_i)$, denoted by B_i:

$$H(x_i, B_i) = \inf_{\gamma\in\mathcal{P}(x_i,B_i)} \sup_{t\in[0,1]} V(\gamma(t)) - V(x_i),$$

  where $\mathcal{P}(x_i, B_i)$ denotes the set of continuous paths from $[0, 1]$ to $\mathcal{D}$ such that $\gamma(0) = x_i$ and $\gamma(1) \in B_i$. For each $i \in \{1, \ldots, m_0 - 1\}$, the communication height $H(x_i, B_i)$ is generally reached at a unique order-one saddle point $s_{j(i)}$:

$$H(x_i, B_i) = V(s_{j(i)}) - V(x_i).$$

Note that we implicitly assume here that x_0 is the global minimum and the other local minima $(x_i)_{1\leqslant i\leqslant m_0-1}$ are ordered such that

$$H(x_1, B_1) > \cdots > H(x_{m_0-1}, B_{m_0-1}),$$

in order to ensure that $\lambda_1 < \lambda_2 < \cdots < \lambda_{m_0}$.

One ingredient for the proof of the results provided in Helffer, Klein and Nier (2004) is to introduce the Witten Laplacian on 1-forms, which can be identified with the following operator on vector fields: for $w : \mathcal{D} \mapsto \mathbb{R}^d$,

$$(\Delta^{(1)}_{f,h} w)_i = -h^2 \Delta w_i + (|\nabla f|^2 - h\Delta f) w_i + 2h((\nabla^2 f) w)_i.$$

Note that if $u : \mathcal{D} \to \mathbb{R}$ is an eigenvector for $\Delta_{f,h}$, then $h\nabla u + u\nabla f$ is either 0 or an eigenvector for $\Delta^{(1)}_{f,h}$. Then one can see that the eigenvalues of $\Delta_{f,h}$ are the square of the singular values of the operator $d_{f,h} = h\nabla + \nabla f$ which maps 0-forms to 1-forms, since $\Delta_{f,h} = d^*_{f,h} d_{f,h}$ (where $d^*_{f,h} = -h \operatorname{div} + \nabla f$ is the L^2-adjoint of $d_{f,h}$). Using a basis of quasi-modes associated with the small eigenvalues of $\Delta_{f,h}$ and $\Delta^{(1)}_{f,h}$, one is then able to analyse the singular values of $d_{f,h}$ restricted to the finite-dimensional subspaces spanned by these quasi-modes. The quasi-modes are built using the following important fact: the eigenvectors of $\Delta_{f,h}$ (resp. of $\Delta^{(1)}_{f,h}$) are concentrated around the local minima (resp. the order-one saddle points) of the function f. This can be proved using so-called semi-classical Agmon estimates (Simon 1984, Helffer and Sjöstrand 1984). For generalizations to manifolds with boundaries and to Witten Laplacians on p-forms, we refer to Helffer and Nier (2006) and Le Peutrec, Nier and Viterbo (2013).

Bovier *et al.* (2004*a*, 2004*b*) developed another analytic approach to proving (2.83). In particular, they proved that, for $i \in \{1, \ldots, m_0 - 1\}$,

$$\lambda_i^{-1} = \mathbb{E}_{x_i}(\tau_{B_i})(1 + \mathrm{O}(\mathrm{e}^{-\beta c}))$$

for some positive constant $c > 0$, where $\mathbb{E}_{x_i}(\tau_{B_i})$ denotes the mean time to reach B_i starting from x_i.

We would like to emphasize that we do not provide these results with all the required assumptions (in particular, it is assumed that the communication heights are all different, and that the communication height is attained at a unique saddle point). For more details of this subject, we refer for example to the lecture notes by Helffer and Nier (2005), and also to the nice review by Berglund (2013).

From a modelling viewpoint, this spectral approach gives the cascade of relevant time scales to reach from a local minimum x_i (for $i \in \{1, \ldots, m_0 - 1\}$) any other local minimum which is lower in energy (and in particular the global minimum x_0). This does not provide us with any information about the typical time scale to go from any local minimum to any other local minimum (for example from the global minimum to the second lower

minimum). We will come back to this in Sections 6.3 and 6.4, where we use a spectral approach on bounded domains (rather than on the full state space $\mathcal{D}$) to derive a Markov state model; see also the article by Lelièvre and Nier 2015. We also refer to the works by Maier and Stein (1993, 1997), Naeh, Klosek, Matkowsky and Schuss (1990) and Matkowsky and Schuss (1979) for other approaches to this problem using formal expansions of singularly perturbed partial differential equations. An interesting work relating spectral gaps for the infinitesimal generator restricted to metastable states with the global spectral gap is that of Menz and Schlichting (2014). Notice also that spectral approaches are very relevant to energetic barriers but not to entropic barriers (for which the Morse assumption may typically be wrong: see Figure 1.1(c, d)).

For numerical purposes, these results are interesting for the following reason. As explained above, the existence of small eigenvalues can be related to the existence of metastable states, characterized as the regions where the eigenfunctions of $\mathcal{L}$ are constant. This can be used to identify the metastable states. More importantly, if one wants to define an optimal Galerkin projection of the operator $\mathcal{L}$ onto a discrete space, it is natural to introduce the Galerkin basis consisting of the eigenvectors associated with the smallest eigenvalues. This indeed leads to the smallest errors (in the $L^2(\nu)$-norm) for the semi-group $\mathrm{e}^{t\mathcal{L}}$. By using the Galerkin projection, one can thus build from the original dynamics an approximated Markov dynamics in a discrete state space (a so-called Markov state model, or kinetic Monte Carlo model: see Bowman, Pande and Noé 2014). This has been extensively investigated by C. Schütte and his collaborators, starting with the seminal work of Schütte (1998). In fact, Schütte *et al.* look at the eigenvalues close to 1 for the so-called transfer operator $P_t = \mathrm{e}^{t\mathcal{L}}$ (for a well-chosen lag time $t > 0$), which is equivalent to looking at the small positive eigenvalues of $-\mathcal{L}$. In practice, the eigenfunctions can be approximated using the committor functions associated with core sets (*i.e.*, disjoint subsets of the configuration space $\mathcal{D}$). For more details of this approach, we refer to the recent book by Sarich and Schütte (2013). We will come back to the relation between original Langevin or overdamped Langevin dynamics with a discrete state-space Markov process in Section 6.3 below, where another projection technique will be presented using the notion of quasi-stationary distributions.

For Langevin dynamics, similar questions can be raised. For example, the Eyring–Kramers law (2.83) can be generalized to Langevin dynamics: see Maier and Stein (1993, 1997), Naeh *at al.* (1990), Matkowsky and Schuss (1979), Hérau, Hitrik and Sjöstrand (2011) or Bouchet and Reygner (2015). In general, the eigenvectors cannot be written in a separated form in terms of positions and momenta. Metastable states should therefore be defined both in terms of positions and velocities. This raises some difficulties from

modelling and numerical viewpoints. We refer to Sarich and Schütte (2013) for discussions on this aspect.

Remark 2.29. In this section we have discussed the spectral properties of $\mathcal{L}$ in the weighted $L^2(\nu)$-space since this is a very natural functional space for overdamped Langevin dynamics, in particular in terms of self-adjointness. There are also interesting results concerning the spectrum in the weighted space $L^\infty_{\mathcal{K}}$, where $\mathcal{K}$ is a Lyapunov function, as introduced in Section 2.4. We refer for example to Sarich and Schütte (2013, Section 4.3.1) for a discussion and references on this question. See also Cattiaux, Guillin and Zitt (2013) for related questions on the link between Lyapunov conditions and spectral gaps, and Cattiaux, Guillin and Wu (2009) for the link between Lyapunov conditions and logarithmic Sobolev inequalities.

3. Computing thermodynamic averages

In this section we discuss the errors in the practical estimation of canonical averages. As already mentioned in Section 1 (see in particular the discussion after (1.5)), the challenging part is to sample the canonical measure $\nu(\mathrm{d}q) = Z_\nu^{-1}\,\mathrm{e}^{-\beta V(q)}\,\mathrm{d}q$ on the position space $\mathcal{D}$. Without restriction of generality, we can therefore restrict ourselves to observables depending only on q. The typical way to estimate canonical averages $\int_{\mathcal{D}} \varphi\,\mathrm{d}\nu$ is to first discretize in time a dynamics ergodic with respect to ν (such as the Langevin (1.9) or overdamped Langevin (1.8) dynamics), and then to compute averages along one realization of this dynamics. The long-time convergence of such an estimator requires the discretized dynamics to be ergodic as well, although the corresponding ergodic measure generally differs from ν because of the time discretization.

In practice, choosing a good initial condition q_0 (namely such that $V(q_0)$ is not too large) may be difficult. This is why a burn-in phase is usually performed, during which the sampled configurations are discarded. Actual averages start after this initial relaxation. We will not discuss this practical aspect further.

Note also that it is possible to run several realizations of the discretized dynamics in parallel, which provides faster results in terms of wall clock time. Some care has to be taken in order to ensure that the initial conditions, and thus the realizations, are indeed independent (for instance by performing independent burn-in phases for each realizations). This computational overhead sometimes motivates practitioners to estimate average properties with a single, long trajectory.

To perform the numerical analysis allowing us to quantify errors made in the trajectorial averages of discretized dynamics, we first review results ensuring the ergodicity of continuous dynamics in Section 3.1. We next

turn to the discretization of continuous dynamics and the ergodic properties of the resulting numerical schemes in Section 3.2, together with a general discussion of the sources of errors. A precise description of each source of error is finally given in the remaining sections: the bias on the invariant measure is studied in Section 3.3, while the quantification (and the possible reduction) of the statistical error is the topic of Section 3.4.

3.1. Ergodicity of continuous dynamics

One of the major aims of molecular simulation is the estimation of averages with respect to high-dimensional probability measures through time averages along a realization of a stochastic differential equation ergodic with respect to the target probability measure π on $\mathcal{X}$:

$$\mathrm{d}x_t = b(x_t)\,\mathrm{d}t + \sigma(x_t)\,\mathrm{d}W_t. \tag{3.1}$$

For a given observable φ, the estimator under consideration reads

$$\widehat{\varphi}_t = \frac{1}{t}\int_0^t \varphi(x_s)\,\mathrm{d}s.$$

3.1.1. Almost sure convergence of $\widehat{\varphi}_t$

Let us first discuss the convergence of the above estimator, in the case when the generator $\mathcal{L}$ of the process (3.1) is elliptic or hypoelliptic. The results by Kliemann (1987) show that, when the dynamics admits an invariant probability measure π with a positive density with respect to the Lebesgue measure, then, for a given observable $\varphi \in L^1(\pi)$, the following almost sure convergence holds for any initial condition $x_0 \in \mathcal{X}$:

$$\lim_{t\to+\infty} \widehat{\varphi}_t = \int_{\mathcal{X}} \varphi(x)\,\pi(\mathrm{d}x).$$

As a consequence, the invariant measure is unique.

3.1.2. Asymptotic variance of $\widehat{\varphi}_t$

Next, convergence rates can be obtained provided a central limit theorem holds. A first step is to show that the asymptotic variance is well defined:

$$\varsigma_\varphi^2 = \lim_{t\to+\infty} t\,\mathrm{Var}_\pi(\widehat{\varphi}_t), \tag{3.2}$$

where

$$t\,\mathrm{Var}_\pi(\widehat{\varphi}_t) = t\,\mathbb{E}_\pi[(\widehat{\varphi}_t)^2 - (\mathbb{E}_\pi\widehat{\varphi}_t)^2] = t\,\mathbb{E}_\pi\left[\left(\frac{1}{t}\int_0^t \Pi\varphi(x_s)\,\mathrm{d}s\right)^2\right].$$

In these expressions, the expectations are with respect to initial conditions $x_0 \sim \pi$ and for all realizations of the stochastic dynamics. Further, we have

introduced the projection operator Π:

$$\Pi\varphi = \varphi - \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi. \tag{3.3}$$

Note finally that we used the invariance of π to write $\mathbb{E}_\pi \widehat{\varphi}_t = \mathbb{E}_\pi \varphi$. To prove that (3.2) is indeed well defined, we rewrite it in terms of the semigroup $\mathrm{e}^{t\mathcal{L}}$. We expand $t \operatorname{Var}_\pi(\widehat{\varphi}_t)$ as

$$\begin{aligned} t \operatorname{Var}_\pi(\widehat{\varphi}_t) &= \frac{1}{t} \int_0^t \int_0^t \mathbb{E}_\pi[\Pi\varphi(x_s)\Pi\varphi(x_r)] \, \mathrm{d}r \, \mathrm{d}s \\ &= \frac{2}{t} \int_0^t \int_0^s \mathbb{E}_\pi[\Pi\varphi(x_s)\Pi\varphi(x_r)] \, \mathrm{d}r \, \mathrm{d}s, \end{aligned}$$

and use the stationarity of the process (valid since the dynamics is started under the invariant measure π) to write, for $s \geqslant r$,

$$\mathbb{E}_\pi[\Pi\varphi(x_s)\Pi\varphi(x_r)] = \mathbb{E}_\pi[\Pi\varphi(x_{s-r})\Pi\varphi(x_0)].$$

Then,

$$\begin{aligned} t \operatorname{Var}(\widehat{\varphi}_t) &= 2 \int_0^t \left(1 - \frac{s}{t}\right) \mathbb{E}_\pi[\Pi\varphi(x_s)\Pi\varphi(x_0)] \, \mathrm{d}s \\ &= 2 \int_0^t \left(1 - \frac{s}{t}\right) \int_{\mathcal{X}} (\mathrm{e}^{s\mathcal{L}}\Pi\varphi)(x)\Pi\varphi(x) \, \pi(\mathrm{d}x) \, \mathrm{d}s. \end{aligned}$$

Let us now assume that decay estimates for the semigroup $\mathrm{e}^{s\mathcal{L}}$ are available in a functional space $\mathscr{E} \subset L^2_0(\pi)$, typically

$$\|\mathrm{e}^{s\mathcal{L}}\|_{\mathcal{B}(\mathscr{E})} \leqslant C \, \mathrm{e}^{-\kappa s}, \quad \text{for all } t \geqslant 0, \tag{3.4}$$

for some positive constants C and κ. Various ways of proving such inequalities are provided in Section 2, with $\mathscr{E}$ being $L^2_0(\pi)$ (see Propositions 2.3 and 2.20), $H^1(\pi) \cap L^2_0(\pi)$ (see Theorem 2.15) or $L^\infty_{\mathcal{K},0}(\mathcal{X})$ (see (2.78)). If $\varphi \in \mathscr{E}$, then the dominated convergence theorem shows that

$$\lim_{t\to\infty} t \operatorname{Var}_\pi(\widehat{\varphi}_t) = \varsigma^2_\varphi,$$

with

$$\varsigma^2_\varphi = 2 \int_0^{+\infty} \int_{\mathcal{X}} (\mathrm{e}^{s\mathcal{L}}\Pi\varphi)\Pi\varphi \, \mathrm{d}\pi \, \mathrm{d}s = -2 \int_{\mathcal{X}} (\mathcal{L}^{-1}\Pi\varphi)\Pi\varphi \, \mathrm{d}\pi. \tag{3.5}$$

Recall that $\mathcal{L}^{-1}$ is a well-defined bounded operator on $\mathscr{E}$ when decay estimates such as (3.4) hold (see Proposition 2.1 and Corollaries 2.4, 2.16, 2.21 and 2.27).

It is instructive at this stage to compare the asymptotic variance (3.5) to the one obtained by averages of independent and identically distributed (i.i.d.) random variables $(x^n)_{n\geqslant 1}$ with common law π. When $\varphi \in L^2(\pi)$, a

central limit theorem holds true for the estimator

$$\widehat{\varphi}_N^{\mathrm{iid}} = \frac{1}{N}\sum_{n=1}^{N} \varphi(x^n),$$

whose asymptotic variance is

$$\varsigma^2_{\varphi,\mathrm{iid}} = \lim_{N\to+\infty} N\mathbb{E}_\pi\left[(\widehat{\Pi\varphi}_N^{\mathrm{iid}})^2\right] = \int_{\mathcal{X}} (\Pi\varphi)^2 \,\mathrm{d}\pi.$$

We write the asymptotic variance (3.5) for time average estimates with ergodic SDEs in terms of the reference variance $\varsigma^2_{\varphi,\mathrm{iid}}$ as

$$\varsigma^2_\varphi = \theta_{\mathrm{corr},\varphi}\varsigma^2_{\varphi,\mathrm{iid}}, \tag{3.6}$$

and interpret $\theta_{\mathrm{corr},\varphi}$ as some correlation time. Then the mean-square error of the estimator $\widehat{\varphi}_t$ asymptotically behaves as $\varsigma^2_\varphi/t = \varsigma^2_{\varphi,\mathrm{iid}}\theta_{\mathrm{corr},\varphi}/t$, so that in order to have an estimator of the same quality as that based on N i.i.d. samples, integration times of order $t = N\theta_{\mathrm{corr},\varphi}$ should be considered. When the operator $\mathcal{L}$ is self-adjoint on $L^2(\pi)$, the norm of $\mathcal{L}^{-1}$ on $L^2_0(\pi)$ is equal to $1/\lambda_1$, λ_1 being the first non-zero eigenvalue of $-\mathcal{L}$ on $L^2(\pi)$ (see the discussion in Section 2.5). This provides the following upper bound for the correlation time: $\theta_{\mathrm{corr},\varphi} \leqslant 1/\lambda_1$. Equality holds when φ is an eigenvector of $-\mathcal{L}$ associated with λ_1.

3.1.3. Central Limit Theorems

Note that (3.5) shows that it is possible to define the asymptotic variance without any reference to the evolution semigroup $\mathrm{e}^{t\mathcal{L}}$, by directly considering

$$\varsigma^2_\varphi = -2\int_{\mathcal{X}} (\mathcal{L}^{-1}\Pi\varphi)\Pi\varphi \,\mathrm{d}\pi.$$

The integral on the right-hand side is well defined once the solution Φ of the Poisson equation

$$-\mathcal{L}\Phi = \Pi\varphi \tag{3.7}$$

belongs to a functional space $\mathscr{E} \subset L^2(\pi)$. This is the case when $\varphi \in \mathscr{E}$ and $\mathcal{L}^{-1}$ is a bounded operator on $\mathscr{E}$; see the discussion after (3.5). However, there are more general frameworks to define solutions of Poisson equations; see for instance Pardoux and Veretennikov (2005) and references therein.

In fact, it was proved by Bhattacharya (1982) that a central limit theorem holds once the Poisson equation (3.7) has a solution in $L^2(\pi)$ and the initial conditions are distributed according to π. In this case,

$$\lim_{t\to\infty} \sqrt{t}\,\widehat{\Pi\varphi}_t = \mathcal{N}(0,\varsigma^2_\varphi) \text{ in law.}$$

This result can be extended to cover the case when the initial conditions are not distributed according to the invariant measure: see also Bhattacharya

(1982). The bottom line of the proof is to use Itô calculus to rewrite $\sqrt{t}\,\widehat{\Pi\varphi}_t$ as

$$\sqrt{t}\,\widehat{\Pi\varphi}_t = \frac{\Phi(x_0) - \Phi(x_t)}{\sqrt{t}} + \mathcal{M}_t, \quad \mathcal{M}_t = \frac{1}{\sqrt{t}} \int_0^t \nabla\Phi(x_s)^T \sigma(x_s)\, \mathrm{d}W_s.$$

The first term converges to 0 in law as $t \to +\infty$, while a central limit theorem for martingales can be applied to obtain the asymptotic behaviour of the second term (see for instance Komorowski, Landim and Olla 2012 or Ethier and Kurtz 1986): $\mathcal{M}_t$ converges in law to a Gaussian distribution with variance

$$\tilde{\varsigma}^2 = \lim_{t\to+\infty} \mathbb{E}\left(\frac{1}{t}\int_0^t \nabla\Phi(x_s)^T\sigma(x_s)\sigma(x_s)^T\nabla\Phi(x_s)\,\mathrm{d}s\right).$$

By ergodicity of the dynamics,

$$\tilde{\varsigma}^2 = \int_{\mathcal{X}} \nabla\Phi^T\sigma\sigma^T\nabla\Phi\,\mathrm{d}\pi.$$

Now, recalling the expression (1.12) for the generator, a simple computation shows that

$$\mathcal{L}(\Phi^2) - 2\Phi\mathcal{L}\Phi = \sigma\sigma^T : \nabla^2(\Phi^2) - 2\Phi\sigma\sigma^T : \nabla^2\Phi = \nabla\Phi^T\sigma\sigma^T\nabla\Phi.$$

The invariance of π, expressed as (1.17), then leads to

$$\tilde{\varsigma}^2 = \int_{\mathcal{X}} \nabla\Phi^T\sigma\sigma^T\nabla\Phi\,\mathrm{d}\pi = -2\int_{\mathcal{X}} \Phi\mathcal{L}\Phi\,\mathrm{d}\pi = \varsigma_\varphi^2,$$

which allows us to conclude.

3.2. Numerical discretizations and their ergodic behaviour

In practice, continuous dynamics such as (3.1) are discretized in time using a numerical scheme. The simplest choice is to resort to an Euler–Maruyama discretization with a fixed time step $\Delta t > 0$. In this case $x_{n\Delta t}$ is approximated by x^n, constructed iteratively from a given initial condition $x^0 = x_0$ as

$$x^{n+1} = x^n + b(x^n)\,\Delta t + \sigma(x^n)\sqrt{\Delta t}\,G^n, \tag{3.8}$$

where G^n is a sequence of i.i.d. Gaussian random variables with covariance matrix Id_d. Hereafter, the evolution operator (also called the transition operator) associated with a general one-step numerical scheme is denoted by

$$P_{\Delta t}\varphi(x) = \mathbb{E}(\varphi(x^{n+1}) \mid x^n = x). \tag{3.9}$$

It is the discrete counterpart of the semigroup $\mathrm{e}^{\Delta t\mathcal{L}}$. The transition operator $P_{\Delta t}$ describes how the values of a given function evolve on average over one time step.

We discuss in this section the various types of errors arising from the time discretization. We first recall the standard measures of error on finite time intervals (weak versus strong errors). We then turn to our main concern, error estimates on long time averages. As discussed below, these errors can be decomposed, for sufficiently long times, into (i) a systematic part, related to errors on the invariant measure due to the time discretization, and (ii) a statistical error dictated by a central limit theorem, with an asymptotic variance close to that of the underlying continuous process. The quality of the numerical schemes for ergodic dynamics is therefore mainly determined by the size of the systematic errors, as well as their stability.

3.2.1. Standard numerical analysis of SDEs

There are many good review articles and textbooks on the numerical analysis of SDEs, such as Kloeden and Platen (1992), Platen (1999) and Milstein and Tretyakov (2004). Two types of errors are considered for discretizations of SDEs, as follows.

(i) *Weak error estimates.* There exists $\alpha > 0$ such that, for any C^∞ test function with compact support φ and finite time horizon $T > 0$, there are $C \geqslant 0$ and $\Delta t^* > 0$ (the latter two constants depending on φ and T *a priori*) such that, for any $\Delta t \in (0, \Delta t^*]$,

$$\sup_{0 \leqslant n \leqslant T/\Delta t} |\mathbb{E}[\varphi(x^n)] - \mathbb{E}[\varphi(x_{n\Delta t})]| \leqslant C\Delta t^\alpha.$$

In fact, when the numerical method is stable (which is the case when b and σ are globally Lipschitz), such error estimates can be deduced from the error over one time step, as stated in Theorem 2.1 of Milstein and Tretyakov (2004). In essence, the order α is determined by the formal equality

$$P_{\Delta t} = \mathrm{e}^{\Delta t \mathcal{L}} + \mathrm{O}(\Delta t^{\alpha+1}).$$

To make the functional setting more precise, this should be understood as $P_{\Delta t}\varphi = \mathrm{e}^{\Delta t \mathcal{L}}\varphi + \Delta t^{\alpha+1} r_{\varphi,\Delta t}$ for a function φ growing at most polynomially, whose derivatives of order at most $2\alpha + 2$ grow at most polynomially, and with a remainder term $r_{\varphi,\Delta t}$ such that there exists $p \in \mathbb{N}$ for which $\|r_{\varphi,\Delta t}/(1 + |x|^p)\|_{L^\infty} \leqslant K$, for Δt sufficiently small.

(ii) *Strong error estimates in L^p-norm.* There exists $\alpha \in \mathbb{R}_+$ such that, for any time horizon T, there is $C \geqslant 0$ and $\Delta t^* > 0$ (the latter two constants depending on T *a priori*) such that, for any $0 \leqslant \Delta t \leqslant \Delta t^*$,

$$\sup_{0 \leqslant n \leqslant T/\Delta t} (\mathbb{E}|x^n - x_{n\Delta t}|^p)^{1/p} \leqslant C\Delta t^\alpha.$$

In this case the Gaussian increments used in a numerical scheme such as (3.8) must be induced by the Brownian motion for the continuous

dynamics (3.1), that is,

$$G^n = \frac{W_{(n+1)\Delta t} - W_{n\Delta t}}{\sqrt{\Delta t}}.$$

Typically, the constant C is obtained via some (discrete) Gronwall estimate, as in the standard numerical analysis of ordinary differential equations, and hence increases exponentially with time.

As an example, let us mention that, when the functions b and σ in (3.1) are globally Lipschitz, the weak and strong errors of the Euler–Maruyama scheme (3.8) are respectively 1 and 1/2.

Note that it is also possible to reduce errors due to time discretization by resorting to extrapolation methods as in Lamberton and Pagès (2002) or multilevel Monte Carlo methods (Giles 2008, Giles 2015).

3.2.2. Convergence of time averages

The above error estimates are not relevant to long-time convergence since the prefactor C is not uniformly controlled in time. Therefore, additional techniques should be introduced to control the quality of the approximation of average properties presented by the time averages traditionally considered in molecular simulation:

$$\widehat{\varphi}_{N_{\rm iter},\Delta t} = \frac{1}{N_{\rm iter}} \sum_{n=0}^{N_{\rm iter}-1} \varphi(x^n). \tag{3.10}$$

A first issue is to prove that the numerical scheme is indeed ergodic with respect to some probability measure, which depends in general on the time step Δt. Even if the associated continuous dynamics is ergodic, the ergodicity of the discretized dynamics cannot be guaranteed. A typical obstruction is the presence of non-globally Lipschitz drifts which induce a transient behaviour of the Markov chain; see for instance the examples provided in Mattingly *et al.* (2002, Section 6.3) for $b(x) = -x^3$ and $\sigma(x) = 1$.

In order to prove the existence of an invariant probability measure $\pi_{\Delta t}$ for the numerical scheme, it is standard to resort to results such as Theorem 2.25. Proving that a minorization condition holds on any compact set (see Assumption 2.24) is usually not too difficult, thanks to the presence of Gaussian increments; see for instance Lemma 3.4 below for an Euler–Maruyama discretization of the overdamped Langevin dynamics. Some assumptions on the drift, on the other hand, are usually required to state a Lyapunov condition such as (2.62): see Mattingly *et al.* (2002). When these conditions are met, the convergence of the law of x^n to the invariant measure is exponentially fast (with respect to the iteration index, the time step Δt being fixed, as stated in Theorem 2.25).

The next step is to prove that the numerical scheme is irreducible. Usually, irreducibility is defined with respect to the Lebesgue measure, in which

case this property reads: for any Borel subset $S \subset \mathcal{X}$ of positive Lebesgue measure, and any initial condition $x^0 \in \mathcal{X}$, we have $\mathbb{P}_{x^0}(\tau_S < +\infty) > 0$, where $\tau_S = \inf\{n \geqslant 0, x^n \in S\}$ is the first entry time into the set S. Note that $\mathbb{P}_{x^0}(\tau_S < +\infty) > 0$ is implied by the following property: for any Borel subset $S \subset \mathcal{X}$ of positive Lebesgue measure and any initial condition $x^0 \in \mathcal{X}$, there exists $n \geqslant 0$ (depending on x^0 and S) such that $P^n(x^0, S) > 0$. For discretizations of SDEs, the latter inequality is often easy to establish. It holds, for instance, with $n = 1$ for the Euler–Maruyama scheme (3.8) when σ has full rank d at each point $x \in \mathcal{X}$, since

$$G^1 = \sigma(x^0)^{-1}\left(\frac{x^1 - x^0 - \Delta t\, b(x^0)}{\sqrt{\Delta t}}\right) =: \mathcal{G}_{x^0}(x^1),$$

so that

$$P(x^0, S) = \mathbb{P}(G^1 \in \mathcal{G}_{x_0}^{-1}(S)) > 0.$$

When σ is degenerate, several iterates may be necessary; see for instance Bou-Rabee and Owhadi (2010), where it is proved that $P^2(x^0, S) > 0$ for some discretization of Langevin dynamics.

Once the existence of an invariant probability measure and the irreducibility of the scheme are proved, the almost-sure convergence of long-time averages over one realization already follows (see Chapter 17 in Meyn and Tweedie (2009)): for almost all initial conditions $x^0 \in \mathcal{X}$,

$$\widehat{\varphi}_{N_{\rm iter},\Delta t} \xrightarrow[N_{\rm iter}\to+\infty]{\text{a.s.}} \int_{\mathcal{X}} \varphi \,\mathrm{d}\pi_{\Delta t}.$$

Let us also mention that, when irreducibility holds with respect to the Lebesgue measure, the measure $\pi_{\Delta t}$ is absolutely continuous with respect to the Lebesgue measure. In addition, when the transition kernel $P(x, \mathrm{d}y)$ is absolutely continuous with respect to the Lebesgue measure $\mathrm{d}y$, the above almost sure convergence occurs for all initial conditions: see Corollary 1 in Tierney (1994), based on Nummelin (1984).

As in the continuous case, when decay estimates on the evolution operator $P_{\Delta t}$ hold as in Theorem 2.25, it can easily be shown that the asymptotic variance is well defined when $x^0 \sim \pi_{\Delta t}$, provided $\mathcal{K} \in L^2(\pi_{\Delta t})$ and $\varphi \in L^\infty_{\mathcal{K}}(\mathcal{X})$. A way to prove that $\mathcal{K} \in L^2(\pi_{\Delta t})$ is to check whether the assumptions of Theorem 2.25 are satisfied for the Lyapunov function $\mathcal{K}^2$ instead of $\mathcal{K}$, so that the integrability condition follows from (2.64). In practice, it is often convenient to establish minorization conditions on arbitrary compact sets, and to obtain Lyapunov conditions for the family of functions $\mathcal{K}_n(x) = 1 + |x|^n$ (for integers $n \geqslant n_0$ sufficiently large). Note that $\mathcal{K}_n^2 = \mathcal{K}_{2n} + 2(\mathcal{K}_n - 1)$, so that the $L^2(\pi_{\Delta t})$ integrability of $\mathcal{K}_n$ follows once Lyapunov estimates hold for $\mathcal{K}_{2n}$.

To state the expression of the asymptotic variance when Theorem 2.25 holds for a Lyapunov function $\mathcal{K} \in L^2(\pi_{\Delta t})$, let us introduce the equivalent of the projection operator (3.3) for the invariant measure $\pi_{\Delta t}$, namely

$$\Pi_{\Delta t}\varphi = \varphi - \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi_{\Delta t}.$$

Note that if $x^0 \sim \pi_{\Delta t}$, then $x^n \sim \pi_{\Delta t}$ for all $n \geqslant 1$ and the following stationarity property holds:

$$\mathbb{E}_{\pi_{\Delta t}}\left[\Pi_{\Delta t}\varphi(x^n)\Pi_{\Delta t}\varphi(x^m)\right] = \mathbb{E}_{\pi_{\Delta t}}\left[\Pi_{\Delta t}\varphi(x^{|n-m|})\Pi_{\Delta t}\varphi(x^0)\right]$$

for all $n, m \geqslant 0$. Then,

$$\begin{aligned}
N_{\mathrm{iter}} \operatorname{Var}_{\pi_{\Delta t}}(\widehat{\varphi}_{N_{\mathrm{iter}}}) &= N_{\mathrm{iter}} \, \mathbb{E}_{\pi_{\Delta t}}\left[(\Pi_{\Delta t}\widehat{\varphi}_{N_{\mathrm{iter}}})^2\right] \\
&= \frac{1}{N_{\mathrm{iter}}} \sum_{n,m=0}^{N_{\mathrm{iter}}-1} \mathbb{E}_{\pi_{\Delta t}}\left[\Pi_{\Delta t}\varphi(x^n)\Pi_{\Delta t}\varphi(x^m)\right] \\
&= \frac{1}{N_{\mathrm{iter}}} \sum_{n=0}^{N_{\mathrm{iter}}-1} \mathbb{E}_{\pi_{\Delta t}}\left[(\Pi_{\Delta t}\varphi(x^n))^2\right] \\
&\quad + \frac{2}{N_{\mathrm{iter}}} \sum_{0 \leqslant m < n \leqslant N_{\mathrm{iter}}-1} \mathbb{E}_{\pi_{\Delta t}}\left[\Pi_{\Delta t}\varphi(x^n)\Pi_{\Delta t}\varphi(x^m)\right] \\
&= \mathbb{E}_{\pi_{\Delta t}}\left[(\Pi_{\Delta t}\varphi)^2\right] + 2 \sum_{1 \leqslant n \leqslant N_{\mathrm{iter}}-1} \left(1 - \frac{n}{N_{\mathrm{iter}}}\right) \mathbb{E}_{\pi_{\Delta t}}\left[\Pi_{\Delta t}\varphi(x^n)\Pi_{\Delta t}\varphi(x^0)\right],
\end{aligned}$$

where we used the stationarity property in the last step. Since

$$\mathbb{E}_{\pi_{\Delta t}}\left[\Pi_{\Delta t}\varphi(x^n)\Pi_{\Delta t}\varphi(x^0)\right] = \int_{\mathcal{X}} (\Pi_{\Delta t}\varphi)(P_{\Delta t}^n \Pi_{\Delta t}\varphi) \, \mathrm{d}\pi_{\Delta t},$$

and we assume (2.67) to hold, a dominated convergence argument shows that

$$\lim_{N_{\mathrm{iter}} \to +\infty} N_{\mathrm{iter}} \operatorname{Var}_{\pi_{\Delta t}}(\widehat{\varphi}_{N_{\mathrm{iter}},\Delta t}) = \varsigma^2_{\varphi,\Delta t}$$

with (compare with (3.5))

$$\begin{aligned}
\varsigma^2_{\varphi,\Delta t} &= \mathbb{E}_{\pi_{\Delta t}}\left[(\Pi_{\Delta t}\varphi)^2\right] + 2 \sum_{n=1}^{+\infty} \mathbb{E}_{\pi_{\Delta t}}\left[(\Pi_{\Delta t}\varphi)(x^n)(\Pi_{\Delta t}\varphi)(x^0)\right] \\
&= \int_{\mathcal{X}} (\Pi_{\Delta t}\varphi) \left[\mathrm{Id} + 2 \sum_{n=1}^{+\infty} P_{\Delta t}^n\right] \Pi_{\Delta t}\varphi \, \mathrm{d}\pi_{\Delta t} \\
&= \int_{\mathcal{X}} (\Pi_{\Delta t}\varphi) \left[2(\mathrm{Id} - P_{\Delta t})^{-1} - \mathrm{Id}\right] \Pi_{\Delta t}\varphi \, \mathrm{d}\pi_{\Delta t}. \qquad (3.11)
\end{aligned}$$

The expression for the asymptotic variance suggests a definition of a correlation time, as for continuous dynamics. By the same reasoning as in (3.6), using the variance obtained with i.i.d. samples as a reference,

$$\varsigma^2_{\varphi,\Delta t} = N_{\mathrm{corr},\Delta t,\varphi}\varsigma^2_{\varphi,\mathrm{iid},\Delta t} \quad \text{where } \varsigma^2_{\varphi,\mathrm{iid},\Delta t} = \int_{\mathcal{X}} (\Pi_{\Delta t}\varphi)^2 \, \mathrm{d}\pi_{\Delta t}.$$

The number $N_{\mathrm{corr},\Delta t,\varphi}$ measures the correlation between the samples: intuitively, it gives the number of steps which should be performed to have a new configuration sufficiently decorrelated from the previous one. As discussed below, the correlation time $\theta_{\mathrm{corr},\varphi}$ is related to the number of correlation steps $N_{\mathrm{corr},\Delta t,\varphi}$ for the discrete dynamics since in the limit of small Δt, $\theta_{\mathrm{corr},\varphi} \simeq N_{\mathrm{corr},\Delta t,\varphi}\Delta t$.

Finally, let us mention that, as in the continuous case, a central limit theorem in fact holds true for time averages computed using ergodic Markov chains once the solution Φ of the Poisson equation

$$(\mathrm{Id} - P_{\Delta t})\Phi = \Pi_{\Delta t}\varphi \tag{3.12}$$

belongs to $L^2(\pi_{\Delta t})$; see Theorem 17.4.4 in Meyn and Tweedie (2009). If $\varphi \in L^\infty_{\mathcal{K}}(\mathcal{X})$ and $\mathcal{K} \in L^2(\pi_{\Delta t})$ (see the above discussion for the latter condition), one way to guarantee that $\Phi \in L^2(\pi_{\Delta t})$ is to resort to Corollary 2.26.

3.2.3. Error analysis on time averages: general decomposition

Let us now discuss the topic of main interest for this section, namely error estimates for the computation of thermodynamic averages $\mathbb{E}_\pi(\varphi) = \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi$ with (3.10), for a given one-step discretization of the continuous dynamics (3.1) ergodic with respect to π. We decompose the error as the sum of two contributions:

$$\widehat{\varphi}_{N_{\mathrm{iter}},\Delta t} - \mathbb{E}_\pi(\varphi) = \big(\widehat{\varphi}_{N_{\mathrm{iter}},\Delta t} - \mathbb{E}_{\pi_{\Delta t}}(\varphi)\big) + \big(\mathbb{E}_{\pi_{\Delta t}}(\varphi) - \mathbb{E}_\pi(\varphi)\big). \tag{3.13}$$

The first term is a statistical error arising from the finiteness of the number of time steps N_{iter}, while the second term is a systematic error (or bias), which persists in the limit $N_{\mathrm{iter}} \to +\infty$, and is due to the use of finite time steps $\Delta t > 0$. Let us discuss each term more precisely.

(i) According to the central limit theorem for Markov chains (which holds when the Poisson equation (3.12) can be solved in $L^2(\pi_{\Delta t})$: see the above discussion), the statistical error behaves in the limit $N_{\mathrm{iter}} \to +\infty$ as a Gaussian random variable with asymptotic variance given by (3.11). In practice, this asymptotic regime is attained when $N_{\mathrm{iter}} \gg N_{\mathrm{corr},\Delta t,\varphi}$. However, $N_{\mathrm{corr},\Delta t,\varphi}$ is often very large because of the metastability of the underlying continuous dynamics (so that $\theta_{\mathrm{corr},\varphi}$ is large), which makes it difficult in practice to ensure that the central limit theorem actually holds for the values of N_{iter} achievable with computer simulations.

When the asymptotic regime can be considered to have been attained, $\widehat{\varphi}_{N_{\rm iter},\Delta t} - \mathbb{E}_{\pi_{\Delta t}}(\varphi)$ is of order

$$\frac{\varsigma_{\varphi,\Delta t}}{\sqrt{N_{\rm iter}}} = \frac{\varsigma_{\varphi,\Delta t}\sqrt{\Delta t}}{\sqrt{T}}.$$

This reformulation highlights the fact that the statistical error is of the order of the inverse of the square root of the 'physical' simulation time $T = N_{\rm iter}\Delta t$. Indeed, weakly consistent discretizations of SDEs are such that

$$\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\varphi = -\mathcal{L}\varphi + \mathrm{O}(\Delta t).$$

In addition, the variance $\varsigma^2_{\varphi,\Delta t}$ can be rewritten as

$$\Delta t\, \varsigma^2_{\varphi,\Delta t} = 2\int_{\mathcal{X}} (\Pi_{\Delta t}\varphi)\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)^{-1} \Pi_{\Delta t}\varphi \,\mathrm{d}\pi_{\Delta t} + \mathrm{O}(\Delta t).$$

In view of the expression ς^2_φ of the asymptotic variance (3.5) of the continuous dynamics, this suggests the following convergence result:

$$\lim_{\Delta t \to 0} \Delta t\, \varsigma^2_{\varphi,\Delta t} = \varsigma^2_\varphi. \tag{3.14}$$

A rigorous proof of this convergence is provided by Theorem 5.6. In fact, it typically holds that $\theta_{{\rm corr},\varphi} = N_{{\rm corr},\Delta t,\varphi}\Delta t + \mathrm{O}(\Delta t)$.

The interpretation of (3.14) is that the asymptotic variance of time averages computed using numerical methods is, to first order in Δt, related to the asymptotic variance of the time averages computed with the continuous process. This motivates direct study of the variance (3.5) rather than its discrete counterpart, which we therefore do in Section 3.4.

(ii) The second term in (3.13) is a systematic error (or bias) related to the fact that the invariant measure of the numerical scheme $\pi_{\Delta t}$ is different from the canonical measure π. Although this is not obvious, the expression of the formal correction function $h_{\Delta t}$ defined as $\pi_{\Delta t} = h_{\Delta t}\pi$ is encoded in the asymptotic expansion of the one-step evolution operator $P_{\Delta t}$, provided some ergodicity conditions are met. A typical result is that $h_{\Delta t} = 1 + \mathrm{O}(\Delta t^p)$ for some integer p. In fact, it is often possible to make the leading term in the bias precise as follows (see Theorem 3.3 below):

$$\int_{\mathcal{X}} \varphi \,\mathrm{d}\pi_{\Delta t} = \int_{\mathcal{X}} \varphi \,\mathrm{d}\pi + \Delta t^p \int_{\mathcal{X}} \varphi f \,\mathrm{d}\pi + \mathrm{O}(\Delta t^{p+1})$$

for some function f to be made precise. As made clear in Theorem 3.3, $p \geqslant \alpha$, where α is the weak order of the method (in the sense that $P_{\Delta t} = \mathrm{e}^{\Delta t\mathcal{L}} + \mathrm{O}(\Delta t^{\alpha+1})$). In some cases, it even holds that $p \geqslant \alpha + 1$.

An important remark is that the bias becomes noticeable only for sufficiently long integration times $T = N_{\text{iter}}\Delta t$, namely when $\varsigma_\varphi/\sqrt{T} \sim \Delta t^p$. More precisely,

$$N_{\text{iter}} \sim \frac{\varsigma_\varphi^2}{\Delta t^{2p+1}}.$$

This motivates the fact that the statistical error often dominates in actual simulations. However, it is desirable to have biases as small as possible in order to use larger time steps (while still satisfying stability constraints). We now analyse the bias in the next section.

3.3. Analysis of the bias

In this section we present the analysis of the bias (second term on the right-hand side of (3.13)). We start with a presentation of the general strategy, and then apply it to the estimation of canonical averages with discretizations of overdamped Langevin or Langevin dynamics.

3.3.1. General strategy

Assume that the continuous dynamics, with generator $\mathcal{L}$, admits a unique invariant measure π. Consider next a given numerical scheme, described in terms of its discrete evolution operator $P_{\Delta t}$ defined in (3.9), and admitting an invariant measure $\pi_{\Delta t}$ (we do not make any uniqueness assumption on $\pi_{\Delta t}$ at this stage).

Assume that, for a C^∞ function φ and a given integer p, the evolution operator can be expanded as

$$P_{\Delta t}\varphi = \varphi + \Delta t\, \mathcal{A}_1\varphi + \Delta t^2 \mathcal{A}_2\varphi + \cdots + \Delta t^{p+1}\mathcal{A}_{p+1}\varphi + \Delta t^{p+2} r_{\varphi,\Delta t}, \quad (3.15)$$

for some remainder term $r_{\varphi,\Delta t}$ (which depends on p, although we henceforth omit this dependence). The operators $\mathcal{A}_k$ are identified in practice by Taylor expansions of $\varphi(x^{n+1})$ around $\varphi(x^n)$, with a remainder term typically expressed as an integral. Note that, by consistency of the discretization, it typically holds that

$$\mathcal{A}_1 = \mathcal{L},$$

although in principle it is possible to construct numerical schemes approximating π for which this is not the case. More generally, the method is of weak order p when $\mathcal{A}_k = \mathcal{L}^k/k!$ for $1 \leqslant k \leqslant p$. However, there may be discrete dynamics for which $\mathcal{A}_k = a_k\mathcal{L}^k$ with a prefactor $a_k \neq 1/k!$ for some $1 \leqslant k \leqslant p$. In this case the dynamics is not of weak order p, but the invariant measure may nonetheless be correct up to error terms of order Δt^{p+1}. Moreover, as made precise in Theorem 3.3, there are situations in which the operators $\mathcal{A}_k$ are different from $\mathcal{L}^k$ but the invariant measure $\pi_{\Delta t}$ is still close to π; see for instance Abdulle, Vilmart and Zygalakis (2015) in the context of Langevin dynamics.

We next need some functional estimates on the operator $\mathcal{A}_1$ appearing in (3.15). To this end, we introduce the following set of C^∞ functions.

Definition 3.1 (smooth functions). Consider scale functions $\mathcal{K}_n : \mathcal{X} \to [1, +\infty)$ such that

$$\mathcal{K}_n \leqslant \mathcal{K}_{n+1}, \quad \text{for all } n \geqslant 0.$$

The space $\mathcal{S}$ is composed of all functions $\varphi \in C^\infty(\mathcal{X})$ for which, for any $k \in \mathbb{N}^d$, there exists $m \in \mathbb{N}$ such that $\partial^k \varphi \in L^\infty_{\mathcal{K}_m}(\mathcal{X})$.

In the simple case when $\mathcal{X}$ is bounded, it is possible to choose $\mathcal{K}_n = \mathbf{1}$ for all $n \geqslant 0$, in which case $\mathcal{S} = C^\infty(\mathcal{X})$. For unbounded spaces, a typical choice is $\mathcal{K}_n(x) = 1 + |x|^n$. In this case, the above definition means that the functions in $\mathcal{S}$, as well as all their derivatives, grow at most polynomially. The set $\mathcal{S}$ can then be shown to be dense in $L^2(\pi)$ when π has moments of all orders, so that all the operators arising in the error estimations can be considered as operators on $L^2(\pi)$, defined on the core $\mathcal{S}$.

In order to state the regularity result we rely on, we restrict the space of smooth functions to those with average 0 with respect to the invariant measure π of the continuous process, namely

$$\mathcal{S}_0 = \Pi\mathcal{S} = \left\{ \varphi \in \mathcal{S} \;\middle|\; \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi = 0 \right\}, \tag{3.16}$$

where Π is defined in (3.3). We next consider the following assumption on the generator of the continuous dynamics. Recall that operators are considered on $L^2(\pi)$ and that adjoints are taken with respect to the corresponding scalar product.

Assumption 3.2 (stability of smooth functions by inverse operators). The space $\mathcal{S}$ is dense in $L^2(\pi)$ (in particular $\mathcal{K}_n \in L^2(\pi)$ for any $n \geqslant 0$) and the operators $\mathcal{A}_1^{-1}$ and $(\mathcal{A}_1^*)^{-1}$ leave $\mathcal{S}_0$ invariant.

Let us recall that $\mathcal{A}_1^*$ denotes the adjoint of $\mathcal{A}_1$ with respect to the scalar product in $L^2(\mu)$ (see (2.7) and Section 2.3.2) and that typically $\mathcal{A}_1 = \mathcal{L}$. The invariance of $\mathcal{S}_0$ by an operator $\mathcal{T}^{-1}$ has to be understood in the following sense: when $\varphi \in \mathcal{S}_0$, the unique solution Φ of the equation $\mathcal{T}\Phi = \varphi$ belongs to $\mathcal{S}_0$.

Error estimates on averages of smooth functions can finally be stated as follows.

Theorem 3.3 (error estimates on the invariant measure). Suppose that Assumption 3.2 is satisfied, and that an expansion such as (3.15) holds for any $\varphi \in \mathcal{S}$ and a given $p \in \mathbb{N}$, with a remainder $r_{\varphi,\Delta t}$ for which there exist $K \geqslant 0$, $m \in \mathbb{N}$ and $\Delta t^* > 0$ (all depending on φ and p) such that

$$\|r_{\varphi,\Delta t}\|_{L^\infty_{\mathcal{K}_m}} \leqslant K, \quad \text{for all } \Delta t \leqslant \Delta t^*.$$

Assume in addition that the operators $\mathcal{A}_k$ leave $\mathcal{S}$ invariant for any $k \geqslant 1$, that

$$\int_{\mathcal{X}} \mathcal{A}_k \varphi \, \mathrm{d}\pi = 0, \quad \text{for all } k \in \{1, \ldots, p\}, \tag{3.17}$$

and that $g_{p+1} = \mathcal{A}_{p+1}^* \mathbf{1} \in \mathcal{S}_0$. Finally, assume that the numerical scheme admits an invariant measure $\pi_{\Delta t}$ which integrates all scale functions:

$$\int_{\mathcal{X}} \mathcal{K}_n \, \mathrm{d}\pi_{\Delta t} < +\infty, \quad \text{for all } n \geqslant 0.$$

Then, there exists $L > 0$ such that, for any $\Delta t \in (0, \Delta t^*]$,

$$\int_{\mathcal{X}} \varphi \, \mathrm{d}\pi_{\Delta t} = \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi + \Delta t^p \int_{\mathcal{X}} \varphi f_{p+1} \, \mathrm{d}\pi + \Delta t^{p+1} R_{\varphi, \Delta t}, \tag{3.18}$$

with $|R_{\varphi,\Delta t}| \leqslant L$ and where

$$f_{p+1} = -(\mathcal{A}_1^*)^{-1} g_{p+1} \in \mathcal{S}_0. \tag{3.19}$$

Let us first comment on the assumptions of the theorem. The condition that the operators $\mathcal{A}_k$ leave $\mathcal{S}$ invariant is usually very easy to check since these operators are typically differential operators with C^∞ coefficients when b and σ in (3.1) are C^∞. Moreover, in order to obtain the expression of $g_{p+1} = \mathcal{A}_{p+1}^* \mathbf{1}$, it is convenient to resort to integration by parts to determine the function g_{p+1} such that, for any test function $\varphi \in \mathcal{S}$,

$$\int_{\mathcal{X}} \mathcal{A}_{p+1} \varphi \, \mathrm{d}\pi = \int_{\mathcal{X}} g_{p+1} \varphi \, \mathrm{d}\pi. \tag{3.20}$$

It can usually be checked by direct inspection that $g_{p+1} \in \mathcal{S}$. In addition, by considering $\varphi = \mathbf{1}$ in (3.20), it follows that g_{p+1} automatically has average 0 with respect to π when $\mathcal{A}_{p+1}\mathbf{1} = 0$, which is the case for differential operators. Finally, let us emphasize once again that the important condition which determines the order of the error is (3.17). This condition holds when $\mathcal{A}_k$ is proportional to $\mathcal{L}^k$, but can be satisfied for more general operators.

The interpretation of Theorem 3.3 is as follows. First, averages with respect to $\pi_{\Delta t}$ are correct up to errors of order Δt^p. Second, we can give an explicit expression of the leading-order term in the error, which can then be eliminated either by Romberg extrapolation or by a numerical estimate (see (3.32) below). Let us finally mention that it is possible to obtain bounds on $R_{\varphi,\Delta t}$ in terms of weighted L^∞-norms of φ and a given number of its derivatives when more precise estimates for $\mathcal{L}^{-1}$ are available.

Proof. The proof is obtained by a generalization of the proof of Theorem 2.13 in Leimkuhler *et al.* (2015). A similar result is provided in Abdulle *et al.* (2015). Note first that, since f_{p+1} has average 0 with respect to π, it is sufficient to establish (3.18) for functions with average 0 with respect to π, upon considering $\varphi - \mathbb{E}_\pi(\varphi)$.

The first step is to prove (3.18) for functions $\varphi \in (P_{\Delta t} - \mathrm{Id})\mathcal{S}$. This step motivates the expression for the correction function f_{p+1}. Consider $\phi \in \mathcal{S}$. The invariance of $\pi_{\Delta t}$ by the discretized dynamics implies that

$$\int_{\mathcal{X}} \left[\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t} \right) \phi \right] \mathrm{d}\pi_{\Delta t} = 0. \tag{3.21}$$

On the other hand, (3.15) and (3.17) give

$$\int_{\mathcal{X}} \left[\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t} \right) \phi \right] \mathrm{d}\pi = \Delta t^p \int_{\mathcal{X}} \mathcal{A}_{p+1}\phi \, \mathrm{d}\pi + \Delta t^{p+1} \int_{\mathcal{X}} r_{\phi,\Delta t} \, \mathrm{d}\pi,$$

and, for a given function $f \in \mathcal{S}$,

$$\int_{\mathcal{X}} \left[\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t} \right) \phi \right] f \, \mathrm{d}\pi = \int_{\mathcal{X}} (\mathcal{A}_1\phi) f \, \mathrm{d}\pi + \Delta t \int_{\mathcal{X}} \left[\frac{P_{\Delta t} - \mathrm{Id} - \Delta t \, \mathcal{A}_1}{\Delta t^2} \phi \right] f \, \mathrm{d}\pi,$$

so that

$$\begin{aligned} &\int_{\mathcal{X}} \left[\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t} \right) \phi \right] (1 + \Delta t^p f) \, \mathrm{d}\pi \\ &\quad = \Delta t^p \int_{\mathcal{X}} (\mathcal{A}_{p+1}\phi + (\mathcal{A}_1\phi) f) \, \mathrm{d}\pi \\ &\qquad + \Delta t^{p+1} \int_{\mathcal{X}} \left(r_{\phi,\Delta t} + \left[\frac{P_{\Delta t} - \mathrm{Id} - \Delta t \, \mathcal{A}_1}{\Delta t^2} \phi \right] f \right) \mathrm{d}\pi. \end{aligned}$$

The second term on the right-hand side of the previous equality is indeed of order Δt^{p+1} since the integrand of the corresponding integral is bounded in some weighted L^∞-space. In order to choose f so that the first term on the right-hand side vanishes for all test functions $\phi \in \mathcal{S}$, we rewrite it as

$$\int_{\mathcal{X}} (\mathcal{A}_{p+1}\phi + (\mathcal{A}_1\phi) f) \, \mathrm{d}\pi = \int_{\mathcal{X}} (g_{p+1} + \mathcal{A}_1^* f) \phi \, \mathrm{d}\pi.$$

This suggests choosing $f = f_{p+1} = -(\mathcal{A}_1^*)^{-1} g_{p+1}$, which is well defined by our assumptions on $\mathcal{A}_1$ since $g_{p+1} \in \mathcal{S}_0$. With this choice,

$$\begin{aligned} &\int_{\mathcal{X}} \left[\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t} \right) \phi \right] (1 + \Delta t^p f_{p+1}) \, \mathrm{d}\pi \\ &\quad = \Delta t^{p+1} \int_{\mathcal{X}} \left(r_{\phi,\Delta t} + \left[\frac{P_{\Delta t} - \mathrm{Id} - \Delta t \, \mathcal{A}_1}{\Delta t^2} \phi \right] f_{p+1} \right) \mathrm{d}\pi. \end{aligned} \tag{3.22}$$

so that (3.18) holds for $\varphi = (\mathrm{Id} - P_{\Delta t})\phi / \Delta t$.

The second step of the proof is to extend (3.18) to all functions in $\mathcal{S}_0$. Of course, we would like to replace ϕ with $\Delta t(\mathrm{Id} - P_{\Delta t})^{-1}\varphi$ in the previous estimates. However, there are two obstructions to this approach: (i) the inverse $\Delta t(\mathrm{Id} - P_{\Delta t})^{-1}$ is well defined only on spaces of functions with average 0 with respect to $\pi_{\Delta t}$, and (ii), when this is the case, we typically do not have any control on the derivatives of $\Delta t(\mathrm{Id} - P_{\Delta t})^{-1}\varphi$, but only on

the function itself (by results such as Corollary 2.26). Our strategy is to construct an operator $Q_{\Delta t}$ which leaves $\mathcal{S}_0$ invariant and is an approximate inverse of $(\mathrm{Id} - P_{\Delta t})/\Delta t$ on $\mathcal{S}_0$. Since $(\mathrm{Id} - P_{\Delta t})\varphi/\Delta t = \mathcal{A}_1\varphi + \mathrm{O}(\Delta t)$, we expect the inverse operator to be $\mathcal{A}_1^{-1}$ at dominant order in Δt.

The first task is to restrict $\mathrm{Id} - P_{\Delta t}$ to $\mathcal{S}_0$ via $\Pi(\mathrm{Id} - P_{\Delta t})\Pi$, using the projection operator Π defined in (3.3). Then (3.21) implies

$$\int_{\mathcal{X}} \left[\Pi\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\right)\Pi\phi\right] \mathrm{d}\pi_{\Delta t} = -\frac{1}{\Delta t}\int_{\mathcal{X}} (P_{\Delta t}\phi)\,\mathrm{d}\pi, \tag{3.23}$$

for $\phi \in \mathcal{S}_0$ (so that $\Pi\phi = \phi$). Further, using the fact that f_{p+1} is of average 0 with respect to π, (3.22) leads to

$$\begin{aligned}\int_{\mathcal{X}} \left[\Pi\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\right)\Pi\phi\right](1 + \Delta t^p f_{p+1})\,\mathrm{d}\pi &= -\frac{1}{\Delta t}\int_{\mathcal{X}} (P_{\Delta t}\phi)\,\mathrm{d}\pi \\ &+ \Delta t^{p+1}\int_{\mathcal{X}} \left(r_{\phi,\Delta t} + \left[\frac{P_{\Delta t} - \mathrm{Id} - \Delta t\,\mathcal{A}_1}{\Delta t^2}\phi\right] f_{p+1}\right)\mathrm{d}\pi. \end{aligned} \tag{3.24}$$

This shows that, for any $\phi \in \mathcal{S}_0$,

$$\begin{aligned}\int_{\mathcal{X}} \left[\Pi\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\right)\Pi\phi\right] \mathrm{d}\pi_{\Delta t} &= \int_{\mathcal{X}} \left[\Pi\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\right)\Pi\phi\right](1 + \Delta t^p f_{p+1})\,\mathrm{d}\pi \\ &- \Delta t^{p+1}\int_{\mathcal{X}} \left(r_{\phi,\Delta t} + \left[\frac{P_{\Delta t} - \mathrm{Id} - \Delta t\,\mathcal{A}_1}{\Delta t^2}\phi\right] f_{p+1}\right)\mathrm{d}\pi.\end{aligned}$$

We next consider an approximate inverse operator $Q_{\Delta t}$ which leaves $\mathcal{S}_0$ invariant, and is such that

$$\Pi\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\right)\Pi Q_{\Delta t}\varphi = \varphi + \Delta t^{p+1}\widetilde{r}_{\varphi,\Delta t}, \quad \text{for all } \varphi \in \mathcal{S}_0, \tag{3.25}$$

with $\|\widetilde{r}_{\varphi,\Delta t}\|_{L^\infty_{\mathcal{K}_m}}$ uniformly bounded with respect to Δt for some integer m. Such operators are constructed as follows. The fundamental idea is to truncate the formal series expansion of the inverse of the operator $A + \Delta t\,B = A(\mathrm{Id} + \Delta t\,A^{-1}B)$ in powers of $A^{-1}B$:

$$A^{-1} - \Delta t\,A^{-1}BA^{-1} + \Delta t^2\,A^{-1}BA^{-1}BA^{-1} + \cdots.$$

In the present situation, we set $A = \Pi\mathcal{A}_1\Pi$ and $B = \Pi\mathcal{A}_2\Pi + \cdots + \Delta t^{p-1}\Pi\mathcal{A}_{p+1}\Pi$. Note that perturbative arguments cannot be used to make sense of the formal infinite series since B usually involves differential operators of higher order than A, so that B cannot be controlled by A (in contrast to the situation we will encounter later on in Theorem 5.2). Introducing $\widetilde{\mathcal{A}}_k = \Pi\mathcal{A}_k\Pi$ for notational simplicity, we consider

$$\widetilde{Q}_{\Delta t} = \widetilde{\mathcal{A}}_1^{-1}\sum_{n=0}^{p}(-1)^n\Delta t^n\,(B\widetilde{\mathcal{A}}_1^{-1})^n.$$

We next remove operators with powers of Δt larger than or equal to $p+1$ (arising from the higher-order terms in the expression of $B\widetilde{\mathcal{A}}_1^{-1}$) in order to write down an expression for $Q_{\Delta t}$ involving only powers of Δt smaller than or equal to p. Finally, the so-constructed operator,

$$Q_{\Delta t} = \widetilde{\mathcal{A}}_1^{-1} - \Delta t \widetilde{\mathcal{A}}_1^{-1}\widetilde{\mathcal{A}}_2\widetilde{\mathcal{A}}_1^{-1} + \Delta t^2(\widetilde{\mathcal{A}}_1^{-1}\widetilde{\mathcal{A}}_2\widetilde{\mathcal{A}}_1^{-1}\widetilde{\mathcal{A}}_2\widetilde{\mathcal{A}}_1^{-1} - \widetilde{\mathcal{A}}_1^{-1}\widetilde{\mathcal{A}}_3\widetilde{\mathcal{A}}_1^{-1}) + \Delta t^3\mathcal{Q}_3 + \cdots + \Delta t^p\mathcal{Q}_p,$$

is a well defined operator which leaves $\mathcal{S}_0$ invariant (since its action is given by the application of at most p operations of the form $\widetilde{\mathcal{A}}_k\widetilde{\mathcal{A}}_1^{-1}$, and a final application of $\widetilde{\mathcal{A}}_1^{-1}$) and is such that (3.25) holds true.

In order to conclude, it remains to replace ϕ with $Q_{\Delta t}\varphi$, and gather all the higher-order terms in $R_{\varphi,\Delta t}$. This gives the desired equality (3.18) for functions in $\mathcal{S}_0$. □

We next illustrate the strategy presented here applied to overdamped Langevin and Langevin dynamics. We postpone numerical illustration of the results of Theorem 3.3 to Figure 5.1 in Section 5.3. Indeed, as explained in Section 3.3.4, the leading-order correction term can be computed as some integrated correlation function, for which we provide elements of numerical analysis later on, in Section 5.3.

3.3.2. Application to overdamped Langevin dynamics

For overdamped Langevin dynamics, we consider for simplicity the case of a simple Euler–Maruyama discretization of a dynamics defined on the compact position space $\mathcal{D} = (L\mathbb{T})^d$:

$$q^{n+1} = q^n - \Delta t\, \nabla V(q^n) + \sqrt{\frac{2\Delta t}{\beta}}\, G^n. \tag{3.26}$$

The results below can be easily generalized to other discretization schemes. On the other hand, an extension to unbounded position spaces is not trivial (see Kopec 2015), although several results can be stated in a weaker form than the results given below.

The first task is to find an expansion of the transition operator $P_{\Delta t}$ defined in (3.9) in powers of Δt: see (3.15). Simple but lengthy computations show that (see Section 4.9.2 in Leimkuhler *et al.* 2015)

$$P_{\Delta t}\varphi = \varphi + \Delta t\, \mathcal{L}\varphi + \Delta t^2 \mathcal{A}_2\varphi + \Delta t^3 r_{\varphi,\Delta t}, \tag{3.27}$$

where $\|r_{\varphi,\Delta t}\|_{L^\infty} \leqslant K$ for Δt sufficiently small and

$$\mathcal{A}_2 = \frac{1}{2}(\mathcal{L}^2 + \mathcal{R}_2), \tag{3.28}$$

with

$$\mathcal{R}_2\varphi = \frac{2}{\beta}\nabla^2 V : \nabla^2\varphi + \frac{1}{\beta}\nabla(\Delta V)\cdot\nabla\varphi - \nabla V^T(\nabla^2 V)\nabla\varphi.$$

We want to apply Theorem 3.3 with $p = 1$. In summary, there are three types of assumptions to be checked:

(i) assumptions on the generator $\mathcal{L} = \mathcal{A}_1$ of the continuous dynamics;

(ii) assumptions on the operator $\mathcal{A}_2$, whose action is given by (3.28);

(iii) assumptions on the invariant measure of the numerical scheme.

First, recall that, as discussed after Definition 3.1, $\mathcal{K}_n(q) = \mathbf{1}$ for all $n \geqslant 0$ when $\mathcal{D}$ is bounded, so that $\mathcal{S} = C^\infty(\mathcal{D})$. Standard results of elliptic regularity then show that $\mathcal{L}^{-1}$ is a well-defined operator from

$$\mathcal{S}_0 = \left\{ f \in C^\infty(\mathcal{D}) \,\middle|\, \int_{\mathcal{D}} f \,\mathrm{d}\nu = 0 \right\}$$

to itself. For dynamics in the full space $\mathcal{D} = \mathbb{R}^d$, additional assumptions on the potential are needed to obtain the stability of $\mathcal{S}$ with the choice $\mathcal{K}_n(q) = 1 + |q|^n$, and the proof of this stability result is much more involved (see Kopec 2015).

It can now be easily checked that g_2 is well defined and belongs to $\mathcal{S}$ (recall that, as discussed after Theorem 3.3, it automatically has average 0 with respect to π). To obtain the precise expression for g_2, we use integration by parts to compute

$$\begin{aligned}\int_{\mathcal{D}} \mathcal{R}_2\varphi \,\mathrm{d}\nu &= -\frac{1}{\beta}\int_{\mathcal{D}} \nabla\left(\Delta V - \frac{\beta}{2}|\nabla V|^2\right)\cdot\nabla\varphi \,\mathrm{d}\nu \\ &= \int_{\mathcal{D}} \mathcal{L}\left(\Delta V - \frac{\beta}{2}|\nabla V|^2\right)\varphi \,\mathrm{d}\nu.\end{aligned}$$

Using that, for all $\varphi \in \mathcal{S}$,

$$\int_{\mathcal{D}} \mathcal{A}_2\varphi \,\mathrm{d}\nu = \frac{1}{2}\int_{\mathcal{D}} \mathcal{R}_2\varphi \,\mathrm{d}\nu = \int_{\mathcal{D}} g_2\varphi \,\mathrm{d}\nu,$$

one obtains

$$g_2 = \frac{1}{2}\mathcal{L}\left(\Delta V - \frac{\beta}{2}|\nabla V|^2\right).$$

Since $\mathcal{A}_1 = \mathcal{L} = \mathcal{L}^*$ on $L^2(\nu)$, we in fact obtain the analytical expression of the correction function f_2 defined in (3.19):

$$f_2 = -\frac{1}{2}\left(\Delta V - \frac{\beta}{2}|\nabla V|^2 - \frac{a_{\beta,V}}{2}\right),$$

where

$$a_{\beta,V} = \int_{\mathcal{D}} \Delta V \, \mathrm{d}\nu = \beta \int_{\mathcal{D}} |\nabla V|^2 \, \mathrm{d}\nu$$

is a constant ensuring that f_2 is of average 0 with respect to ν.

Finally, to prove that an invariant measure $\pi_{\Delta t}$ exists for the numerical scheme, we rely on Theorem 2.25. The Lyapunov condition of Assumption 2.23 is trivially satisfied since the position space is compact. Therefore, the only property to prove is the minorization condition of Assumption 2.24. Since the space is bounded, it readily follows from (3.26) that, for any Borel subset $S \subset \mathcal{D}$,

$$\begin{aligned}
\mathbb{P}(q^1 \in S | q^0 = q) &\geqslant \mathbb{P}\left(\sqrt{\frac{2\Delta t}{\beta}} G^1 \in S - q + \Delta t \nabla V(q) \right) \\
&\geqslant \inf_{|\mathcal{Q}| \leqslant |\mathcal{D}| + \|\nabla V\|_{L^\infty}} \mathbb{P}\left(\sqrt{\frac{2\Delta t}{\beta}} G^1 \in S - \mathcal{Q} \right) \\
&= \left(\frac{\beta}{4\pi \Delta t} \right)^{d/2} \inf_{|\mathcal{Q}| \leqslant |\mathcal{D}| + \|\nabla V\|_{L^\infty}} \int_{S-\mathcal{Q}} \exp\left(-\frac{\beta |g|^2}{4\Delta t} \right) \mathrm{d}g,
\end{aligned}$$

where the first inequality is due to the fact that the contributions associated with periodic images of q are not taken into account. The minorization condition then follows by defining the measure

$$\widetilde{\lambda}_{\Delta t}(S) = \inf_{|\mathcal{Q}| \leqslant |\mathcal{D}| + \|\nabla V\|_{L^\infty}} \int_{S-\mathcal{Q}} \exp\left(-\frac{\beta |g|^2}{4\Delta t} \right) \mathrm{d}g,$$

normalizing it as $\lambda_{\Delta t}(S) = \widetilde{\lambda}_{\Delta t}(S) / \widetilde{\lambda}_{\Delta t}(\mathcal{D})$, and introducing

$$\eta_{\Delta t} = (4\pi \Delta t / \beta)^{-d/2} \widetilde{\lambda}_{\Delta t}(\mathcal{D}).$$

We state below a minorization condition stronger than is needed for Assumption 2.24, but which will prove useful later on in Section 5.3. We call it a 'uniform minorization condition' since, in contrast to the above computations, the lower bound η and the probability measure λ do not depend on the time step Δt provided it is sufficiently small. To obtain such a condition, we have to consider evolutions over fixed physical times $T \simeq n\Delta t > 0$, which amounts to iterating the elementary evolution $P_{\Delta t}$ over $\lceil T/\Delta t \rceil$ time steps (where $\lceil x \rceil$ denotes the smallest integer larger than x).

Lemma 3.4 (uniform minorization condition). Consider the evolution operator $P_{\Delta t}$ associated with the Euler–Maruyama discretization (3.26) on the position space $\mathcal{D} = (L\mathbb{T})^d$, and a given integration time $T > 0$. There exist $\Delta t^*, \eta > 0$ and a probability measure λ such that, for any bounded,

measurable non-negative function f, for any $\Delta t \in (0, \Delta t^*]$ and for any $q \in \mathcal{D}$,

$$(P_{\Delta t}^{\lceil T/\Delta t \rceil} f)(q) \geqslant \eta \int_{\mathcal{D}} f \, \mathrm{d}\lambda.$$

Such estimates were obtained in Bou-Rabee and Hairer (2013) in unbounded spaces $\mathcal{D}$ for a class of Metropolis–Hastings schemes whose proposition kernel is (3.26). See also Fathi, Homman and Stoltz (2014) and Fathi and Stoltz (2015) for related results on bounded spaces, as well as Leimkuhler *et al.* (2015) and Redon, Stoltz and Trstanova (2016) for discretizations of Langevin dynamics.

Proof. It is sufficient to prove the result for indicator functions of Borel sets $S \subset \mathcal{D}$. We therefore aim to prove that

$$\mathbb{P}(q^{\lceil T/\Delta t \rceil} \in S \mid q^0 = q) \geqslant \eta \, \lambda(S),$$

for a probability measure λ and a constant $\eta > 0$. The idea of the proof is to explicitly rewrite q^n as a perturbation of the reference evolution corresponding to $\nabla V = 0$. Since we consider C^∞ potentials and the position space is compact, the perturbation can be uniformly controlled. More precisely, an immediate induction argument based on (3.26) gives

$$q^n = q^0 + \mathcal{G}^n + \mathcal{F}^n,$$

with

$$\mathcal{G}^n = \sqrt{\frac{2\Delta t}{\beta}} \sum_{k=0}^{n-1} G^k, \quad \mathcal{F}^n = -\Delta t \sum_{k=0}^{n-1} \nabla V(q^k).$$

Let us now fix $n = \lceil T/\Delta t \rceil$ (so that $T \leqslant n\Delta t \leqslant T + \Delta t$). Then $|\mathcal{F}^n| \leqslant \|\nabla V\|_{L^\infty}(T + \Delta t)$, while $\mathcal{G}^n$ is a Gaussian random variable with covariance matrix $2n\Delta t\beta^{-1} \, \mathrm{Id}_d$. Therefore,

$$\begin{aligned} \mathbb{P}(q^n \in S \mid q^0 = q) &\geqslant \mathbb{P}(\mathcal{G}^n \in S - q - \mathcal{F}^n) \\ &= \left(\frac{\beta}{2\pi n \Delta t}\right)^{d/2} \int_{S-q-\mathcal{F}^n} \exp\left(-\frac{\beta |g|^2}{2n\Delta t}\right) \mathrm{d}g. \end{aligned}$$

In the latter expression and below, we assume that the random variable $\mathcal{G}^n$ has values in $\mathbb{R}^d$ rather than $\mathcal{D}$, and regard $S - q - \mathcal{F}^n$ as a subset of $\mathbb{R}^d$ rather than $\mathcal{D}$. This amounts to neglecting the periodic images, and henceforth reduces the probability on the right-hand side of the above inequality. This is not a problem, however, since we seek a lower bound. Now, for Δt sufficiently small, we have $T/2 \leqslant n\Delta t \leqslant 2T$, so

$$\left(\frac{1}{n\Delta t}\right)^{d/2} \int_{S-q-\mathcal{F}^n} \exp\left(-\frac{\beta |g|^2}{2n\Delta t}\right) \mathrm{d}g \geqslant \left(\frac{1}{2T}\right)^{d/2} \int_{S-q-\mathcal{F}^n} \exp\left(-\frac{\beta |g|^2}{T}\right) \mathrm{d}g.$$

Since the state space is compact, there exists $R > 0$ such that $|q + \mathcal{F}^n| \leqslant R$ for any $q \in \mathcal{D}$. The conclusion then easily follows by defining the probability measure λ as

$$\lambda(S) = Z_R^{-1} \inf_{|Q| \leqslant R} \int_{S+Q} \exp\left(-\frac{\beta |g|^2}{T}\right) \mathrm{d}g,$$

for all $S \subset \mathcal{D}$, and setting $\eta = Z_R (4\pi T/\beta)^{-d/2}$. □

The proof of Lemma 3.4 shows that the measure λ has a positive density with respect to the Lebesgue measure. Application of Theorem 2.25, together with an argument similar to that used to obtain (2.78) from the decay at multiples of a given time, gives the following exponential convergence.

Corollary 3.5 (ergodicity of Euler–Maruyama for compact spaces). There exists $\Delta t^* > 0$ such that, for any $\Delta t \in (0, \Delta t^*]$, the Markov chain associated with $P_{\Delta t}$ has a unique invariant probability measure $\nu_{\Delta t}$. This measure admits a density with respect to the Lebesgue measure. Moreover, there exist $C, \kappa > 0$ such that, for all functions $f \in L^\infty(\mathcal{D})$,

$$\left\| P_{\Delta t}^n f - \int_{\mathcal{D}} f \, \mathrm{d}\nu_{\Delta t} \right\|_{L^\infty} \leqslant C \, \mathrm{e}^{-\kappa n \Delta t} \, \|f\|_{L^\infty}, \quad \text{for all } n \in \mathbb{N}. \tag{3.29}$$

The fact that $\nu_{\Delta t}$ admits a density with respect to Lebesgue measure $\mathrm{d}q$ follows from the minorization condition stated in Lemma 3.4, which ensures that the Markov chain is irreducible with respect to Lebesgue measure.

3.3.3. Application to Langevin dynamics

Many numerical schemes have been proposed for Langevin dynamics. The first ones, such as the BBK scheme proposed in Brünger, Brooks and Karplus (1984), were often obtained by modifications of numerical schemes for deterministic Hamiltonian dynamics. Various other schemes have been proposed since then; see for instance Skeel and Izaguirre (2002), Melchionna (2007), Bussi and Parrinello (2007), Thalmann and Farago (2007) and Leimkuhler and Matthews (2013*a*) for recent suggestions.

One systematic way to derive numerical schemes of arbitrary order is to resort to splitting schemes to analytically integrate the various parts of the dynamics. However, proving rigorous ergodicity results on the corresponding numerical schemes has so far been done only for compact position spaces (see Leimkuhler *et al.* 2015). For unbounded position spaces, implicit schemes should be considered (see Mattingly *et al.* 2002, Kopec 2015), in which case some geometric convergence can be achieved. Alternatively, it is always possible to superimpose a Metropolis–Hastings step upon the discretization under consideration in order to stabilize the numerical method and ensure the existence of an invariant measure.

We restrict ourselves in the rest of this section to compact position spaces $\mathcal{D} = (L\mathbb{T})^d$. Let us first decompose the generator of Langevin dynamics as $\mathcal{L} = A + B + \gamma C$, with

$$A = M^{-1}p \cdot \nabla_q, \quad B = -\nabla V(q) \cdot \nabla_p, \quad C = -M^{-1}p \cdot \nabla_p + \frac{1}{\beta}\Delta_p. \quad (3.30)$$

First-order splitting schemes are obtained via a Lie–Trotter splitting of the elementary evolutions generated by $A, B, \gamma C$. There are six possible schemes of the general form

$$P_{\Delta t}^{Z,Y,X} = \mathrm{e}^{\Delta t Z}\, \mathrm{e}^{\Delta t Y}\, \mathrm{e}^{\Delta t X},$$

with all possible permutations (Z, Y, X) of $(A, B, \gamma C)$. For instance, the numerical scheme associated with $P_{\Delta t}^{B,A,\gamma C}$ is

$$\widetilde{p}^{n+1} = p^n - \Delta t\, \nabla V(q^n), \quad (3.31a)$$

$$q^{n+1} = q^n + \Delta t\, M^{-1}\widetilde{p}^{n+1}, \quad (3.31b)$$

$$p^{n+1} = \alpha_{\Delta t}\widetilde{p}^{n+1} + \sqrt{\frac{1-\alpha_{\Delta t}^2}{\beta}}\, M\, G^n, \quad (3.31c)$$

where $\alpha_{\Delta t} = \exp(-\gamma M^{-1}\Delta t)$, and $(G^n)_{n\geqslant 0}$ are i.i.d. Gaussian random vectors with identity covariance. Second-order splitting schemes are obtained by a Strang splitting of the elementary evolutions generated by $A, B, \gamma C$. There are again six possible schemes, which are of the general form

$$P_{\Delta t}^{Z,Y,X,Y,Z} = \mathrm{e}^{\Delta t Z/2}\, \mathrm{e}^{\Delta t Y/2}\, \mathrm{e}^{\Delta t X}\, \mathrm{e}^{\Delta t Y/2}\, \mathrm{e}^{\Delta t Z/2},$$

with the same possible orderings as for first-order schemes. In fact, as already proved in Bou-Rabee and Owhadi (2010), second-order accuracy on the invariant measure can be obtained by resorting to a first-order splitting between the Hamiltonian and the Ornstein–Uhlenbeck parts, and discretizing the Hamiltonian part with a second-order scheme. This corresponds to evolution operators of so-called geometric Langevin algorithm type, such as

$$P_{\Delta t}^{B,A,B,\gamma C} = \mathrm{e}^{\Delta t B/2}\, \mathrm{e}^{\Delta t A}\, \mathrm{e}^{\Delta t B/2}\, \mathrm{e}^{\gamma \Delta t C}.$$

The latter scheme amounts to first performing one iteration of the Verlet scheme ($\mathrm{e}^{\Delta t B/2}\, \mathrm{e}^{\Delta t A}\, \mathrm{e}^{\Delta t B/2}$) and then updating the momenta via an analytic integration of the Ornstein–Uhlenbeck process (which corresponds to $\mathrm{e}^{\gamma \Delta t C}$). Among all these splitting schemes, some lead to smaller errors on the invariant measure. Moreover, only certain geometric Langevin algorithms and second-order splitting schemes admit invariant measures consistent with ν in the overdamped regime $\gamma \to 0$. See Leimkuhler *et al.* (2015) for more precise statements.

Remark 3.6 (stability). Stability is a necessary condition for the existence of an invariant measure $\mu_{\Delta t}$ for the numerical scheme under consideration. This property is guaranteed by the existence of a Lyapunov function $\mathcal{K}$ for the transition operator $P_{\Delta t}$ (see Theorem 2.25), which typically requires the time step Δt to be sufficiently small. For discretizations of Langevin dynamics, the stability conditions are similar to those obtained for discretizations of the Hamiltonian dynamics, which correspond to $\gamma = 0$; see Leimkuhler and Matthews (2013*b*, 2015). Note indeed that all the splitting schemes introduced above reduce to a symplectic discretization of the Hamiltonian dynamics when $\gamma = 0$. Recall also that the numerical stability of these schemes requires the time step Δt to be sufficiently small. For one-dimensional systems with a harmonic potential $V(q) = \omega^2 q^2/2$, the stability constraint reads $\omega \Delta t < 2$. For more general systems, this means that the time step has to be of the order of a fraction of the period of the fastest vibrational modes.

As for overdamped Langevin dynamics, Theorem 3.3 can be used to obtain error estimates on the invariant measure (*i.e.* (3.18) with π replaced by μ), with either $p = 1$ for first-order splitting schemes or $p = 2$ for second-order splitting schemes and geometric Langevin algorithms. We consider the hierarchy of scale functions $\mathcal{K}_n(q, p) = 1 + |p|^{2n}$. There are three types of assumptions to be checked.

(i) *Assumptions on the generator $\mathcal{A}_1 = \mathcal{L}$ of the continuous dynamics.* The stability of the space of smooth functions $\mathcal{S}_0$ under $\mathcal{L}^{-1}$ follows from careful analysis of the proof presented in Talay (2002), as provided in Kopec (2015, Appendix A) for unbounded position spaces (the adaptation to compact position spaces allowing simplification of some arguments).

(ii) *Assumptions on the invariant measure $\mu_{\Delta t}$ of the numerical scheme.* It can be proved that there exists an invariant measure $\mu_{\Delta t}$ such that $\mathcal{K}_n \in L^1(\mu_{\Delta t})$ for all $n \geqslant 0$. In fact, similarly to Corollary 3.5, a much stronger result can be stated, namely the exponential convergence of the law to a unique invariant measure, in any space $L^\infty_{\mathcal{K}_n}(\mathcal{E})$ and with a rate independent of the time step (see Leimkuhler *et al.* 2015). The proof of the existence and uniqueness of the invariant measure can be somewhat simplified if no uniform convergence rate is sought (as done in Mattingly *et al.* 2002).

(iii) *Assumptions on the differential operators arising in the expansion in powers of Δt of the evolution operator $P_{\Delta t}$.* The expressions of these operators are conveniently obtained for splitting schemes with the Baker–Campbell–Hausdorff formula; see for instance the presentation

in Hairer *et al.* (2006, Section III.4.2). Such computations are reported in Leimkuhler and Matthews (2013*a*) and Leimkuhler *et al.* (2015).

3.3.4. Removing the systematic error

The estimate (3.18) not only allows us to estimate the order of magnitude of the error on average properties arising from the discretization of the dynamics, but also provides an expression of the leading-order term in the difference as

$$\int_{\mathcal{X}} \varphi \, \mathrm{d}\pi_{\Delta t} - \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi = \Delta t^p \int_{\mathcal{X}} \varphi f_{p+1} \, \mathrm{d}\pi + \mathrm{O}(\Delta t^{p+1}).$$

There are two principal strategies for reducing/removing the leading-order term on the right-hand side of the above equality.

(1) Using Romberg extrapolation as already suggested in Talay and Tubaro (1990), by performing simulations with two different time steps, in order to eliminate the leading-order error term by an appropriate linear combination of the associated estimators.

(2) Directly estimating the correction term by reformulating it as an integrated correlation function. This is possible when $\mathcal{A}_1 = \mathcal{L}$, namely for schemes of weak order at least one. Indeed, the expression (3.19) for f_{p+1} leads to

$$\begin{aligned} \int_{\mathcal{X}} \varphi f_{p+1} \, \mathrm{d}\pi &= \int_{\mathcal{X}} \Pi\varphi f_{p+1} \, \mathrm{d}\pi = - \int_{\mathcal{X}} (\mathcal{L}^{-1}\Pi\varphi) g_{p+1} \, \mathrm{d}\pi \\ &= \int_0^{+\infty} \int_{\mathcal{X}} (\mathrm{e}^{t\mathcal{L}}\Pi\varphi) g_{p+1} \, \mathrm{d}\pi \, \mathrm{d}t \\ &= - \int_0^{+\infty} \mathbb{E}[\varphi(x_t) g_{p+1}(x_0)] \, \mathrm{d}t, \end{aligned} \tag{3.32}$$

where we have implicitly assumed that decay estimates hold for the semigroup $\mathrm{e}^{t\mathcal{L}}$ (which allows us to rewrite the resolvent $-\mathcal{L}^{-1}$ as a time integral of the semigroup), and where the expectation $\mathbb{E}_\pi$ is over all initial conditions $x_0 \sim \pi$ and for all realizations of the dynamics with generator $\mathcal{L}$. Recall that the expression of the function g_{p+1} is usually not too difficult to obtain once the expansion (3.15) has been worked out: see the discussion after (3.20). The integrated correlation on the right-hand side is then approximated as described in Section 5 (see Theorem 5.6). This strategy has been tested on a simple case in Leimkuhler *et al.* (2015): see Figure 5.1 in Section 5.3 below.

An alternative strategy consists in using a Metropolis–Hastings algorithm with the numerical scheme as a proposal, in order to completely remove

the systematic error due to the time discretization. This is straightforward for reversible dynamics such as overdamped Langevin dynamics. When an Euler–Maruyama scheme is used, one obtains the so-called Metropolis-adjusted Langevin algorithm (MALA) in the statistics literature, known as smart MC in the chemistry literature (Rossky, Doll and Friedman 1978, Roberts and Rosenthal 1998). Another advantage of superimposing a Metropolis–Hastings procedure upon a discretization of overdamped Langevin dynamics is that it stabilizes the numerical scheme even for non-globally Lipschitz forces ∇V: an invariant probability measure exists by construction, which ensures the recurrence of the Markov chain. In contrast, numerical discretizations which are not stabilized by a Metropolis–Hastings procedure may be transient, as discussed in Section 3.2.

For Langevin dynamics, some care has to be taken since the transition kernel associated with the Hamiltonian dynamics is irregular (because the noise acts only on momenta), which raises some difficulties in the definition of the Metropolis–Hastings ratio. It is nonetheless possible to use a Metropolis–Hastings procedure for schemes based on a splitting between the Ornstein–Uhlenbeck process on the momenta and a reversible discretization of the Hamiltonian part, upon reverting momenta when rejecting proposed moves; see Lelièvre *et al.* (2010, Section 2.2.3.2) for a precise discussion.

However, Metropolis corrections are usually not implemented in popular molecular dynamics packages such as NAMD since it is not always possible or desirable to use a Metropolis correction. First, the average acceptance probability in the Metropolis step for MALA or related algorithms in general decreases exponentially with the dimension of the system for a *fixed* time step. In fact, the time step should be reduced as some inverse power of the system size in order to maintain a constant acceptance rate (see Roberts, Gelman and Gilks 1997, Roberts and Rosenthal 1998). There are two ways to limit the decrease of the acceptance probability.

(i) Change the dynamics or the measure used to compute the Metropolis ratio. For the Metropolization of the Hamiltonian dynamics, see the works by Izaguirre and Hampton (2004) and Akhmatskaya and Reich (2005) where the Hamiltonian H in the canonical measure is changed to $H + \Delta t^2 \widetilde{H}$, with $\widetilde{H}$ the first-order correction arising from backward error analysis. For the Metropolization of overdamped Langevin dynamics, see Fathi and Stoltz (2015).

(ii) Evolve only subparts of the system (in the spirit of a Gibbs sampler), as advocated by Bou-Rabee and Vanden-Eijnden (2012).

However, the latter strategy may complicate the implementation of parallel algorithms for the simulation of very large systems, especially if long-range

potentials are used (as acknowledged in Bou-Rabee and Vanden-Eijnden 2012, Remark 2.5). Second, the variance of the computed averages may increase since rejections occur, and the samples along the numerical trajectory are therefore more correlated in general than for rejection-free dynamics. Lastly, the Metropolis procedure requires that the invariant measure π of the continuous dynamics is known. This is the case for Langevin and overdamped Langevin dynamics with forces $-\nabla V(q)$. However, for non-reversible systems subjected to external forcings such as a temperature gradient or a non-gradient force (see Section 5), the invariant measure of the continuous dynamics is not known.

3.4. Variance reduction

As already mentioned in (3.14), the variance of trajectory averages computed with discretizations of SDEs are consistent with the variance of trajectory averages computed with the continuous dynamics they approximate. In this section we discuss standard ways to reduce this variance in order to decrease the statistical error in the estimated averages (3.10).

A first important distinction should be made between target-oriented variance reduction, which corresponds to reducing the asymptotic variance ς_φ^2 defined in (3.5) for a given observable φ; and a general-purpose reduction, for which the aim is to decrease

$$\sup_{\varphi \in L^2(\pi)} \frac{\varsigma_\varphi^2}{\|\varphi\|_{L^2(\pi)^2}} = \sup_{\|\varphi\|_{L^2(\pi)} \leqslant 1} 2 \int_{\mathcal{X}} \Pi\varphi(-\mathcal{L}^{-1}\Pi\varphi)\, \mathrm{d}\pi. \tag{3.33}$$

In (3.33), we choose $\varphi \in L^2(\pi)$ for simplicity, but the same question may be raised for specific subsets of $L^2(\pi)$.

For Markov chains or Markov processes, the fact that the variance is large is due to the fact that the correlation time of the dynamics is large (see the discussion after (3.6)), which itself is due to the metastability of the process. The idea is therefore to modify the dynamics in order to remove, or at least reduce the metastability, while still being able to reconstruct canonical averages.

We present in this section standard variance reduction techniques for Markov processes. These techniques are extensions of methods used for i.i.d. sequences, for which a review is provided in Caflisch (1998). The most famous techniques are antithetic variables, stratification, control variate methods and importance sampling. We present flavours of the latter three approaches in this section.

3.4.1. Stratification

Stratification is a way of decomposing a difficult sampling problem into several easier ones. Ideally, the phase space should be decomposed into the

collection of all metastable states, corresponding to local minima of the potential energy function, and these regions should be independently sampled. The local averages in each region should then be reweighted according to the canonical weight of the region itself. This method is therefore a general-purpose variance reduction technique.

There are two major ways to make this idea practical, depending on whether the considered regions overlap.

(i) When there is some overlap between the regions, bridge sampling methods such as MBAR (Shirts and Chodera 2008) can be used. The method is based on several works in statistics (Geyer 1994, Meng and Wong 1996, Kong *et al.* 2003, Tan 2004).

(ii) Non-overlapping regions can also be constructed as the level sets of some real-valued function of the configuration of the system. In this case, the sampling is performed by constraining the dynamics on the level set, and varying the constant defining the level set in order to sample the full phase space. This method is known as thermodynamic integration, with a reconstruction performed by computing the free energy: see Section 4.5.

3.4.2. Control variate

The control variate method is a classical technique for variance reduction, although it is not so often used in molecular simulation. Consider the case when thermodynamic averages such as (1.3) or (1.4) are estimated by ergodic averages of a stochastic dynamics with generator $\mathcal{L}$. Recall that we denote the invariant measure by π and the configuration of the system by $x \in \mathcal{X}$, so that thermodynamic averages read $\mathbb{E}_\pi(\varphi)$.

Let us first explain the general principle of control variates applied to a simple example. Consider a given observable φ, and introduce an observable ϕ such that $\mathbb{E}_\pi(\phi) = 0$. Then,

$$\mathbb{E}_\pi(\varphi) = \mathbb{E}_\pi(\varphi - \phi).$$

The idea now is to choose ϕ such that $\mathrm{Var}_\pi(\varphi - \phi)$ is much smaller than $\mathrm{Var}_\pi(\varphi)$. Of course, the optimal choice is $\phi = \varphi - \mathbb{E}_\pi(\varphi)$, in which case $\mathrm{Var}_\pi(\varphi - \phi) = 0$. Note however that the optimal control variate depends on the quantity of interest, $\mathbb{E}_\pi(\varphi)$, which is not available.

A systematic way of constructing admissible control variates (that is, functions with average 0 with respect to π) is to choose them in the image of the generator $\mathcal{L}$. Indeed, the invariance of the measure π, formulated as (1.17), leads to the following equality: for any C^∞ and compactly supported test function Φ,

$$\mathbb{E}_\pi(\varphi) = \mathbb{E}_\pi(\varphi - \mathcal{L}\Phi) = \int_{\mathcal{X}} (\varphi - \mathcal{L}\Phi)\, \mathrm{d}\pi. \tag{3.34}$$

The optimal choice corresponds to the solution of the Poisson equation

$$\mathcal{L}\Phi = \varphi - \mathbb{E}_\pi(\varphi). \tag{3.35}$$

The solvability of this equation is ensured by Corollaries 2.4, 2.16, 2.21 and 2.27, depending on the dynamics at hand and/or the integrability properties of the observable φ under consideration. By construction, the estimator based on (3.34)–(3.35), namely

$$\frac{1}{t}\int_0^t (\varphi(x_s) - \mathcal{L}\Phi(x_s))\,\mathrm{d}s,$$

has a variance equal to 0 since $\varphi(x) - \mathcal{L}\Phi(x) = \mathbb{E}_\pi(\varphi)$ for any value of $x \in \mathcal{X}$. Such approaches were first suggested for Markov chains in the computational statistics literature (Andradóttir, Heyman and Ott 1993, Henderson 1997) but have also been proposed in the statistical physics literature (Assaraf and Caffarel 1999), where they are known as the 'zero-variance principle'.

In practice, it is generally impossible to solve the Poisson equation (3.35) exactly. However, it is possible to approximate the ideal function Φ in (3.35) on a basis of K trial functions $\Phi_1, \ldots, \Phi_K$. This amounts to optimizing the coefficients $a = (a_1, \ldots, a_K) \in \mathbb{R}^K$ such that the variance of the estimator

$$\widehat{\varphi}_t(a_1, \ldots, a_K) = \frac{1}{t}\int_0^t \phi_{a_1,\ldots,a_K}(x_s)\,\mathrm{d}s, \quad \phi_{a_1,\ldots,a_K} = \varphi - \sum_{k=1}^K a_k \mathcal{L}\Phi_k$$

is minimal.

3.4.3. Importance sampling

The basic idea of importance sampling is to change the measure which is sampled into a measure which is easier to sample. In the context of molecular simulation, this is most commonly done by changing the potential energy function V in the dynamics to a modified potential $V + \widetilde{V}$. However, there are alternative strategies: see for instance Redon *et al.* (2016), where the kinetic energy function is modified for Langevin dynamics. The modified overdamped Langevin dynamics associated with the potential $V + \widetilde{V}$, namely

$$\mathrm{d}q_t = -\nabla(V + \widetilde{V})(q_t)\,\mathrm{d}t + \sqrt{\frac{2}{\beta}}\,\mathrm{d}W_t, \tag{3.36}$$

is then ergodic for the modified probability measure $\nu_{\widetilde{V}} = Z_{\widetilde{V}}^{-1}\,\mathrm{e}^{-\beta\widetilde{V}}\nu$. Likewise, the modified Langevin dynamics

$$\begin{aligned}
\mathrm{d}\widetilde{q}_t &= M^{-1}\widetilde{p}_t\,\mathrm{d}t, \\
\mathrm{d}\widetilde{p}_t &= -\nabla(V + \widetilde{V})(\widetilde{q}_t)\,\mathrm{d}t - \gamma\, M^{-1}\widetilde{p}_t\,\mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}}\,\mathrm{d}W_t
\end{aligned}$$

is ergodic for the modified canonical probability measure $\mu_{\widetilde{V}} = Z_{\widetilde{V}}^{-1}\,\mathrm{e}^{-\beta\widetilde{V}}\mu$. The fundamental observation to retrieve averages with respect to ν or μ with realizations of the modified dynamics is that

$$\int_{\mathcal{D}} \varphi(q)\,\nu(\mathrm{d}q) = \int_{\mathcal{E}} \varphi(q)\,\mu(\mathrm{d}q\,\mathrm{d}p) = \frac{\int_{\mathcal{D}} \varphi\,\mathrm{e}^{\beta\widetilde{V}}\,\mathrm{d}\nu_{\widetilde{V}}}{\int_{\mathcal{D}} \mathrm{e}^{\beta\widetilde{V}}\,\mathrm{d}\nu_{\widetilde{V}}} = \frac{\int_{\mathcal{E}} \varphi\,\mathrm{e}^{\beta\widetilde{V}}\,\mathrm{d}\mu_{\widetilde{V}}}{\int_{\mathcal{E}} \mathrm{e}^{\beta\widetilde{V}}\,\mathrm{d}\mu_{\widetilde{V}}}. \quad (3.37)$$

Therefore, the following estimator is considered:

$$\widehat{\varphi}_t^{\widetilde{V}} = \frac{\int_0^t \varphi(\widetilde{q}_s)\,\mathrm{e}^{\beta\widetilde{V}(\widetilde{q}_s)}\,\mathrm{d}s}{\int_0^t \mathrm{e}^{\beta\widetilde{V}(\widetilde{q}_s)}\,\mathrm{d}s}. \quad (3.38)$$

After discretization in time, the estimator $\widehat{\varphi}_t^{\widetilde{V}}$ is approximated by

$$\widehat{\varphi}_{N_{\mathrm{iter}},\Delta t}^{\widetilde{V}} = \frac{\sum_{n=0}^{N_{\mathrm{iter}}} \varphi(\widetilde{q}^n)\,\mathrm{e}^{\beta\widetilde{V}(\widetilde{q}^n)}}{\sum_{n=0}^{N_{\mathrm{iter}}} \mathrm{e}^{\beta\widetilde{V}(\widetilde{q}^n)}},$$

where $\widetilde{q}^n$ is an approximation of $\widetilde{q}_{n\Delta t}$.

In order for importance sampling to be efficient, the weights $\mathrm{e}^{\beta\widetilde{V}(q^n)}$ should not be too degenerate. This can be quantified in various ways, for instance through the so-called efficiency factor

$$\frac{\left(\sum_{n=0}^{N_{\mathrm{iter}}} \mathrm{e}^{\beta\widetilde{V}(q^n)}\right)^2}{N_{\mathrm{iter}} \sum_{n=0}^{N_{\mathrm{iter}}} \mathrm{e}^{2\beta\widetilde{V}(q^n)}} \in [0,1].$$

The fact that this number is indeed between 0 and 1 can be seen via the Cauchy–Schwarz inequality. The efficiency factor counts the fraction of significant values on average. It should be as close as possible to 1, which is indeed the case if the weights are of similar magnitudes. In the statistics literature, the efficiency factor corresponds to the effective sample size of Kong, Liu and Wong (1994) divided by the number of sampled values.

When the observable φ under consideration is fixed (target-oriented variance reduction), it is possible to optimize the importance sampling procedure in order to minimize the variance of the estimator. We illustrate this fact for samples $\widetilde{q}^n$ independently and identically distributed from a measure with density $\mathrm{e}^{-\beta(V(q)+\widetilde{V}(q))}$, whose normalization constant fortunately does not need to be known. Consider the estimator

$$\widehat{\varphi}_{N_{\mathrm{iter}}}^{\mathrm{iid},\widetilde{V}} = \frac{\sum_{n=0}^{N_{\mathrm{iter}}} \varphi(\widetilde{q}^n)\,\mathrm{e}^{\beta\widetilde{V}(\widetilde{q}^n)}}{\sum_{n=0}^{N_{\mathrm{iter}}} \mathrm{e}^{\beta\widetilde{V}(\widetilde{q}^n)}}, \quad \widetilde{q}^n \sim \nu_{\widetilde{V}} \ \text{ i.i.d.}$$

A simple computation based on the equality[9] $\nu_{\widetilde{V}}/\nu = \mathrm{e}^{-\beta\widetilde{V}}\,Z_0/Z_{\widetilde{V}}$ shows

[9] With some abuse of notation, we denote the measures $\nu_{\widetilde{V}}(\mathrm{d}q)$ and their densities by

that

$$\sqrt{N_{\rm iter}}(\widehat{\varphi}^{{\rm iid},\widetilde{V}}_{N_{\rm iter}} - \mathbb{E}_\nu(\varphi)) = \frac{\sqrt{N_{\rm iter}}\left[\frac{1}{N_{\rm iter}}\sum_{n=0}^{N_{\rm iter}}[\varphi(\widetilde{q}^n) - \mathbb{E}_\nu(\varphi)]\frac{\nu}{\nu_{\widetilde{V}}}(\widetilde{q}^n)\right]}{\frac{1}{N_{\rm iter}}\sum_{n=0}^{N_{\rm iter}}\frac{\nu}{\nu_{\widetilde{V}}}(\widetilde{q}^n)}.$$

By the law of large numbers, the denominator almost surely converges to 1 as $N_{\rm iter} \to +\infty$, while, by the central limit theorem, the numerator converges in law to a Gaussian distribution with variance

$$\varsigma^2_{\widetilde{V}}(\varphi) = \int_{\mathcal{D}} \frac{(\Pi\varphi)^2\nu^2}{\nu_{\widetilde{V}}},$$

where we introduced $\Pi\varphi = \varphi - \mathbb{E}_\nu(\varphi)$. By Slutsky's theorem, the random variable

$$\sqrt{N_{\rm iter}}\big(\widehat{\varphi}^{{\rm iid},\widetilde{V}}_{N_{\rm iter}} - \mathbb{E}_\nu(\varphi)\big)$$

therefore converges in law to a Gaussian distribution with variance $\varsigma^2_{\widetilde{V}}(\varphi)$. The Cauchy–Schwarz inequality on $L^2(\nu_{\widetilde{V}})$ now shows that

$$\varsigma^2_{\widetilde{V}}(\varphi) \geqslant \left(\int_{\mathcal{D}} \frac{|\Pi\varphi|\nu}{\nu_{\widetilde{V}}}\,\nu_{\widetilde{V}}\right)^2 = \left(\int_{\mathcal{D}} |\Pi\varphi|\,{\rm d}\nu\right)^2,$$

with equality if and only if $|\Pi\varphi|\nu/\nu_{\widetilde{V}} \propto 1$. The optimal biased measure for i.i.d. sampling is thus

$$\nu_{\widetilde{V}}({\rm d}q) = \frac{|\Pi\varphi(q)|\,\nu({\rm d}q)}{\int_{\mathcal{D}}|\Pi\varphi|\,{\rm d}\nu}$$

which formally corresponds to the potential

$$\widetilde{V}(q) = -\frac{1}{\beta}\log|\Pi\varphi(q)|.$$

The latter expression is singular at configurations q such that $\Pi\varphi(q) = 0$. More importantly, the expression of $\widetilde{V}$ depends on $\mathbb{E}_\nu(\varphi)$, and is therefore not available as such. We will extend such importance sampling approaches to averages over the path space in Section 6.2, and will also show how to approximate the optimal bias in this context.

Let us finally discuss general-purpose variance reduction in the context of importance sampling. In view of the definition (3.33), a reduction in the variance for any observable φ amounts to a decrease in the operator norm of the *symmetric* part of $\mathcal{L}^{-1}$. Note indeed that only the symmetric part of of $\mathcal{L}^{-1}$ matters in the right-hand side of (3.33). For overdamped Langevin dynamics, the generator is self-adjoint: $\mathcal{L} = \mathcal{L}^*$ when these operators are

the same symbol. The ratio $\nu_{\widetilde{V}}/\nu$ is the Radon–Nikodym derivative of $\nu_{\widetilde{V}}$ with respect to ν.

considered on $L^2(\nu)$. General variance reduction therefore amounts to increasing the spectral gap of the operator (see (2.82)). This can be done by choosing $\widetilde{V}$ to erase local minima in V which degrade the Poincaré/LSI constants (see Section 2.2). In Section 4 we discuss good choices of $\widetilde{V}$ to overcome such metastability issues. When the generator $\mathcal{L}$ is not self-adjoint, there are *a priori* no simple relationships between the symmetric and antisymmetric parts of $\mathcal{L}$ and its inverse. However, some results can be obtained for specific dynamics, such as overdamped Langevin dynamics perturbed by a divergence-free non-reversible drift: see Duncan, Lelièvre and Pavliotis (2016) and Rey-Bellet and Spiliopoulos (2016).

4. Free energy computations and adaptive importance sampling methods

This section is devoted to presentation of a sampling method of the canonical measure, which was developed in the field of molecular dynamics but could also be used in many other contexts. As explained in Section 1.2, sampling the canonical measure is difficult because the dimension is large, and the measure is multimodal. The high-dimensionality of the problem entails the use of stochastic dynamics to sample the target measure, and the multimodality of the target implies the metastability of the process (and thus the slow convergence of time averages). In this section we present techniques which circumvent the difficulties raised by the metastability of the sampling dynamics, by using two ideas.

- Use an importance sampling method (see Section 3.4.3) by changing the original potential V to[10] $V - F \circ \xi$, where F is the so-called *free energy* associated with the *reaction coordinate* ξ (see Section 4.1 for precise definitions).
- Since the free energy is not available in practice, instead of the biasing potential $F \circ \xi$, an approximation $F_t \circ \xi$ is used, where F_t is an estimate of the free energy at time t, given all the configurations visited so far: F_t is updated on-the-fly in order to better approximate the free energy F as time goes on (this is the adaptive feature of the approach).

Methods of this type are thus called *free energy adaptive biasing techniques*, and encompass many algorithms (Lelièvre, Rousset and Stoltz 2007): the adaptive biasing force (Darve and Porohille 2001, Hénin and Chipot 2004), the Wang–Landau algorithm (Wang and Landau 2001*a*, Wang and Landau 2001*b*), metadynamics (Laio and Parrinello 2002, Bussi, Laio and Parrinello 2006), and the self-healing umbrella sampling method (Marsili *et al.* 2006, Dickson *et al.* 2010).

[10] Here and in the following, $\circ$ denotes the composition operator.

This section is organized as follows. In Section 4.1 we introduce the notion of free energy. In Section 4.2 we present in detail the two families of adaptive techniques: adaptive biasing potential (ABP) and adaptive biasing force (ABF) methods, in a rather general setting. Section 4.3 is devoted to the mathematical analysis of an ABF approach based on entropy techniques and logarithmic Sobolev inequalities (in the spirit of the techniques introduced in Section 2.2). For pedagogical purposes, this mathematical analysis is presented in a simple setting. Extensions and related works are mentioned in Section 4.4. The focus of this section is thus on adaptive techniques. We will show in particular how an analysis based on partial differential equations is very useful for understanding the efficiency of the methods. However, we will also give a quick overview of other free energy calculation techniques in Section 4.5.

4.1. Free energy

4.1.1. Reaction coordinate

In order to introduce the notions of free energy and reaction coordinate, let us go back to the simple two-dimensional experiments of Section 1.2 reported in Figure 1.1. On the time evolution of the x-coordinate of the stochastic process, one observes metastable features: the process remains trapped for a very long time in some interval around -1 before hopping to another metastable region, around 1. The transition between the two metastable states can be inferred from the time evolution of the x-coordinate. The x-coordinate is thus a slow variable of the system, in the sense that the typical time scale on which hopping events occur is much larger than the typical time scale of vibrations within metastable regions. In this simple two-dimensional example, a reaction coordinate is thus the function $\xi(x, y) = x$ (where we let (x, y) denote the two components of the position q). Generally speaking, a reaction coordinate is a function

$$\xi : \mathbb{R}^d \to \mathbb{R}^m$$

with $m \leqslant d$, such that $(\xi(q_t))_{t\geqslant 0}$ is a metastable process. Here and in the following, we assume for simplicity that the position space $\mathcal{D}$ is $\mathbb{R}^d$, but the generalization to a general space $\mathcal{D}$ is straightforward. The function ξ thus indexes the transitions between some metastable states. In addition we require the dimension m of the reaction coordinate to be as small as possible, in order to have a low-dimensional representation of the original high-dimensional system via $(\xi(q_t))_{t\geqslant 0}$ (for the numerical methods presented in this section, m is typically required to be less than 4, even though some methods have been proposed to deal with higher-dimensional reaction coordinates; see for example Piana and Laio 2007). In practice, it could be an angle in a protein, for instance, which characterizes the conformation

of the molecule, or the position of a defect in a material. For choosing a reaction coordinate ξ, the practitioner needs some *a priori* knowledge of the system, since the reaction coordinate should characterize the conformations of the system which correspond to some metastable states of the dynamics.

4.1.2. Free energy

Once a reaction coordinate ξ has been chosen, one can introduce the free energy associated with ξ. Let us again go back to the simple two-dimensional example of Figure 1.1, and to the specific choice $\xi(x, y) = x$. One can define the marginal distribution of

$$\nu(\mathrm{d}x\,\mathrm{d}y) = Z_\nu^{-1} \exp(-\beta V(x, y))\,\mathrm{d}x\,\mathrm{d}y,$$

in x, namely the probability measure on $\mathbb{R}$:

$$\left(Z_\nu^{-1} \int_{\mathbb{R}} \exp(-\beta V(x, y))\,\mathrm{d}y \right) \mathrm{d}x.$$

Notice that the integration is performed only over y in the previous definition. Then, the free energy $F(x)$ is defined so that $\exp(-\beta F(x))$ is the density of this marginal law:

$$F(x) = -\beta^{-1} \ln\left(Z_\nu^{-1} \int_{\mathbb{R}} \exp(-\beta V(x, y))\,\mathrm{d}y \right).$$

In a more general setting, a similar definition holds.

Definition 4.1 (free energy). The free energy $F : \mathbb{R}^m \to \mathbb{R}$ associated with the canonical measure $\nu(\mathrm{d}q) = Z_\nu^{-1} \exp(-\beta V(q))\,\mathrm{d}q$ and the reaction coordinate $\xi : \mathbb{R}^d \to \mathbb{R}^m$ is defined by

$$F(z) = -\beta^{-1} \ln\left(\int_{\Sigma(z)} \exp(-\beta V(q)) \delta_{\xi(q)-z}(\mathrm{d}q) \right), \quad \text{for all } z \in \mathbb{R}^m, \quad (4.1)$$

where $\Sigma(z) = \{q \in \mathbb{R}^d,\, \xi(q) = z\}$.

The free energy F (as the potential V) is actually defined up to an additive constant, and is such that $Z_F^{-1} \exp(-\beta F(z))\,\mathrm{d}z$, where

$$Z_F = \int_{\mathbb{R}^m} \exp(-\beta F(z))\,\mathrm{d}z = Z_\nu,$$

by definition (4.1), is the image of the measure ν by ξ (or equivalently the marginal distribution of ν along ξ). Since the free energy F is defined up to an additive constant, quantities of interest are actually *free energy differences* between two states indexed by ξ.

In Definition 4.1, we used the delta measure $\delta_{\xi(q)-z}(\mathrm{d}q)$, which is a measure supported by the submanifold $\Sigma(z)$. One way to define this measure is

the following: the delta measure $\delta_{\xi(q)-z}(\mathrm{d}q)$ is such that '$\delta_{\xi(q)-z}(\mathrm{d}q)\,\mathrm{d}z = \mathrm{d}q$'. More precisely, for all C^∞ and compactly supported test functions $A : \mathbb{R}^m \to \mathbb{R}$ and $B : \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} A \circ \xi(q) B(q)\,\mathrm{d}q = \int_{\mathbb{R}^m} A(z) \left(\int_{\Sigma(z)} B(q) \delta_{\xi(q)-z}(\mathrm{d}q) \right) \mathrm{d}z. \tag{4.2}$$

The delta measure can be understood using a regularization procedure: for any C^∞ and compactly supported test function $A : \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\Sigma(z)} A(q) \delta_{\xi(q)-z}(\mathrm{d}q) = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} A(q) \delta^\varepsilon(\xi(q) - z)\,\mathrm{d}q,$$

where δ^ε is a C^∞ approximation of the identity (when $\varepsilon \to 0$, δ^ε converges to the Dirac in the sense of distributions).

Remark 4.2 (co-area formula). The delta measure can also be defined in terms of the surface measure $\sigma_{\Sigma(z)}$ over the submanifold $\Sigma(z) \subset \mathbb{R}^d$, equipped with the Riemannian structure from $\mathbb{R}^d$:

$$\delta_{\xi(q)-z}(\mathrm{d}q) = (\det G)^{-1/2}\,\mathrm{d}\sigma_{\Sigma(z)},$$

where

$$G = (\nabla \xi)^T \nabla \xi, \tag{4.3}$$

which has components

$$G_{\alpha,\beta} = \nabla \xi_\alpha \cdot \nabla \xi_\beta, \quad \text{for all } (\alpha, \beta) \in \{1, \ldots, m\}^2.$$

The equivalence between this definition and the previous one is a consequence of the *co-area formula* (Ambrosio, Fusco and Pallara 2000, Evans and Gariepy 1992, Lelièvre *et al.* 2010). Note that it is implicitly assumed here and in the following that $\det G \neq 0$ (where G is defined by (4.3)). Contrary to the surface measure $\sigma_{\Sigma(z)}$, the measure $\delta_{\xi(q)-z}(\mathrm{d}z)$ does not only depend on ξ via its level set $\Sigma(z)$, but also depends on the value of $\nabla \xi$ over $\Sigma(z)$. Note that in the simple case where $m = 1$ (one-dimensional reaction coordinate), $\det G = |\nabla \xi|^2$. Further, when $\xi(q_1, \ldots, q_d) = (q_1, \ldots, q_m)$, then $\det G = 1$ and $\delta_{\xi(q)-z}(\mathrm{d}q) = \mathrm{d}q_{m+1} \cdots \mathrm{d}q_n$ is simply the surface measure on the hyperplane

$$\Sigma(z) = \{q \in \mathbb{R}^d,\, (q_1, \ldots, q_m) = (z_1, \ldots, z_m)\}.$$

Once F has been defined by (4.1), the fact that $Z_F^{-1} \exp(-\beta F(z))\,\mathrm{d}z$ is the image of the measure ν by ξ is actually a direct consequence of (4.2). Indeed, for any C^∞ and compactly supported test functions $A : \mathbb{R}^m \to \mathbb{R}$

and $B : \mathbb{R}^d \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}_\nu(A(\xi(q))B(q)) &= \int_{\mathbb{R}^d} A\circ\xi(q)B(q)Z_\nu^{-1}\exp(-\beta V(q))\,\mathrm{d}q \\ &= \int_{\mathbb{R}^m} A(z)\left(\int_{\Sigma(z)} B(q)Z_\nu^{-1}\exp(-\beta V(q))\delta_{\xi(q)-z}(\mathrm{d}q)\right)\mathrm{d}z \\ &= \int_{\mathbb{R}^m} A(z)\frac{\int_{\Sigma(z)} B(q)Z_\nu^{-1}\exp(-\beta V(q))\delta_{\xi(q)-z}(\mathrm{d}q)}{Z_F^{-1}\exp(-\beta F(z))}Z_F^{-1}\exp(-\beta F(z))\,\mathrm{d}z.\end{aligned}$$

Note that

$$Z_F^{-1}\exp(-\beta F(z)) = \int_{\Sigma(z)} Z_\nu^{-1}\exp(-\beta V(q))\delta_{\xi(q)-z}(\mathrm{d}q),$$

so the fraction in the above integral defines the average of B with respect to a probability measure supported by $\Sigma(z)$. More precisely, from this formula, one immediately gets that the marginal law of ν along ξ is $Z_F^{-1}\exp(-\beta F(z))$, and that the conditional probability of ν, given that $\xi(x) = z$, is given by

$$\nu(\mathrm{d}q|\xi(q) = z) = \frac{Z_\nu^{-1}\exp(-\beta V(q))\delta_{\xi(q)-z}(\mathrm{d}q)}{Z_F^{-1}\exp(-\beta F(z))}. \tag{4.4}$$

Observe that, up to an additive constant,

$$F(z) = -\beta^{-1}\ln Z_z, \tag{4.5}$$

where

$$Z_z = \int_{\Sigma(z)} Z_\nu^{-1}\exp(-\beta V(q))\delta_{\xi(q)-z}(\mathrm{d}q)$$

is the partition function of the probability measure $\nu(\mathrm{d}q|\xi(q) = z)$. This justifies the name 'free energy' (see Lelièvre *et al.* 2010, Section 1.3.1.2).

As a consequence, the free energy F can be seen as a 'coarse-grained' potential function (for the degrees of freedom $\xi(q)$), which is thermodynamically consistent with the original Boltzmann–Gibbs measure ν, since the Boltzmann–Gibbs measure associated with F (*i.e.* $Z_F^{-1}\exp(-\beta F(z))\,\mathrm{d}z$) is exactly the marginal law of ν along ξ. In fact, computing the free energy is now the aim of many molecular dynamics simulations (see Chipot and Pohorille 2007), since it gives a reduced description of the system along specific degrees of freedom (the reaction coordinate).

4.2. Free energy adaptive biasing techniques

Let us now discuss why the free energy is useful for sampling purposes.

4.2.1. Free energy biased dynamics

As explained above, the main idea of free energy biasing techniques is to use an importance sampling approach by replacing the original potential V

by the biased potential $V - F \circ \xi$, where F is the free energy associated with the reaction coordinate ξ. The intuition behind this idea is that by doing so, the metastable features of the original potential along ξ will be removed. Indeed, by the very definition of F, the marginal of the Boltzmann–Gibbs distribution $Z^{-1} \exp(-\beta(V - F \circ \xi)(q))\, \mathrm{d}q$ in ξ has density

$$\begin{aligned} &\int_{\Sigma(z)} \exp[-\beta(V(q) - F \circ \xi(q))] \delta_{\xi(q)-z}(\mathrm{d}q) \\ &\qquad = \exp(\beta F(z)) \int_{\Sigma(z)} \exp(-\beta V(q)) \delta_{\xi(q)-z}(\mathrm{d}q) \\ &\qquad = \exp(\beta F(z)) \exp(-\beta F(z)) = 1 \end{aligned} \tag{4.6}$$

(up to a normalizing constant). This means that along ξ the biased measure $Z^{-1} \exp(-\beta(V - F \circ \xi)(q))\, \mathrm{d}q$ is uniform. For the normalizing constant $Z = \int_{\mathbb{R}^d} \exp(-\beta(V - F \circ \xi)(q))\, \mathrm{d}q$ to be finite, one needs to consider a situation where ξ lives in a compact space (for example the torus if one thinks of an angle as a reaction coordinate) or to restrict the bias to a compact domain of $\mathbb{R}^m$. The idea is thus as follows. If the free energy F were known, it would be a good biasing potential since the biased Boltzmann–Gibbs measure $Z^{-1} \exp(-\beta(V - F \circ \xi)(q))\, \mathrm{d}q$ has a marginal law in ξ which is the uniform law. It is thus expected that the associated sampling dynamics (say Langevin or overdamped Langevin with $V - F \circ \xi$ as a potential) will no longer be metastable along the reaction coordinate ξ.

Before explaining how to make this idea practical by changing F to an approximation of F in the previous reasoning, let us check that F is indeed a good biasing potential on the simple two-dimensional examples we considered in Section 1.2. In Figure 4.1 we again plot the x-coordinate for overdamped Langevin dynamics as a function of time, but using the biased potential $V - F \circ \xi$ instead of the original potential V, with $\xi(x, y) = x$ as a reaction coordinate. In these simple two-dimensional examples, the free energy can be very easily computed by a simple integration in y, for a fixed value of x. By comparing the results with those of Figure 1.1, it is obvious that the biased dynamics yields a much better coverage of the configuration space: the system now freely moves from $x = -1$ to $x = 1$, both in the energetic and the entropic barrier cases.

4.2.2. Updating formulae

Let us now explain the second key idea of free energy adaptive biasing techniques, namely the adaptive feature of these numerical methods. As explained above, we would like to use $V - F \circ \xi$ as a potential instead of V. The problem is that F is unknown (it is actually the aim of many molecular dynamics computations to compute F). The idea is then to use $V - F_t \circ \xi$ as a potential, and to update F_t in such a way that F_t better approximates F as time goes on.

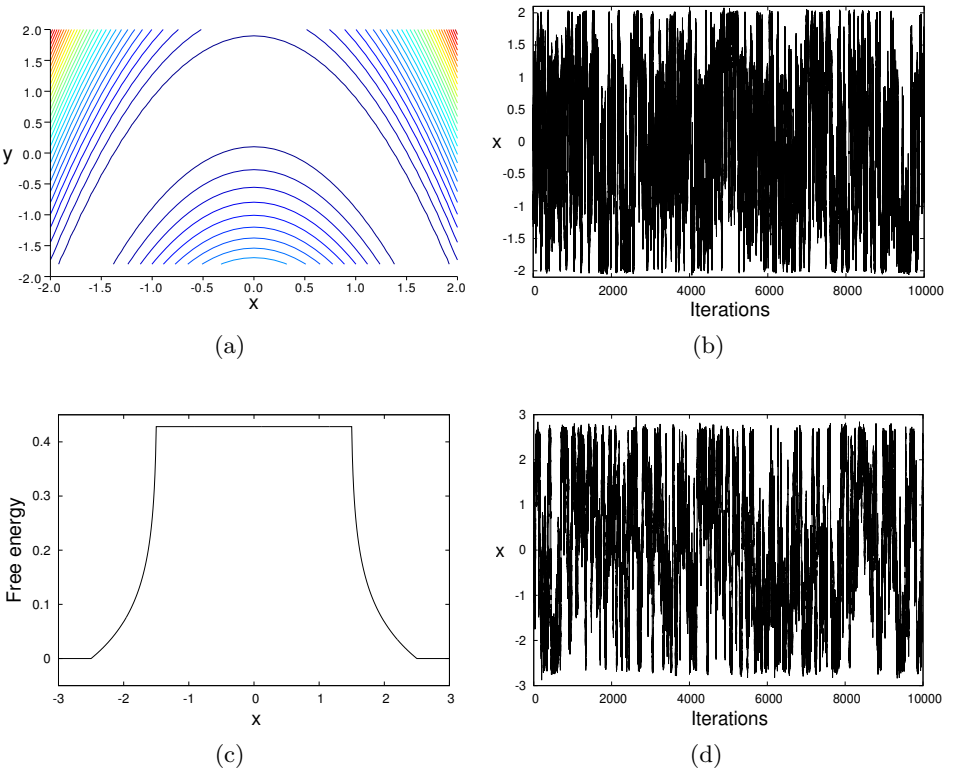


Figure 4.1. To be compared with Figure 1.1. (a, b) Energetic barrier case, (c, d) entropic barrier case. (a) Level sets of the free energy biased potential. (b) Time evolution of the x-coordinate on the biased potential ($\beta = 4$). (c) Free energy. (d) Time evolution of the x-coordinate on the biased potential ($\beta = 10$).

There are two different strategies. *Adaptive biasing potential* (ABP) techniques rely on the following formula for F:

$$\exp(-\beta F(z)) = \int_{\Sigma(z)} \exp(-\beta V(q))\delta_{\xi(q)-z}(\mathrm{d}q), \quad \text{for all } z \in \mathbb{R}^m.$$

Thus an approximation of F at time t (up to an additive constant) is given by using the empirical marginal distribution along ξ of the configurations visited so far:

$$\exp(-\beta F_t^\varepsilon(z)) \simeq \int_0^t \delta^\varepsilon(\xi(q_s) - z)\,\mathrm{d}s, \quad \text{for all } z \in \mathbb{R}^m,$$

where $(q_s)_{s\geqslant 0}$ follows Langevin (1.9) or overdamped Langevin (1.8) dynamics with the potential V and δ^ε is a C^∞ approximation of the identity. Now, imagine that at a given time t, an approximation F_t^ε of F is used to bias the dynamics: the potential seen by the particles is $V - F_t^\varepsilon \circ \xi$ instead of V.

Then, a natural updating formula for F_t^ε is

$$\exp(-\beta F_t^\varepsilon(z)) = \int_0^t \delta^\varepsilon(\xi(q_s) - z) \frac{\exp(-\beta F_s^\varepsilon(\xi(q_s)))}{\int_{\mathbb{R}^m} \exp(-\beta F_s^\varepsilon(\tilde{z}))\mathrm{d}\tilde{z}} \, \mathrm{d}s, \tag{4.7}$$

for all $z \in \mathbb{R}^m$, where $(q_s)_{s\geqslant 0}$ now follows Langevin or overdamped Langevin dynamics with the biased potential $V - F_s^\varepsilon \circ \xi$ at time s. The factor

$$\frac{\exp(-\beta F_s^\varepsilon(\xi(q_s)))}{\int_{\mathbb{R}^m} \exp(-\beta F_s^\varepsilon(\tilde{z}))\,\mathrm{d}\tilde{z}}$$

is introduced in order to unbias the sample q_s which, if immediate equilibrium with respect to the biased potential were reached, would be distributed according to $Z_s^{-1} \exp[-\beta(V(q) - F_s^\varepsilon \circ \xi(q))]\,\mathrm{d}q$ (this is the unbiasing formula (3.37) mentioned above). The updating strategy (4.7) is one example of an ABP technique. We refer to Fort, Jourdain, Lelièvre and Stoltz (2015*b*) for more details on this specific example, and to Dickson *et al.* (2010) for a similar approach using a different normalization technique on F_s^ε. For other ABP methods, we refer to Wang and Landau (2001*a*, 2001*b*), Laio and Parrinello (2002), Bussi *et al.* (2006) and Marsili *et al.* (2006), and to Fort *et al.* (2014, 2015*a*, 2015*b*) for associated numerical analysis. The fundamental idea of all these ABP techniques is to penalize the states already visited in ξ in order to favour transitions to new regions in ξ. As the simulation continues, the penalization strength on F_t^ε tends to zero, and all the states in ξ are equally probable. In (4.7), the penalization strength is the multiplicative factor $\left(\int_{\mathbb{R}^m} \exp(-\beta F_s^\varepsilon(\tilde{z}))\,\mathrm{d}\tilde{z}\right)^{-1}$. Indeed, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} F_t^\varepsilon(z) = -\frac{1}{\beta} \delta^\varepsilon(\xi(q_t) - z) \frac{\exp(-\beta[F_t^\varepsilon(\xi(q_t)) - F_t^\varepsilon(z)])}{\int_{\mathbb{R}^m} \exp(-\beta F_t^\varepsilon(\tilde{z}))\,\mathrm{d}\tilde{z}}. \tag{4.8}$$

Because of the prefactor $\delta^\varepsilon(\xi(q_t) - z)$, the numerator

$$\exp(-\beta(F_t^\varepsilon(\xi(q_t)) - F_t^\varepsilon(z)))$$

is close to one. Further, it can be checked that $\left(\int_{\mathbb{R}^m} \exp(-\beta F_t^\varepsilon(\tilde{z}))\,\mathrm{d}\tilde{z}\right)^{-1}$ behaves as $1/t$ in the large t limit; see Fort *et al.* (2015*b*) for a precise statement in a discrete setting. In view of (4.8), this means that there is a vanishing rate of adaptation of the estimated free energy. More precisely, the adaptation rate decays as $1/t$ for the updating strategy (4.7). Other vanishing adaptation rates can be used while still ensuring convergence: see Fort *et al.* (2015*a*, 2015*b*).

The second strategy is to approximate ∇F instead of F. Such techniques are thus called *adaptive biasing force* (ABF) methods. Indeed, note that in Langevin or overdamped Langevin dynamics, only ∇F is needed and not F. To introduce this updating strategy, we need a formula for ∇F. To give a heuristic for this formula, let us go back to the two-dimensional example

of Section 1.2, where $q = (x, y) \in \mathbb{R}^2$ and $\xi(x, y) = x$. In this case, as mentioned above, we have

$$F(x) = -\beta^{-1} \ln\left(\int_{\mathbb{R}} \exp(-\beta V(x, y))\, \mathrm{d}y\right).$$

Thus, by a simple computation, we obtain

$$F'(x) = \int_{\mathbb{R}} \partial_x V(x, y) \frac{\exp(-\beta V(x, y))\, \mathrm{d}y}{\int_{\mathbb{R}} \exp(-\beta V(x, y))\, \mathrm{d}y}. \tag{4.9}$$

We thus see that $F'(x)$ is the average of $\partial_x V(q)$ with respect to the conditional probability measure

$$\nu(\cdot|\xi(q) = x) = \frac{\exp(-\beta V(x, y))\, \mathrm{d}y}{\int_{\mathbb{R}} \exp(-\beta V(x, y))\, \mathrm{d}y},$$

that is,

$$F'(x) = \mathbb{E}_\nu(\partial_x V(q)|\xi(q) = x).$$

This simple observation can be generalized to any reaction coordinate, and we have the following result (Sprik and Ciccoti 1998, den Otter and Briels 1998).

Lemma 4.3. Let $F : \mathbb{R}^m \to \mathbb{R}$ be the free energy associated with the Boltzmann–Gibbs distribution $\nu(\mathrm{d}q) = Z_\nu^{-1} \exp(-\beta V(q))\, \mathrm{d}q$ and the reaction coordinate $\xi : \mathbb{R}^d \to \mathbb{R}^m$, as introduced in Definition 4.1. Then,

$$\nabla F(z) = \int_{\Sigma(z)} f(q) \frac{\exp(-\beta V(q)) \delta_{\xi(q)-z}(\mathrm{d}q)}{\exp(-\beta F(z))}, \quad \text{for all } z \in \mathbb{R}^m, \tag{4.10}$$

with $f = (f_1, \ldots, f_m)$ defined as

$$f_\alpha = \sum_{\gamma=1}^m G^{-1}_{\alpha,\gamma} \nabla \xi_\gamma \cdot \nabla V - \beta^{-1} \operatorname{div}\left(\sum_{\gamma=1}^m G^{-1}_{\alpha,\gamma} \nabla \xi_\gamma\right), \quad \text{for all } \alpha \in \{1, \ldots, m\}, \tag{4.11}$$

where $G^{-1}_{\alpha,\gamma}$ is the (α, γ)-component of the inverse of the matrix G defined by (4.3).

In probabilistic terms, we have $\nabla F(z) = \mathbb{E}_\nu(f(q)|\xi(q) = z)$, where the family of conditional measures $\nu(\mathrm{d}q|\xi(q) = z)$ is defined in (4.4). The gradient field ∇F (or, depending on the convention at hand, its opposite) is called *the mean force*, while f (or its opposite) is called *the local mean force* or *the instantaneous force*. In the simple two-dimensional case mentioned above, when $\xi(x, y) = x$, we simply have $f = \partial_x V$, and (4.10) is consistent with the simple computation (4.9) made above. For a one-dimensional reaction

coordinate, equation (4.11) simplifies to

$$f = \frac{\nabla\xi}{|\nabla\xi|^2} \cdot \nabla V - \beta^{-1}\mathrm{div}\left(\frac{\nabla\xi}{|\nabla\xi|^2}\right). \tag{4.12}$$

The proof of Lemma 4.3 is a consequence of the identity $\delta_{\xi(q)-z}(\mathrm{d}q)\,\mathrm{d}z = \mathrm{d}q$ and can be read in Lelièvre *et al.* (2010, Section 3.2.2).

Adaptive biasing force methods rely on the formula (4.10) for ∇F, which we rewrite for convenience in the form

$$\nabla F(z) = \int_{\Sigma(z)} f(q)\,\frac{\exp(-\beta V(q))\delta_{\xi(q)-z}(\mathrm{d}q)}{\int_{\Sigma(z)}\exp(-\beta V(q))\delta_{\xi(q)-z}(\mathrm{d}q)}, \quad \text{for all } z \in \mathbb{R}^m.$$

The fundamental observation is that, whatever the biasing potential $F_t : \mathbb{R}^m \to \mathbb{R}$, we still obtain

$$\nabla F(z) = \int_{\Sigma(z)} f(q)\,\frac{\exp(-\beta(V - F_t\circ\xi)(q))\delta_{\xi(q)-z}(\mathrm{d}q)}{\int_{\Sigma(z)}\exp(-\beta(V - F_t\circ\xi)(q))\delta_{\xi(q)-z}(\mathrm{d}q)}. \tag{4.13}$$

for all $z \in \mathbb{R}^m$. This is because the conditional measure

$$\frac{\exp(-\beta(V - F_t\circ\xi)(q))\delta_{\xi(q)-z}(\mathrm{d}q)}{\int_{\Sigma(z)}\exp(-\beta(V - F_t\circ\xi)(q))\delta_{\xi(q)-z}(\mathrm{d}q)}$$

does not depend on F_t (the factor $\exp(\beta F_t\circ\xi)(q) = \exp(\beta F_t(z))$ cancels). This explains the principle of the updating formula for ABF approaches, which is to consider the biasing force

$$\Gamma_t(z) = \mathbb{E}(f(q_t)|\xi(q_t) = z) \tag{4.14}$$

at time t, where q_t follows Langevin or overdamped Langevin dynamics with the biased force

$$-\nabla V(q_t) + \sum_{\alpha=1}^{m}[\Gamma_t(\xi(q_t))]_\alpha \nabla\xi_\alpha(q_t)$$

at time t. We will give an example of such ABF dynamics below: see (4.18). The fundamental property of ABF strategies is thus to approximate the mean force in the visited states in ξ, and to use this mean force as a bias in order to favour transitions to new regions in ξ.

Note that ABP and ABF strategies are different in nature. Even though the formula (4.10) for ∇F is simply the derivative of the formula (4.1) for F, the updating formula (4.14) for ABF is not related to the derivative of the updating formula (4.7) for ABP (in particular, Γ_t is not a gradient field in general).

Such free energy adaptive biasing techniques are now used on a daily basis in the molecular dynamics community, in order to efficiently compute the free energy associated with a reaction coordinate (Chipot and Pohorille 2007). These approaches are in principle useful in other contexts, once

the sampling of a multimodal measure is involved (see for example Chopin, Lelièvre and Stoltz 2012 for an example in computational statistics, for Bayesian inference).

The practical implementation of all these ideas involves three types of approximations: (i) discretization in time of the dynamics, (ii) discretization in space of the values taken by the ξ-variable (at least if ξ takes continuous values), and (iii) an approximation procedure of the average values considered in both approaches. The first approximation is related to the numerical analysis of time discretization schemes, already discussed in Section 3.2. The second approximation is rather standard: classical projection techniques on finite-dimensional spaces can be used (*e.g.* piecewise constant functions, finite element approximations, spectral approaches), and standard numerical analysis approaches give the associated error. Approximation (iii) can be performed using two approaches: time averages over a trajectory following the biased dynamics (this is what was chosen in (4.7)) or averages over many replicas (also called walkers in this context) which all follow the biased dynamics with the same bias, but driven by independent noises. For example, the discretization of (4.14) in this case would be

$$\Gamma_t^{K,\varepsilon}(z) = \frac{\sum_{k=1}^K f(q_t^{k,K})\delta^\varepsilon(\xi(q_t^{k,K} - z))}{\sum_{k=1}^K \delta^\varepsilon(\xi(q_t^{k,K} - z))}, \tag{4.15}$$

where K is the total number of walkers, and $\varepsilon > 0$ is (as in (4.7)) the parameter of the C^∞ kernel δ^ε which approximates the Dirac mass. In (4.15), $q_t^{k,K}$ is the position of the kth replica. In practice, a combination of both time averages and averages over replicas can be used. From a numerical point of view, the interest of a multiple walker strategy is that an additional selection procedure can be used to duplicate walkers which visit new regions in ξ, and kill walkers which are in already well-sampled regions (see Lelièvre *et al.* 2007). In addition, as theoretically explored in Lelièvre and Minoukadeh (2011) and numerically illustrated in Minoukadeh, Chipot and Lelièvre (2010), it is expected that in a multiple channel situation (*i.e.*, when more than one pathway links two metastable states), the multiple walkers approach is more efficient since many channels are visited in parallel by the walkers. From a theoretical viewpoint, in the limit of infinitely many walkers (the mean field limit), one ends up with a system involving a conditional expectation as in (4.14); see Jourdain, Lelièvre and Roux (2010) for a study of the limit $K \to \infty$ and $\varepsilon \to 0$ starting from the approximation (4.15). The long-time behaviour of such a non-linear (in the sense of McKean) stochastic differential equation can be analysed using PDE techniques, as explained in the next section. On the other hand, the analysis of approximation techniques using time averages requires probabilistic tools (convergence of stochastic approximation algorithms) to properly analyse

the correlation in time along the trajectories which play a role in the time integrals (see Section 4.4 for a discussion of such analysis).

Remark 4.4 (ABP versus ABF strategies). The question of which of the variants of adaptive free energy biasing techniques is optimal is often raised. Let us make a few comments. The idea of ABP techniques is that they can easily handle reaction coordinates with discrete values, which are sometimes used in practice (to take into account coordination numbers, for example). ABP is also well suited to the case of purely entropic barriers while ABF would only see a zero local mean force in a completely flat landscape. On the other hand, in the case of continuous reaction coordinates, ABF seems more attractive. Indeed it approximates the biasing force directly (in ABP, one needs to differentiate F_t to get the force in the dynamics, and differentiating a quantity computed as an empirical average is always delicate because of statistical noise). In addition, in ABP approaches, there is typically an additional numerical parameter which governs the adaptation mechanism, and which decreases the penalization strength as time goes on. The rate of decay of this penalization strength should be chosen with care: if it is too fast, F_t may converge to the wrong value; if it is too slow, this also slows down the convergence of F_t to F (see the discussion in Fort *et al.* 2014, 2015*a*). This is not required in ABF. This is related to the following remark: if the free energy is actually a constant function along ξ, the ABF method will not adapt the potential (the conditional expectation will always be zero in (4.14)) while ABP will modify the potential (penalizing the visited states will imply a modification of F_t).

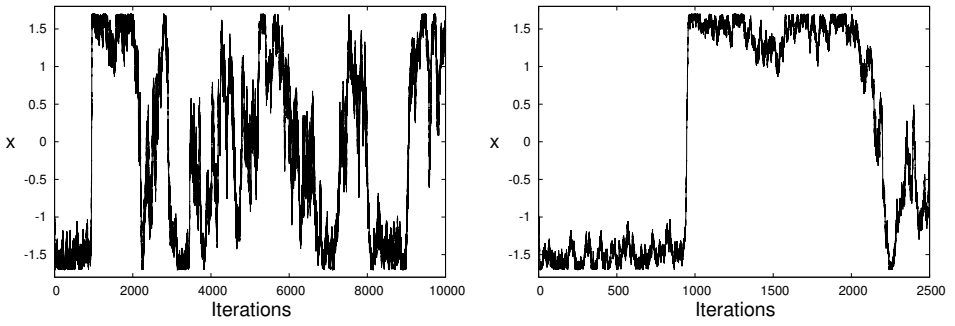
Remark 4.5 (back to the original sampling problem). By using free energy adaptive biasing strategies, in the long-time limit, the free energy biased measure $Z^{-1}\exp(-\beta(V - F\circ\xi)(q))\,\mathrm{d}q$ is sampled instead of the canonical measure ν. In order to recover averages with respect to ν, one can use for example the standard unbiasing strategy for importance sampling (see (3.37)):

$$\int \varphi \,\mathrm{d}\nu = \frac{\int \varphi e^{-\beta F\circ\xi}\,\frac{e^{-\beta(V-F\circ\xi)}}{Z}}{\int e^{-\beta F\circ\xi}\,\frac{e^{-\beta(V-F\circ\xi)}}{Z}}. \tag{4.16}$$

Let us mention that the unbiasing strategy (4.16) is very useful in practice, since it does not require F_t to have converged to F: once a sufficiently good approximation F_t of F has been obtained so that the dynamics visits the modes of ν efficiently, one can fix the biasing potential to this value F_t, and use the formula (4.16) with $F = F_t$ to get unbiased results.

4.2.3. Numerical illustration

We present a simple application of ABF to the model system discussed in Section 1.2. More precisely, let us consider the problem with an energetic



(a) (b)

Figure 4.2. (a) Trajectory of the ABF process on the model problem for an energetic barrier (see Figure 1.1), with $\beta = 80$. The evolution of $\xi(X_t)$ (*i.e.* the abscissa of X_t) is represented as a function of time. (b) Close-up of the first 2500 time units.

barrier presented in Section 1.2 (see Figure 1.1(a, b)), with $\xi(x, y) = x$ as a reaction coordinate, and $\beta = 80$.

In Figure 4.2, one can see that the ABF process goes through three stages. Starting from the left well (the abscissa is around -1.5), the process remains in this well for approximately 1000 iterations. During this first stage, the free energy derivative is approximated for ξ in the vicinity of -1.5 (*i.e.* in the left well). This enables us to quickly leave the left well to go to the right well. In a second stage, the free energy derivative is learned in the right well (the abscissa is around $+1.5$). This again requires about 1000 iterations. In the final stage, the process freely moves from left to right since the free energy derivative is well approximated over the whole range of the reaction coordinate (which is here fixed to $[-1.7, 1.7]$).

In addition, in Figure 4.3 we represent, as a function of β, the average exit time from the left well for the unbiased original overdamped Langevin dynamics, compared with the average exit time for the ABF process. One can see that the exit time is reduced by several orders of magnitude using ABF, which illustrates the ability of the method to escape from metastable states.

4.3. A convergence analysis using entropy techniques

In this section we present a typical example where the use of PDE techniques, and more precisely entropy estimates and logarithmic Sobolev inequalities (see Section 2.2.2), appear to be very useful for understanding the efficiency of a numerical method, namely the adaptive biasing force (ABF) method. For pedagogical purposes, the convergence result is presented in a very simple setting (the two-dimensional case and $\xi(x, y) = x$). Extensions are discussed in the next section.

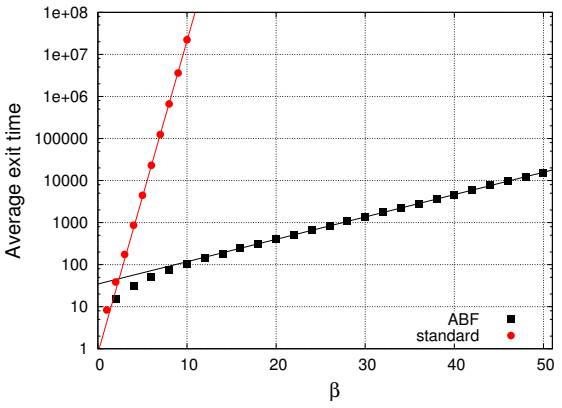


Figure 4.3. Average exit time from the left well as a function of the inverse temperature β, for the ABF process (squares) compared to the unbiased overdamped Langevin dynamics (circles: see also Figure 1.2(a)).

4.3.1. The theoretical setting and the convergence result

Let us again consider the two-dimensional setting of Section 1.2, where $q = (x, y)$ and $\xi(x, y) = x$. Let us assume in addition that q lives in the space $\mathbb{T} \times \mathbb{R}$ (which simply means that $V : (x, y) \in \mathbb{R}^2 \mapsto V(x, y) \in \mathbb{R}$ is 1-periodic in the x-variable). We will comment on these assumptions below. In this case, the mean force is given by $F'(x) = \mathbb{E}_\nu(\partial_x V(q)|\xi(q) = x)$ (see (4.9)). Thus, the free energy biased dynamics on $q_t = (x_t, y_t) \in \mathbb{T} \times \mathbb{R}$ is given by

$$\mathrm{d}x_t = -\partial_x V(x_t, y_t)\,\mathrm{d}t + F'(x_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t^x, \tag{4.17a}$$

$$\mathrm{d}y_t = -\partial_y V(x_t, y_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t^y, \tag{4.17b}$$

$$F'(x) = \mathbb{E}_\nu(\partial_x V(q)|\xi(q) = x). \tag{4.17c}$$

Equations (4.17a) and (4.17b) are simply overdamped Langevin dynamics on $q_t = (x_t, y_t)$ using the free energy biased potential $V(x, y) - F(x)$. Equation (4.17c) is the definition of the derivative of the free energy. As explained above (see (4.14)), the ABF strategy is to replace the measure ν in this definition by the law at time t of the pair (x_t, y_t), which yields

$$\mathrm{d}x_t = -\partial_x V(x_t, y_t)\,\mathrm{d}t + F_t'(x_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t^x, \tag{4.18a}$$

$$\mathrm{d}y_t = -\partial_y V(x_t, y_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t^y, \tag{4.18b}$$

$$F_t'(x) = \mathbb{E}(\partial_x V(x_t, y_t)|x_t = x). \tag{4.18c}$$

The key idea of the approach is that if the system were immediately at equilibrium with respect to the biased potential $V(x, y) - F_t(x)$, then F_t' would be equal to F' (see (4.13)). But the updating strategy is dynamic: F_t always keeps on changing and it is therefore not obvious how the system

behaves in the transient phase of the dynamics (4.18). The objective of the mathematical analysis is to be able to prove that F_t' indeed converges to F', and that, in addition, this convergence typically occurs on much smaller time scales than the convergence to equilibrium for the original simple overdamped Langevin dynamics:

$$\mathrm{d}q_t = -\nabla V(q_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t. \tag{4.19}$$

In order to study the rate of convergence to equilibrium, we will actually analyse the long-time behaviour of the density[11] $\psi(t,x,y)$ of the process (x_t,y_t) solution to (4.18), namely the Fokker–Planck equation associated with (4.18):

$$\begin{aligned}\partial_t\psi &= \partial_x[(\partial_x V - F_t')\psi] + \beta^{-1}\partial_{x,x}\psi \\ &\quad + \partial_y[(\partial_y V)\psi] + \beta^{-1}\partial_{y,y}\psi \qquad \text{for } (x,y)\in\mathbb{T}\times\mathbb{R},\end{aligned} \tag{4.20a}$$

$$F_t'(x) = \frac{\int_{\mathbb{R}}\partial_x V(x,y)\psi(t,x,y)\,\mathrm{d}y}{\int_{\mathbb{R}}\psi(t,x,y)\,\mathrm{d}y} \qquad \text{for } x\in\mathbb{T}. \tag{4.20b}$$

Equation (4.20b) is obtained using the fact that, for fixed time t and $x\in\mathbb{T}$, $\psi(t,x,y)\,\mathrm{d}y/\int_{\mathbb{R}}\psi(t,x,y)\,\mathrm{d}y$ is indeed the conditional law at time t of (x_t,y_t) given that $x_t = x$. In the following, we will let ψ^ξ denote the marginal of ψ along ξ (*i.e.* the density of x_t):

$$\psi^\xi(t,x) = \int_{\mathbb{R}}\psi(t,x,y)\,\mathrm{d}y.$$

In contrast to the Fokker–Planck equation for simple overdamped Langevin dynamics (4.19), that is,

$$\partial_t\phi = \mathrm{div}(\nabla V\phi) + \beta^{-1}\Delta\phi$$

(see (2.12)), the Fokker–Planck equation (4.20) is a *non-linear* parabolic equation. The non-linearity comes from the conditional expectation which appears in the drift term of the stochastic differential equation (4.18).

The following result is proved in Lelièvre, Rousset and Stoltz (2008) in a more general setting.

Theorem 4.6 (convergence of ABF). Let us consider the ABF dynamics (4.18) on the domain $\mathbb{T}\times\mathbb{R}$. Let us assume that

$$\begin{aligned}&\text{There exists } \rho>0 \text{ such that all the conditional}\\ &\text{measures } \nu(\mathrm{d}q|\xi(q)=x) \text{ satisfy LSI}(\rho)\end{aligned} \tag{4.21}$$

[11] The fact that the process (x_t,y_t) admits a density comes from the fact that the Brownian terms in (4.18) imply a smoothing effect on the law of the process.

and that

$$\text{The potential } V \text{ is such that } \|\partial_{x,y} V\|_{L^\infty(\mathbb{T}\times\mathbb{R})} < \infty. \tag{4.22}$$

Let us assume in addition and (without loss of generality) that $\rho \neq 4\pi^2$ and that $\psi^\xi(0,\cdot)$ is a positive function such that[12] $\int_{\mathbb{T}} |\partial_x \ln(\psi^\xi(0,\cdot))|^2 \psi^\xi(0,\cdot) < \infty$. Then there exists a positive constant C such that

$$\sqrt{\int_{\mathbb{T}} |F_t' - F'|^2} \leqslant C \exp(-\beta^{-1} \min(\rho, 4\pi^2) t), \quad \text{for all } t \geqslant 0. \tag{4.23}$$

In our specific context, note that the conditional measure $\nu(\mathrm{d}q|\xi(q) = x)$ is simply the measure

$$\frac{\exp(-\beta V(x,y))\,\mathrm{d}y}{\int_{\mathbb{R}} \exp(-\beta V(x,y))\,\mathrm{d}y}.$$

Notice in addition that if $\rho = 4\pi^2$ in (4.21), one can use this result by replacing ρ with any $\tilde{\rho} < \rho$, so that (4.23) holds in this case for any $\tilde{\rho} < 4\pi^2$. Finally, we will show below (see (4.24)) that ψ^ξ simply satisfies the heat equation on $\mathbb{T}$, so that the assumptions on $\psi^\xi(0,\cdot)$ are not too stringent: if they are not satisfied by $\psi^\xi(0,\cdot)$, one simply has to consider the problem on the time interval $[t_0, +\infty)$ with initial condition $\psi^\xi(t_0,\cdot)$, for a positive time t_0.

Let us comment on Theorem 4.6. This theorem shows that the free energy F_t indeed converges to F (up to an irrelevant additive constant) with rate

$$\beta^{-1} \min(\rho, 4\pi^2).$$

As will be more clear below, the limiting parameter here is ρ, the term $4\pi^2$ being only related to the rate of convergence to equilibrium of a simple diffusion on the torus $\mathbb{T}$. Further, we will actually show that the law of (x_t, y_t) (and not only F_t) converges to its equilibrium value at this rate (see (4.38) and the discussion below). This rate of convergence for ABF has to be compared with (2.29), which shows that the original simple overdamped Langevin dynamics (4.19) converges to equilibrium with rate

$$\beta^{-1} R,$$

where R is the logarithmic Sobolev constant of the measure ν. In summary, by using ABF, R has been replaced by ρ: the logarithmic Sobolev inequality constant of the measure ν has been replaced by the logarithmic Sobolev inequality constant of the measures $\nu(\mathrm{d}q|\xi(q) = x)$. If the function ξ is indeed a good index of the metastable features of the original dynamics,

[12] The quantity $\int_{\mathbb{T}} |\partial_x \ln(\psi^\xi(0,\cdot))|^2 \psi^\xi(0,\cdot)$ is simply $\mathcal{I}(\psi^\xi(0,\cdot)|\psi^\xi_\infty)$, where $\psi^\xi_\infty = 1$ is the long-time limit of $\psi^\xi(t,\cdot)$ and $\mathcal{I}$ the Fisher information (see (2.25)).

it is typically expected that $\rho \gg R$. This inequality is actually a way to quantify the somewhat vague ideas that a good reaction coordinate ξ should be such that 'the metastability of the process $(q_t)_{t\geqslant 0}$ is along ξ' or that 'the directions orthogonal to ξ are fast variables', so that equilibrium is quickly reached along those directions. For example, in the two-dimensional examples of Figure 1.1, the logarithmic Sobolev constant of the measure ν is typically very small, while, for a fixed value of ξ (where $\xi(x,y) = x$), the conditional measures $\nu(\mathrm{d}q|\xi(q) = x)$ are gentle unimodal measures with much larger associated logarithmic Sobolev constants. Unfortunately, logarithmic Sobolev constants are very delicate to evaluate, and it thus seems very difficult to turn this measure of the quality of ξ into a numerical procedure to construct good reaction coordinates. This result should thus be seen as a way to understand theoretically the conditions under which an ABF strategy will be efficient, the measure of efficiency being the rate of convergence to equilibrium.

4.3.2. Proof of Theorem 4.6

The proof of convergence is based on two ingredients: (i) the observation that $\psi^\xi(t,x)$ (which, we recall, is the density at time t of x_t) satisfies the simple diffusion equation

$$\partial_t \psi^\xi = \beta^{-1} \partial_{x,x} \psi^\xi \text{ on } \mathbb{T}, \tag{4.24}$$

which is again a manifestation of the fact that, by using ABF, the potential V is 'flattened' along the reaction coordinate ξ (see also equation (4.6)); (ii) entropy estimates in the vein of those explained in Section 2.2.2, and a two-scale decomposition of the entropy in the spirit of Otto and Reznikoff (2007), Grunewald, Otto, Villani and Westdickenberg (2009) and Lelièvre (2009). Let us now present these ideas in detail.

The first ingredient of the proof is based on a simple computation, which consists in looking at the evolution of the marginal of ψ along ξ, namely,

$$\psi^\xi(t,x) = \int_{\mathbb{R}} \psi(t,x,y)\,\mathrm{d}y.$$

By direct integration in y of (4.20), we obtain

$$\partial_t \psi^\xi(t,x) = \partial_x \left(\int_{\mathbb{R}} \partial_x V(x,y) \psi(t,x,y)\,\mathrm{d}y - F_t'(x)\psi^\xi(t,x) \right) + \beta^{-1} \partial_{x,x} \psi^\xi(t,x)$$

which yields (4.24) since

$$\int_{\mathbb{R}} \partial_x V(x,y) \psi(t,x,y)\,\mathrm{d}y - F_t'(x)\psi^\xi(t,x) = 0,$$

by the definition of $F_t'(x)$. This means that the law of x_t satisfies a simple

heat equation on $\mathbb{T}$, as for a simple Brownian motion on the torus.[13] This explains the interest of the ABF method: along ξ, the time marginal of the process evolves as if the potential V were perfectly flat. The convergence to equilibrium of $\psi^\xi(t,x)$ is thus very easy to analyse, and in particular we have the following convergence in terms of Fisher information, which will be useful later (see Lelièvre *et al.* 2008, Lemma 5.29 for a proof).

Lemma 4.7 (convergence of ψ^ξ). Let φ be the solution of the heat equation on the torus $\mathbb{T}$,

$$\partial_t \varphi = \beta^{-1} \partial_{x,x} \varphi \text{ on } \mathbb{T}, \tag{4.25}$$

with initial condition $\varphi(0,\cdot)$ such that

$$\int_{\mathbb{T}} \varphi(0,\cdot) = 1, \quad \varphi(0,\cdot) \geqslant 0, \quad \mathcal{I}(\varphi(0,\cdot)|\varphi_\infty) < \infty,$$

where $\varphi_\infty \equiv 1$ is the long-time limit of φ. Then, for all $t \geqslant 0$,

$$\mathcal{I}(\varphi(t,\cdot) \mid \varphi_\infty) \leqslant \mathcal{I}(\varphi(0,\cdot) \mid \varphi_\infty) \exp(-\beta^{-1} 8\pi^2 t). \tag{4.26}$$

We recall that

$$\mathcal{I}(\varphi(t,\cdot) \mid \varphi_\infty) = \int_{\mathbb{T}} \left| \partial_x \ln\left(\frac{\varphi(t,x)}{\varphi_\infty(x)}\right)\right|^2 \varphi(t,x)\,\mathrm{d}x$$

is the Fisher information of $\varphi(t,x)\,\mathrm{d}x$ with respect to $\varphi_\infty(x)\,\mathrm{d}x$ (see (2.25)). The rate of convergence to equilibrium for $\psi(t,x,y)$ is limited by the rate of convergence to equilibrium for $\psi^\xi(t,y)$: this explains why there is a $4\pi^2$ appearing in the right-hand side of (4.23). This $4\pi^2$ is simply the first non-zero eigenvalue of the Laplacian on the torus $\mathbb{T}$, which determines the rate of convergence to equilibrium for the heat equation (4.25) on $\mathbb{T}$.

To understand the convergence to equilibrium, it now remains to understand the convergence of the conditional measures $\psi(t,x,y)\,\mathrm{d}y/\psi^\xi(t,x)$. We will use entropy techniques to study their long-time behaviours (see Section 2.2.2). Here comes the second ingredient of the proof, namely a decomposition of the entropy in terms of marginal and conditional measures, very much inspired by Otto and Reznikoff (2007) and Grunewald *et al.* (2009). Let us explain this decomposition. We let ψ_∞ denote the expected long-time limit of $\psi(t,x,y)$, defined by

$$\psi_\infty(x,y) = Z^{-1} \exp(-\beta(V(x,y) - F(x))) \tag{4.27}$$

where

$$Z = \int_{\mathbb{T}^2} \exp(-\beta(V(x,y) - F(x)))\,\mathrm{d}x\,\mathrm{d}y.$$

[13] The reader should not be confused here: this does not mean that the law of the process $(x_t)_{t\geqslant 0}$ (on $\mathcal{C}(\mathbb{R}_+,\mathbb{T})$) is the same as the law of the process $(W_t)_{t\geqslant 0}$. Only the time marginals agree.

The marginal of ψ_∞ along ξ is

$$\psi_\infty^\xi(x) = Z^{-1} \int_{\mathbb{R}} \exp[-\beta(V(x,y) - F(x))]\,\mathrm{d}y = 1$$

and the conditional measure of $\psi_\infty(x,y)\,\mathrm{d}x\,\mathrm{d}y$ given $\xi(x,y) = x$ is

$$\frac{\psi_\infty(x,y)\,\mathrm{d}y}{\int_{\mathbb{R}} \psi_\infty(x,y)\,\mathrm{d}y} = \nu(\cdot|\xi(q) = x).$$

Likewise, as explained above, the marginal of ψ along ξ is

$$\psi^\xi(t,x) = \int_{\mathbb{R}} \psi(t,x,y)\,\mathrm{d}y,$$

and the conditional measure of $\psi(t,x,y)\,\mathrm{d}x\,\mathrm{d}y$ given that $\xi(x,y) = x$ is $\psi(t,x,y)\,\mathrm{d}y/\psi^\xi(t,x)$. Let us now introduce the *total entropy* (see (2.24))

$$E(t) = \mathcal{H}(\psi(t,\cdot)|\psi_\infty),$$

the *macroscopic entropy* (namely the relative entropy of the marginal laws)

$$E_M(t) = \mathcal{H}(\psi^\xi(t,\cdot)|\psi_\infty^\xi),$$

and the *microscopic entropy*

$$E_m(t) = \int_{\mathbb{T}} e_m(t,x)\psi^\xi(t,x)\,\mathrm{d}x, \tag{4.28}$$

where $e_m(t,x)$ is the relative entropy associated with the conditional measures:

$$e_m(t,x) = \mathcal{H}\left(\frac{\psi(t,x,y)\,\mathrm{d}y}{\psi^\xi(t,x)} \,\middle|\, \frac{\psi_\infty(x,y)\,\mathrm{d}y}{\psi_\infty^\xi(x)}\right). \tag{4.29}$$

It is straightforward to check the following relation.

Lemma 4.8 (extensivity of entropy).

$$E(t) = E_M(t) + E_m(t), \quad \text{for all } t \geqslant 0. \tag{4.30}$$

In order to measure the convergence of $\psi(t,\cdot)$ to ψ_∞, we will study the rate of convergence to zero of E. We already know that $E_M(t)$ converges exponentially fast to zero, that is,

$$E_M(t) \leqslant E_M(0)\exp(-\beta^{-1}8\pi^2 t), \quad \text{for all } t \geqslant 0. \tag{4.31}$$

since the logarithmic Sobolev constant of the uniform law on $\mathbb{T}$ is $4\pi^2$ (see for example Bakry *et al.* 2014, Proposition 5.7.5) and it is therefore sufficient to understand the convergence of $E_m(t)$ to zero. The convergence (4.31) can

also be seen as a consequence of Lemma 4.7 since

$$\begin{aligned} E_M(t) = \mathcal{H}(\psi(t,\cdot)|\psi_\infty) &= \beta^{-1} \int_t^\infty \mathcal{I}(\psi(t,\cdot)|\psi_\infty)\,\mathrm{d}t \\ &\leqslant \frac{1}{8\pi^2}\mathcal{I}(\psi(t,\cdot)|\psi_\infty) \\ &= -\frac{\beta}{8\pi^2}\frac{\mathrm{d}E_M}{\mathrm{d}t}(t). \end{aligned}$$

The proof of Theorem 4.6 is then conducted as follows. Starting from (4.30), we have[14]

$$\begin{aligned} \frac{\mathrm{d}E_m}{\mathrm{d}t} &= \frac{\mathrm{d}E}{\mathrm{d}t} - \frac{\mathrm{d}E_M}{\mathrm{d}t} \\ &= -\beta^{-1} \int_{\mathbb{T}} \int_{\mathbb{R}} \left| \nabla \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi + \int_{\mathbb{T}} \int_{\mathbb{R}} (F_t' - F')\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\psi \\ &\quad + \beta^{-1} \int_{\mathbb{T}} \left|\partial_x \ln\left(\frac{\psi^\xi}{\psi_\infty^\xi}\right)\right|^2 \psi^\xi \end{aligned} \tag{4.32}$$

which is easily obtained after some integration by parts, using the fact that ψ satisfies the following equation, which is equivalent to (4.20a):

$$\partial_t \psi = \beta^{-1}\mathrm{div}\left(\psi_\infty \nabla\left(\frac{\psi}{\psi_\infty}\right)\right) + \partial_x((F' - F_t')\psi).$$

We next rely on a very useful formula for $F_t' - F'$.

Lemma 4.9. For all $t \geqslant 0$, we have

$$\beta(F_t' - F') = \int_{\mathbb{T}} \partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\frac{\psi}{\psi^\xi}\,\mathrm{d}y - \partial_x \ln\left(\frac{\psi^\xi}{\psi_\infty^\xi}\right). \tag{4.33}$$

Proof. A simple computation gives (using the fact that $\psi_\infty^\xi = 1$)

$$\begin{aligned} &\int_{\mathbb{R}} \partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\frac{\psi}{\psi^\xi}\,\mathrm{d}y - \partial_x \ln\left(\frac{\psi^\xi}{\psi_\infty^\xi}\right) \\ &\quad = \int_{\mathbb{R}} \partial_x \ln\psi \frac{\psi}{\psi^\xi}\,\mathrm{d}y - \int_{\mathbb{R}} \partial_x \ln\psi_\infty \frac{\psi}{\psi^\xi}\,\mathrm{d}y - \partial_x \ln\psi^\xi \\ &\quad = \int_{\mathbb{R}} \frac{\partial_x \psi}{\psi^\xi}\,\mathrm{d}y + \beta \int_{\mathbb{R}} \partial_x(V - F)\frac{\psi}{\psi^\xi}\,\mathrm{d}y - \partial_x \ln\psi^\xi \\ &\quad = \beta(F_t' - F'), \end{aligned}$$

[14] Here and in the following, even though ψ_∞^ξ is simply the constant function equal to 1 in our setting, we keep the notation ψ_∞^ξ in the various expressions we obtain in order to emphasize the homogeneity of the resulting formulae, which remain the same for more general functions ξ and more general settings.

where in the last line we used

$$\partial_x \ln \psi^\xi = \frac{1}{\psi^\xi}\partial_x \int_{\mathbb{R}} \psi \,\mathrm{d}y = \int_{\mathbb{R}} \frac{\partial_x \psi}{\psi^\xi}\,\mathrm{d}y.$$

This concludes the proof. □

By using (4.33) in (4.32), we obtain

$$\begin{aligned}\frac{\mathrm{d}E_m}{\mathrm{d}t} &= -\beta^{-1}\int_{\mathbb{T}}\int_{\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi \\ &\quad - \beta^{-1}\int_{\mathbb{T}}\int_{\mathbb{R}}\left|\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi + \beta^{-1}\int_{\mathbb{T}}\left(\int_{\mathbb{R}} \partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\psi\,\mathrm{d}y\right)^2 \frac{1}{\psi^\xi}\,\mathrm{d}x \\ &\quad - \beta^{-1}\int_{\mathbb{T}}\int_{\mathbb{R}} \partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\psi + \beta^{-1}\int_{\mathbb{T}}\left|\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2 \psi^\xi.\end{aligned}$$

Note that by the Cauchy–Schwarz inequality, the sum of the two terms on the second line above is non-positive. Therefore, again using (4.33),

$$\frac{\mathrm{d}E_m}{\mathrm{d}t} \leqslant -\beta^{-1}\int_{\mathbb{T}}\int_{\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi - \int_{\mathbb{T}} \partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\psi^\xi (F_t' - F'). \quad (4.34)$$

Further, using assumption (4.21), we know that for any positive time t, and for any $x \in \mathbb{T}$,

$$\mathcal{H}\left(\frac{\psi(t,x,\cdot)}{\psi^\xi(t,x)}\middle|\frac{\psi_\infty(x,\cdot)}{\psi^\xi_\infty(x)}\right) \leqslant \frac{1}{2\rho}\mathcal{I}\left(\frac{\psi(t,x,\cdot)}{\psi^\xi(t,x)}\middle|\frac{\psi_\infty(x,\cdot)}{\psi^\xi_\infty(x)}\right),$$

which is given more explicitly by

$$e_m(t,x) \leqslant \frac{1}{2\rho}\int_{\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi/\psi^\xi}{\psi_\infty/\psi^\xi_\infty}\right)\right|^2 \frac{\psi}{\psi^\xi}\,\mathrm{d}y = \frac{1}{2\rho}\int_{\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \frac{\psi}{\psi^\xi}\,\mathrm{d}y$$

since ψ^ξ and ψ^ξ_∞ do not depend on y. Therefore

$$E_m(t) = \int_{\mathbb{T}} e_m(t,x)\psi^\xi(t,x)\,\mathrm{d}x \leqslant \frac{1}{2\rho}\int_{\mathbb{T}}\int_{\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi.$$

By using this in (4.34), we thus obtain

$$\frac{\mathrm{d}E_m}{\mathrm{d}t} \leqslant -2\beta^{-1}\rho E_m - \int_{\mathbb{T}} \partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\psi^\xi (F_t' - F'). \quad (4.35)$$

To continue, we now need an upper bound on $F_t' - F'$ in terms of the 'distance' (in some sense) between the two conditional measures

$$\psi(t,x,y)\,\mathrm{d}y/\psi^\xi(t,x) \quad \text{and} \quad \psi_\infty(x,y)\,\mathrm{d}y/\psi^\xi_\infty(x).$$

It seems natural that such an estimate holds since

$$F_t'(x) - F'(x) = \int_{\mathbb{R}} \partial_x V(x,y) \frac{\psi(t,x,y)\,\mathrm{d}y}{\psi^\xi(t,x)} - \int_{\mathbb{R}} \partial_x V(x,y) \frac{\psi_\infty(x,y)\,\mathrm{d}y}{\psi_\infty^\xi(x)}$$

is the difference between the averages of the same function $(\partial_x V)$ with respect to the two conditional measures. It is the purpose of the next lemma (which is probably the most technical part of the proof) to obtain such an estimate, the 'distance' being measured in terms of relative entropy.

Lemma 4.10. Suppose assumptions (4.21) and (4.22) are satisfied. Then, for all $t \geqslant 0$ and for all $x \in \mathbb{T}$,

$$|F_t'(x) - F'(x)| \leqslant \|\partial_{x,y} V\|_{L^\infty} \sqrt{\frac{2}{\rho} e_m(t,x)}, \tag{4.36}$$

where e_m is defined by (4.29).

Proof. The proof uses the Talagrand inequality, which relates the Wasserstein distance and the relative entropy between two probability measures. Let us first recall the Wasserstein distance and the Talagrand inequality.

Let us introduce the set of coupling measures $\Pi(\nu_{t,x}, \nu_{\infty,x})$, where, for ease of notation, we let

$$\nu_{t,x}(\mathrm{d}y) = \psi(t,x,y)\,\mathrm{d}y/\psi^\xi(t,x), \quad \nu_{\infty,x}(\mathrm{d}y) = \psi_\infty(x,y)\,\mathrm{d}y/\psi_\infty^\xi(x),$$

respectively, denote the conditional measures of

$$\psi(t,x,y)\,\mathrm{d}x\,\mathrm{d}y, \quad \psi_\infty(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

given that $\xi(x,y) = x$. By definition of $\Pi(\nu_{t,x}, \nu_{\infty,x})$, the measures $\pi \in \Pi(\nu_{t,x}, \nu_{\infty,x})$ are probability measures over $\mathbb{R} \times \mathbb{R}$ such that, for any C^∞ and compactly supported test function $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}\times\mathbb{R}} \varphi(y_1)\pi(\mathrm{d}y_1, \mathrm{d}y_2) = \int_{\mathbb{R}} \varphi(y_1)\nu_{t,x}(\mathrm{d}y_1)$$

and

$$\int_{\mathbb{R}\times\mathbb{R}} \varphi(y_2)\pi(\mathrm{d}y_1, \mathrm{d}y_2) = \int_{\mathbb{R}} \varphi(y_2)\nu_{\infty,x}(\mathrm{d}y_2).$$

In other words, the marginal of $\pi \in \Pi(\nu_{t,x}, \nu_{\infty,x})$ on the first (resp. the second) variable is $\nu_{t,x}$ (resp. $\nu_{\infty,x}$). Now, the Wasserstein distance between $\nu_{t,x}$ and $\nu_{\infty,x}$ is denoted by $\mathcal{W}(\nu_{t,x}, \nu_{\infty,x})$ where, for any measures π_1 and π_2 on $\mathbb{R}^d$, the Wasserstein distance between π_1 and π_2 is

$$\mathcal{W}(\pi_1, \pi_2) = \sqrt{\inf_{\pi \in \Pi(\pi_1,\pi_2)} \int_{\mathbb{R}\times\mathbb{R}} |y_1 - y_2|^2 \pi(\mathrm{d}y_1, \mathrm{d}y_2)}.$$

The Wasserstein distance appears very naturally when estimating the difference $F_t' - F'$. Indeed, for any coupling measure $\pi \in \Pi(\nu_{t,x}, \mu_{\infty,x})$, we have

$$\begin{aligned}|F_t'(x) - F'(x)| &= \left| \int_{\mathbb{R}\times\mathbb{R}} (\partial_x V(x,y) - \partial_x V(x,y'))\pi(\mathrm{d}y, \mathrm{d}y') \right| \\ &\leqslant \|\partial_{x,y} V\|_{L^\infty} \int_{\mathbb{R}\times\mathbb{R}} |y - y'| \pi(\mathrm{d}y, \mathrm{d}y') \\ &\leqslant \|\partial_{x,y} V\|_{L^\infty} \sqrt{\int_{\mathbb{R}\times\mathbb{R}} |y-y'|^2 \pi(\mathrm{d}y, \mathrm{d}y')}.\end{aligned}$$

Now taking the infimum over all $\pi \in \Pi(\nu_{t,x}, \nu_{\infty,x})$, we get

$$|F_t'(x) - F'(x)| \leqslant \|\partial_{x,y} V\|_{L^\infty} W(\nu_{t,x}, \nu_{\infty,x}). \tag{4.37}$$

This is where the assumption (4.22) is needed.

Now, in (4.35), it is not the Wasserstein distance between $\nu_{t,x}$ and $\nu_{\infty,x}$ which appears, but the relative entropy since in view of (4.28),

$$E_m = \int \mathcal{H}(\nu_{t,x}|\nu_{\infty,x}) \psi^\xi(t,x)\,\mathrm{d}x.$$

The Talagrand inequality is exactly what we need since it relates the Wasserstein distance to the relative entropy between two measures. Bobkov and Götze (1999) and Otto and Villani (2000) have proved the following result (though in a more general setting).

Lemma 4.11 (Talagrand inequality). Let π_2 be a probability measure on $\mathbb{R}^d$ which satisfies LSI(ρ). Then, for all probability measures π_1 on $\mathbb{R}^d$,

$$\mathcal{W}(\pi_1, \pi_2) \leqslant \sqrt{\frac{2}{\rho}} \mathcal{H}(\pi_1|\pi_2).$$

Using this result and Assumption (4.21), which exactly states that $\nu_{\infty,x} = \nu(\mathrm{d}q|\xi(q) = x)$ satisfies LSI(ρ), we thus have

$$\mathcal{W}(\nu_{t,x}, \nu_{\infty,x}) \leqslant \sqrt{\frac{2}{\rho}} \mathcal{H}(\nu_{t,x}|\nu_{\infty,x}), \quad \text{for all } t \geqslant 0.$$

By using this inequality in (4.37), we thus obtain (4.36). This concludes the proof of Lemma 4.10. □

Using (4.36) in (4.35), we obtain, using Cauchy–Schwarz,

$$\begin{aligned}\frac{\mathrm{d}E_m}{\mathrm{d}t} &\leqslant -2\beta^{-1}\rho E_m + \sqrt{\int_{\mathbb{T}} \psi^\xi |F_t' - F'|^2} \sqrt{\int_{\mathbb{T}} \left| \partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2 \psi^\xi} \\ &\leqslant -2\beta^{-1}\rho E_m + \|\partial_{x,y} V\|_{L^\infty} \sqrt{\frac{2}{\rho}} \sqrt{E_m} \sqrt{\mathcal{I}(\psi^\xi|\psi^\xi_\infty)}.\end{aligned}$$

By dividing by $2\sqrt{E_m}$ and using the estimate (4.26) on $\mathcal{I}(\psi^\xi|\psi^\xi_\infty)$, we thus obtain

$$\frac{\mathrm{d}\sqrt{E_m}}{\mathrm{d}t} \leqslant -\beta^{-1}\rho\sqrt{E_m} + \|\partial_{x,y}V\|_{L^\infty}\sqrt{\frac{1}{2\rho}\mathcal{I}(\psi^\xi(0,\cdot)|\psi^\xi_\infty)}\exp(-\beta^{-1}4\pi^2 t).$$

Using the fact that $\rho \neq 4\pi^2$, a Gronwall lemma[15] then easily shows that there exists a constant C such that, for all $t > 0$,

$$\sqrt{E_m(t)} \leqslant C\exp(-\beta^{-1}\min(\rho, 4\pi^2)t). \tag{4.38}$$

From this result, the extensivity of the entropy (4.30), and the convergence of the marginals in x (4.31), one immediately deduces that $\sqrt{E(t)}$ (and thus $\|\psi(t,\cdot) - \psi_\infty\|_{L^1(\mathbb{T}\times\mathbb{R})}$ by the Csiszár–Kullback inequality (2.27)) converges exponentially fast to zero with rate $\beta^{-1}\min(\rho, 4\pi^2)$.

To deduce from these results the convergence of

$$\sqrt{\int_{\mathbb{T}} |F_t' - F'|^2}$$

exponentially fast with the same rate, one simply uses the fact that

$$\sqrt{\int_{\mathbb{T}} |F_t' - F'|^2} \leqslant C\sqrt{\int_{\mathbb{T}} |F_t' - F'|^2\psi^\xi} \leqslant C\sqrt{E_m(t)}.$$

The first inequality is a consequence of the fact that $\psi^\xi(0,\cdot)$ is assumed to be positive and ψ^ξ satisfies the heat equation on $\mathbb{T}$, so that ψ^ξ is bounded from below by a positive constant. The second inequality is a consequence of (4.36). This concludes the proof of Theorem 4.6.

4.4. Extensions and related works

4.4.1. Extensions of Theorem 4.6

Let us make a few comments on the specific setting in which we presented the convergence result of Theorem 4.6. First, we assumed that the x-variable lives in the torus in order for ξ to be with values in a compact domain, so that the free energy biased Boltzmann–Gibbs measure with density $\exp(-\beta(V - F\circ\xi))$ can be normalized. The fact that $\int \exp(-\beta(V - F\circ\xi)) < \infty$ is necessary for the dynamics to actually have a stationary state and this has already been discussed above (see (4.6)). In practice, if ξ does not take values in a compact domain, the classical technique is to apply the bias only over a compact domain, by using a restraining potential $\mathcal{W}\circ\xi$ (see Lelièvre *et al.* 2008, Alrachid and Lelièvre 2015 for more details). Second, we assumed that ξ is a linear function of q. Let us consider the case of more

[15] If $\rho = 4\pi^2$, one obtains an upper bound of the form $(C_1 + C_2 t)\exp(-\beta^{-1}4\pi^2 t)$.

general reaction coordinates $\xi : \mathbb{R}^d \to \mathbb{R}$ (we consider one-dimensional ξ for simplicity). For general reaction coordinates, one in fact has to slightly modify the ABF dynamics (4.18) by terms which depend on $|\nabla\xi|$ in order to recover the results of Theorem 4.6 (see Lelièvre *et al.* 2008, equation (10)). These modifications are only necessary for theoretical purposes: they are not used by practitioners. However, from a theoretical viewpoint, without these additional terms we are only able to prove a weaker result (see Lelièvre *et al.* 2008, Section 2.3). With these additional terms, the results of Theorem 4.6 can be proved by rewriting (4.22) as

$$\|\nabla_{\Sigma(\xi(q))} f(q)\|_{L^\infty} < \infty \tag{4.39}$$

(where $\nabla_{\Sigma(z)}$ denotes the surface gradient in $\Sigma(z)$ and f is the local mean force: see (4.12)) and by assuming in addition that there exists $m < \infty$ such that

$$0 < |\nabla\xi(q)| \leqslant m, \quad \text{for all } q \in \mathbb{R}^d.$$

Further, assumption (4.21) can be somewhat weakened. Lelièvre and Minoukadeh (2011) consider the situation when two channels link metastable states, so that the assumption (4.21) is only satisfied on some bounded subinterval of the range of ξ. In this situation, it can be shown that if the free energy profile does not differ too much in each channel, then a similar convergence result can be obtained.

Finally, Alrachid and Lelièvre (2015) have proved the convergence of a variant of the ABF method, where the biasing force is projected onto a gradient field (which makes sense since the expected limit, the mean force, is indeed a gradient).

4.4.2. Related works

The pair of assumptions (4.21) and (4.22) seem to be very useful in other contexts. In Otto and Reznikoff (2007) and Lelièvre (2009), it appears as a set of assumptions in order to prove that a logarithmic Sobolev inequality can be derived for a measure μ if, for a given function ξ, it holds on the marginal probability measure $(\xi * \mu)(\mathrm{d}q)$, on the conditional probability measures $\mu(\mathrm{d}q|\xi(q) = z)$ (which corresponds to (4.21)) and if a so-called bounded coupling assumption of the form (4.39) holds.

This set of assumptions also naturally appears in order to measure the error of a coarse-graining technique, which yields an effective Markov dynamics which approximates the dynamics of $(\xi(q_t))_{t\geqslant 0}$. The aim is thus to reduce the dimension of the original problem by deriving an effective dynamics on the reaction coordinate. This is closely related to the question of the relevance of the free energy to approximate dynamical quantities such as transition times (and not only to compute thermodynamic quantities). We refer to Legoll and Lelièvre (2010) for more results along these lines.

Similar assumptions are also used for Markov chains (in a discrete setting); see for example Madras and Randall (2002) and Jerrum, Son, Tetali and Vigoda (2004).

4.4.3. Open problems

Among the open problems related to free energy adaptive biasing techniques, the extension of the efficiency analysis to Langevin dynamics (as opposed to overdamped Langevin dynamics) remains to be done. As already explained in Section 2, the entropy technique does not apply to Langevin dynamics because of the lack of ellipticity. Hypocoercivity is then a good way to study the long-time behaviour of Langevin dynamics, and it would be interesting to extend the standard results on linear Langevin dynamics presented in Section 2.3 to the non-linear setting of adaptive biasing methods.

Let us also mention that we lack a good understanding of the efficiency of free energy adaptive techniques when time averages are used (instead of averages over multiple replicas) to compute the biasing force. We refer to Fort *et al.* (2015*a*, 2015*b*) for convergence results for the Wang–Landau and self-healing umbrella sampling algorithms, and to Fort *et al.* (2014) for a preliminary study of the efficiency of the Wang–Landau dynamics. See also Benaïm and Bréhier (2016) for recent results on ABP methods.

4.5. Other free energy computation techniques

In this section we have focused on adaptive importance sampling techniques for calculations of free energy differences. As mentioned above, such computations are actually the aim of many molecular dynamics simulations today (see Chipot and Pohorille 2007), and many techniques have been proposed to this end. We refer to Lelièvre *et al.* (2010) for an extensive presentation of such techniques. In this article we focus on adaptive importance sampling techniques, since their analysis typically relies on partial differential equation techniques. In practice, adaptive importance sampling approaches are very popular since they do not require many tuning parameters, and they appear to be rather efficient for a large class of systems. However, let us mention here the three other main classes of techniques for free energy calculations, aside from adaptive importance sampling methods.

Thermodynamic integration (Kirkwood 1935, Carter, Ciccotti, Hynes and Kapral 1989) consists in computing the free energy derivative, and then integrating this derivative to obtain free energy differences. It is based on the formula (4.10). Let us consider the case of a one-dimensional reaction coordinate for simplicity (the generalization to $m > 1$ is straightforward). The principle is to compute $F'(z_i)$ at some points $(z_1, \ldots, z_k)$ and then to use a numerical quadrature to approximate $F(z_k) - F(z_1) = \int_{z_1}^{z_k} F'(z)\,\mathrm{d}z$

by a weighted sum over the values $F'(z_i)$. This requires us to sample the conditional probability measures $\nu(\mathrm{d}q|\xi(q) = z_i)$, which can be done using Langevin or overdamped Langevin dynamics projected onto the submanifold $\Sigma(z_i)$. We refer to Carter *et al.* (1989), Ciccotti, Lelièvre and Vanden-Eijnden (2008) and Lelièvre, Rousset and Stoltz (2012) for more details of the sampling of such conditional measures, and to Lelièvre *et al.* (2010, Chapter 3) for a general presentation of thermodynamic integration methods. Let us mention that the sampling of measures supported on submanifolds (such as the conditional measures $\nu(\mathrm{d}q|\xi(q) = z_i)$) is actually a very important subject in practice. Apart from thermodynamic integration, this question naturally appears when considering molecular systems with constraints: for example, a covalent bond is sometimes modelled using a fixed bond length instead of a stiff potential, which can be useful in diminishing the stability constraint on the time step (see Remark 3.6). In such cases, the canonical measure is defined on a submanifold of $\mathbb{R}^{3N}$ or $(L\mathbb{T})^{3N}$, and the differential operators associated with the ergodic stochastic processes are differential operators on manifolds. For more details on sampling methods with constraints, we refer to Lelièvre *et al.* (2012).

Another free energy calculation method consists in using statistical approaches to estimate directly the marginal density $\exp(-\beta F(z))$. Examples include the *free energy perturbation method* introduced in Zwanzig (1954), and *histogram methods*, such as the MBAR method (Shirts and Chodera 2008). We refer to Lelièvre *et al.* (2010, Sections 2.4 and 2.5) for more details and references.

A more recent class of methods relies on dynamics with an imposed schedule for the reaction coordinate. These technique therefore use *non-equilibrium dynamics*. With suitable exponential reweighting, however, equilibrium properties (such as free energy differences) can be recovered from the non-equilibrium trajectories: see Jarzynski (1997). This is explained in Proposition 6.2 below in a slightly different setting (the so-called alchemical setting); see also Lelièvre *et al.* (2010, Chapter 4).

5. Non-equilibrium systems and transport coefficients

In this section we discuss how to compute transport coefficients such as the mobility, the thermal conductivity or the shear viscosity. At the macroscopic level, transport coefficients relate an external forcing acting on the system (*e.g.* electric field, temperature gradient, velocity field) to an average response expressed through some steady-state flux (*i.e.*, of charged particles, energy, momentum). At the microscopic level, this is modelled by systems in a stationary state, evolving according to perturbations of equilibrium dynamics. Here we consider two paradigmatic situations.

(1) Langevin dynamics perturbed by a constant force term:

$$\mathrm{d}q_t = M^{-1} p_t \, \mathrm{d}t, \tag{5.1a}$$

$$\mathrm{d}p_t = (-\nabla V(q_t) + \eta F) \, \mathrm{d}t - \gamma M^{-1} p_t \, \mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}} \, \mathrm{d}W_t, \tag{5.1b}$$

where $(q_t, p_t) \in \mathcal{E} = \mathbb{T}^d \times \mathbb{R}^d$, $F \in \mathbb{R}^d$, with $|F| = 1$ a given direction, and V a C^∞ periodic potential. The parameter $\eta \in \mathbb{R}$ determines the strength of the external forcing. A non-zero velocity in the direction F is expected in the steady state due to the external force F. Let us emphasize that F does not derive from the gradient of a periodic function: it indeed holds that $F = -\nabla W_F(q)$ with $W_F(q) = -F^T q$, but the function W_F is not periodic. It is precisely because the perturbation is not of gradient type that some particle flux can appear in the steady state.

(2) Langevin dynamics with modified fluctuation:

$$\mathrm{d}q_t = M^{-1} p_t \, \mathrm{d}t, \tag{5.2a}$$

$$\mathrm{d}p_t = -\nabla V(q_t) \, \mathrm{d}t - \gamma M^{-1} p_t \, \mathrm{d}t + \sqrt{2\gamma T_\eta(q)} \, \mathrm{d}W_t, \tag{5.2b}$$

where the temperature $T_\eta : \mathcal{D} \to \mathbb{R}_+$ is a non-negative C^∞ function, of the form

$$T_\eta(q) = T_{\mathrm{ref}} + \eta \widetilde{T}(q)$$

for some C^∞ function $\widetilde{T}$ and a given reference temperature $T_{\mathrm{ref}} > 0$. In order for T_η to remain non-negative, the parameter $\eta \in \mathbb{R}$ should be sufficiently small. Typically, $\widetilde{T}$ is constant and positive on a subdomain $\mathcal{D}_+ \subset \mathcal{D}$, and constant and negative on another subdomain $\mathcal{D}_- \subset \mathcal{D}$, with some C^∞ transition between $\mathcal{D}_+$ and $\mathcal{D}_-$. Some energy flux is expected from the 'hot' part $\mathcal{D}_+$ of the system to the 'cold' one $\mathcal{D}_-$. The model (5.2) we consider here for pedagogical purposes is a simplification of more realistic models of thermal transport such as heat transport in one-dimensional chains; see the model presented in Olla and Letizia (2015), as well as the more general review articles by Bonetto, Lebowitz and Rey-Bellet (2000), Lepri, Livi and Politi (2003) and Dhar (2008).

The two dynamics (5.1) and (5.2) reduce to the standard Langevin dynamics (1.9) when $\eta = 0$. Most of our analysis is illustrated with the dynamics (5.1), but we will occasionally refer to the dynamics (5.2) as well.

As discussed more precisely in Section 5.1, dynamics such as (5.1) and (5.2) model non-equilibrium systems since they are non reversible: the law of forward trajectories is different from the law of backward trajectories.

From a physical point of view, the arrow of time can be read off the trajectories. We do not make a distinction here between non-reversible and non-equilibrium systems, although such a distinction is sometimes made in the physical literature; see for instance (Bertini *et al.* 2015) and references therein.

It is observed that, in general, the response of the system, as encoded by the steady-state average of a physical observable (such as the velocity for (5.1) or the energy flux for (5.2)), is proportional to the strength η of the forcing for small values of η. This corresponds to the so-called linear response regime. By definition, transport coefficients are the proportionality constants relating the response to the forcing. For (5.1), the velocity of the particle in the direction of F at stationary state is proportional to η, and the proportionality constant is called the mobility. For (5.2), the energy flux at stationary state is proportional to η, and the proportionality constant is called the thermal conductivity. It turns out that this linear response constant can in fact be rewritten as some integrated correlation function for an equilibrium dynamics, a celebrated equality known as the Green–Kubo formula (see Proposition 5.4 below).

We start with a general presentation of non-equilibrium dynamics in Section 5.1, with some emphasis on perturbations of equilibrium dynamics such as (5.1) and (5.2). We next show in Section 5.2 how first-order changes in average properties with respect to some forcing parameter (namely η in (5.1) and (5.2)) can be computed, and how these quantities are related to transport coefficients. Error estimates on the computation of transport coefficients are provided in Section 5.3. Finally, we discuss variance reduction techniques in Section 5.4.

5.1. Definition of non-equilibrium dynamics

5.1.1. Abstract characterization of non-equilibrium dynamics

Consider a general stochastic dynamics such as (3.1):

$$\mathrm{d}x_t = b(x_t)\,\mathrm{d}t + \sigma(x_t)\,\mathrm{d}W_t.$$

Recall that the existence and uniqueness of an invariant probability measure π can be proved in such a general setting with the techniques provided in Section 2.4, for instance. From a mathematical viewpoint, equilibrium systems are characterized by the self-adjointness of the generator $\mathcal{L}$ on the weighted Hilbert space $L^2(\pi)$:

$$\int_{\mathcal{X}} (\mathcal{L}f)\, g\,\mathrm{d}\pi = \int_{\mathcal{X}} f\,(\mathcal{L}g)\,\mathrm{d}\pi. \tag{5.3}$$

This expresses the reversibility of the dynamics with respect to the invariant measure of the process. A more probabilistic reformulation of the reversibility is as follows: when $x_0 \sim \pi$, the law of the forward paths $(x_s)_{0\leqslant s\leqslant t}$ is

the same as the law of the backward paths $(x_{t-s})_{0\leqslant s\leqslant t}$ (note that $x_t \sim \pi$ by the invariance of π). Therefore, the arrow of time cannot be read off the trajectories.

In some cases, the reversibility property holds only up to a one-to-one transformation preserving the invariant measure. For example, for Langevin dynamics, reversibility is valid only upon momentum reversal $S(q,p) = (q,-p)$:

$$\int_{\mathcal{X}} (\mathcal{L}f)\, g \,\mathrm{d}\pi = \int_{\mathcal{X}} (f\circ S)\, (\mathcal{L}(g\circ S))\,\mathrm{d}\pi;$$

see Lelièvre *et al.* (2010, Section 2.2.3.1). At the level of trajectories, this means that the law of the forward paths $(q_s,p_s)_{0\leqslant s\leqslant t}$ is the same as the law of the paths $(q_{t-s},-p_{t-s})_{0\leqslant s\leqslant t}$.

In this work, we define non-equilibrium dynamics to be stochastic evolutions for which reversibility properties such as (5.3) no longer hold true. The non-reversibility can be quantified by the entropy production, for which fluctuation theorems hold (Gallavotti and Cohen 1995, Kurchan 1998, Lebowitz and Spohn 1999).

5.1.2. Invariant measures

An important property of non-equilibrium systems is that their invariant measures are in general not analytically known, in contrast to equilibrium dynamics. In addition, the invariant measure depends non-trivially on the details of the dynamics due to long-range correlations, which are generally present in non-equilibrium systems (see for instance Derrida, Lebowitz and Speer 2002). Let us make the latter statement precise for overdamped Langevin dynamics on the compact configuration space $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, for a C^∞ periodic potential V and $\beta = 1$. For the reversible dynamics

$$\mathrm{d}q_t = -V'(q_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \tag{5.4}$$

the unique invariant probability measure is $Z^{-1}\mathrm{e}^{-V(q)}\,\mathrm{d}q$, which depends only on the value of V at the configuration q of interest (apart from a global normalization constant). For the perturbed dynamics

$$\mathrm{d}q_t = (-V'(q_t) + F)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \tag{5.5}$$

where $F \in \mathbb{R}$ is a constant force, it can be shown that there exists a unique invariant probability measure $\psi_F(q)\,\mathrm{d}q$. Indeed, the density ψ_F satisfies the stationary Fokker–Planck equation

$$\frac{\mathrm{d}}{\mathrm{d}q}\left((V'-F)\psi_F + \frac{\mathrm{d}\psi_F}{\mathrm{d}q}\right) = 0, \tag{5.6}$$

which can be solved as

$$\psi_F(q) = Z_F^{-1}\int_{\mathbb{T}} \mathrm{e}^{V(q+y)-V(q)-Fy}\,\mathrm{d}y, \tag{5.7}$$

with Z_F chosen such that $\int_0^1 \psi_F = 1$. It is clear from the expression of the invariant measure that, when $F \neq 0$, the invariant measure depends on the values of V everywhere.

Let us also check that the dynamics (5.5) is a non-equilibrium dynamics by checking that its generator is not reversible on $L^2(\psi_F)$. Since we do not need the precise expression of the invariant measure, we generalize the dynamics in arbitrary dimension, and consider $\mathcal{D} = \mathbb{T}^d$. We first rewrite ψ_F in exponential form as

$$\psi_F(q) = \mathrm{e}^{-U_F(q)},$$

and introduce a stochastic dynamics with a general drift $b(q)$:

$$\mathrm{d}q_t = b(q_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t. \tag{5.8}$$

The generator $\mathcal{L}_b$ of this generalized dynamics reads $\mathcal{L}_b = b \cdot \nabla + \Delta$. A simple computation shows that

$$\begin{aligned}\int_{\mathcal{D}} (\mathcal{L}_b f) g\,\psi_F = \int_{\mathcal{D}} f(\mathcal{L}_b g)\,\psi_F - 2\int_{\mathcal{D}} f[(b + \nabla U_F)\cdot\nabla g]\,\psi_F \\ + \int_{\mathcal{D}} fg[-\mathrm{div}(b) - \Delta U_F + \nabla U_F \cdot (b + \nabla U_F)]\,\psi_F.\end{aligned}$$

When $b(q) = -\nabla V(q) + F$, reversibility holds if and only if

$$F + \nabla(U_F - V) = 0.$$

This condition cannot be satisfied since F does not derive from the gradient of a periodic function. More generally, the above computation shows that dynamics such as (5.8) are reversible if and only if the drift is the gradient of a potential energy function.

5.1.3. Computation of transport coefficients

A transport coefficient ρ relates the magnitude of the response of the system in its steady state (an average current) to the magnitude of the external forcing. We present a specific example in Section 5.2; see in particular (5.10). For the paradigmatic dynamics (5.1) and (5.2), the magnitude of the external forcing is η.

Before embarking on a more detailed analysis, it is useful to classify the current methods for computing transport coefficients, as reviewed in Evans and Morriss (2008) and Tuckerman (2010) for instance, into three main classes.

(i) *Equilibrium techniques* based on Green–Kubo formulae, which are integrated correlation functions of the general form

$$\rho = \int_0^{+\infty} \mathbb{E}_\pi(\varphi(x_t)\phi(x_0))\,\mathrm{d}t,$$

where φ, ϕ are two observables whose expressions depend on the physical context at hand, and where the expectation is taken with respect to all initial conditions distributed according to an invariant probability measure π for the reference dynamics $(x_t)_{t\geqslant 0}$, and for all realizations of this dynamics.

(ii) *Transient methods*, where the system is initially locally perturbed, and the relaxation of this perturbation is monitored as a function of time. The comparison with some assumed macroscopic evolution equation (for instance the heat equation for thermal transport) allows us to identify the physical parameters of the macroscopic evolution (such as the thermal conductivity). See Hulse, Howley and Wilding (2005), for example, for an application of this technique.

(iii) *Non-equilibrium steady-state techniques*, where a forcing is permanently applied to the system. The latter methods can be decomposed into two subcategories: boundary-driven techniques, where the external forcing is imposed only in boundary regions (think of (5.2) with a perturbation $\widetilde{T}$ localized in two subdomains $\mathcal{D}_-, \mathcal{D}_+$), and bulk-driven dynamics, where the perturbation is experienced everywhere in the system (think of (5.1)). In both cases a flux is measured, and the transport coefficient is obtained as the average flux divided by the magnitude η of the external forcing. The expression of the flux function is again defined by analogy with macroscopic laws.

Bulk dynamics are often numerically more efficient since the forcing is applied globally to the system, and therefore the steady state can be reached more rapidly. Further, it is in general impossible to prove the existence and uniqueness of an invariant probability measure for boundary-driven dynamics, except in very simple geometries such as one-dimensional atom chains, or for stochastic lattice gases.

It should be emphasized that the definition of transport coefficients is based on an analogy with macroscopic evolution equations, which are (a system of) partial differential equations. This is clear for transient and steady-state dynamics. It is in fact also the case for equilibrium methods, since the expression of the transport coefficient as some integrated correlation function is, up to algebraic manipulations, a straightforward consequence of linear response results for steady-state non-equilibrium dynamics (see for instance (5.18) below for the specific case treated in Section 5.2).

Note that, in the simplest cases, it is in fact possible to rigorously derive the macroscopic evolution equations from microscopic dynamics. The corresponding system of PDEs is known as the hydrodynamic limit; see for instance Kipnis and Landim (1999) for a pedagogical introduction.

5.2. Linear response for non-equilibrium dynamics

In this section we sketch the derivation of the expression of transport coefficients, in the paradigmatic case of the mobility, which is computed with the dynamics (5.1). The extension to more general dynamics may require more elaborate frameworks, such as the one provided in Hairer and Majda (2010). We let $\mathcal{L}_\eta = \mathcal{L}_0 + \eta\widetilde{\mathcal{L}}$ denote the generator, where $\widetilde{\mathcal{L}} = F \cdot \nabla_p$, and $\mathcal{L}_0$ is given in (2.34):

$$\mathcal{L}_0 = p^T M^{-1} \nabla_q - \nabla V^T \nabla_p + \gamma\left(-p^T M^{-1}\nabla_p + \frac{1}{\beta}\Delta_p\right).$$

5.2.1. Existence and uniqueness of the invariant measure

The first observation is that, by a simple extension of the results of Section 2.4, the dynamics (5.1) admits a unique invariant probability measure. Upon introducing the Lyapunov functions $\mathcal{K}_n(q,p) = 1 + |p|^n$ for $n \geqslant 2$, it is in fact possible to formulate the following convergence result, where some uniformity in the parameter η holds; see Leimkuhler *et al.* (2015) and Joubaud, Pavliotis and Stoltz (2015) for related results.

Proposition 5.1. Consider $\eta_* > 0$. For any $\eta \in [-\eta_*, \eta_*]$, the dynamics (5.1) admits a unique invariant probability measure with a C^∞ density $\psi_\eta(q,p)$ with respect to the Lebesgue measure. Moreover, for any $n \geqslant 2$, there exist $C_n, \lambda_n > 0$ (depending on η_*) such that, for any $\eta \in [-\eta_*, \eta_*]$ and for any $\varphi \in L^\infty_{\mathcal{K}_n}(\mathcal{E})$,

$$\left\| \mathrm{e}^{t\mathcal{L}_\eta}\varphi - \int_{\mathcal{E}} \varphi\psi_\eta \right\|_{L^\infty_{\mathcal{K}_n}} \leqslant C_n\, \mathrm{e}^{-\lambda_n t} \|\varphi\|_{L^\infty_{\mathcal{K}_n}}, \quad \text{for all } t \geqslant 0.$$

Recall that the density of the invariant measure satisfies the Fokker–Planck equation

$$\mathcal{L}_\eta^\dagger \psi_\eta = 0, \quad \int_{\mathcal{E}} \psi_\eta = 1,$$

where, we recall, $\mathcal{L}_\eta^\dagger$ denotes the adjoint of $\mathcal{L}_\eta$ on the flat space $L^2(\mathcal{E})$. The smoothness of ψ_η is a consequence of the hypoellipticiy of $\mathcal{L}_\eta^\dagger$ (see Section 2.3). For notational consistency, we set

$$\psi_0(q,p) = Z^{-1}\, \mathrm{e}^{-\beta H(q,p)}.$$

Note also that, as a corollary of Propositions 2.1 and 5.1, it is possible to define $\mathcal{L}_\eta^{-1}$ as a bounded operator on the subspaces

$$L^\infty_{\mathcal{K}_n,\eta}(\mathcal{E}) = \left\{ \varphi \in L^\infty_{\mathcal{K}_n}(\mathcal{E}) \,\middle|\, \int_{\mathcal{E}} \varphi\, \psi_\eta = 0 \right\}.$$

5.2.2. Linear response

It is expected, from a physical viewpoint, that the application of a non-zero constant force in a given direction induces a non-zero velocity in this direction. At the macroscopic level, the mobility is the proportionality constant between the observed average velocity and the force F. To rigorously define the mobility for the microscopic dynamics (5.1) under consideration, we consider the observable

$$R(q,p) = F^T M^{-1} p. \tag{5.9}$$

The response of interest is the steady-state average $\mathbb{E}_\eta(R)$, where $\mathbb{E}_\eta$ is the expectation with respect to the invariant measure of the non-equilibrium dynamics (5.1). Note that $\mathbb{E}_0(R) = 0$. This allows us to define the mobility in the direction F as the ratio of the average projected velocity R divided by η, in the limit of small forcings (provided this limit exists), that is,

$$\rho_F = \lim_{\eta \to 0} \frac{\mathbb{E}_\eta(R) - \mathbb{E}_0(R)}{\eta} = \lim_{\eta \to 0} \frac{\mathbb{E}_\eta(R)}{\eta}. \tag{5.10}$$

This is the definition of the mobility based on the linear response of non-equilibrium dynamics.

In order to ensure that the limit (5.10) is well defined, and eventually to rewrite ρ_F as some integrated correlation function, we need to characterize, to first order in η, the modification of the density $\psi_\eta(q,p)$ of the invariant measure of the dynamics (5.1) with respect to the reference canonical measure. To this end, it is convenient to work in the Hilbert space $L^2(\psi_0)$. We also introduce the projection operator

$$\Pi_0 f = f - \int_{\mathcal{E}} f \, \psi_0,$$

and the Hilbert space $L^2_0(\psi_0) = \Pi_0 L^2(\psi_0)$. Recall also the notation $\mathcal{A}^*$ for the adjoint of a given operator $\mathcal{A}$ on $L^2(\psi_0)$ (see (2.7)).

Theorem 5.2 (power expansion of the invariant measure). Consider the dynamics (5.1) with generator $\mathcal{L}_\eta = \mathcal{L}_0 + \eta \widetilde{\mathcal{L}}$, where $\mathcal{L}_0$ is given in (2.34) and $\widetilde{\mathcal{L}} = F \cdot \nabla_p$. Let r be the spectral radius of the bounded operator $(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^* \in \mathcal{B}(L^2_0(\psi_0))$:

$$r = \lim_{n \to +\infty} \| [(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*]^n \|^{1/n}.$$

Then, for $|\eta| < r^{-1}$, the unique invariant measure can be written as $\psi_\eta = f_\eta \psi_0$, where $f_\eta \in L^2(\psi_0)$ admits the following expansion in powers of η:

$$f_\eta = \left(1 + \eta (\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*\right)^{-1} \mathbf{1} = \left(1 + \sum_{n=1}^{+\infty} (-\eta)^n [(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*]^n\right) \mathbf{1}. \tag{5.11}$$

The linear term in η in the expression of f_η is denoted by

$$\mathfrak{f}_1 = -(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^* \mathbf{1}. \tag{5.12}$$

As will become clear in the proof of Theorem 5.2, this result can actually be proved for other dynamics than (5.1). This will be made precise in Section 5.2.4. As a corollary of Theorem 5.2, we immediately get a formula for the transport coefficient ρ_F defined by (5.10), using the fact that $\int_{\mathcal{E}} R\psi_0 = 0$:

$$\rho_F = \lim_{\eta \to 0} \frac{\int_{\mathcal{E}} R f_\eta \psi_0}{\eta} = \int_{\mathcal{E}} R \mathfrak{f}_1 \psi_0. \tag{5.13}$$

Note that the measure $f_\eta \psi_0$ is a probability measure. In particular, the normalization constant for ψ_η does not depend on η. This is due to the fact that $f_\eta - \mathbf{1} \in L^2_0(\psi_0)$, so that

$$\int_{\mathcal{E}} \psi_\eta = \int_{\mathcal{E}} \psi_0 = 1.$$

It can also be shown by a direct computation that $f_\eta \geqslant 0$: see Remark 5.3 below.

Proof. Let us first show that $(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*$ is a bounded operator on $L^2_0(\psi_0)$. To this end, we show that its adjoint $\widetilde{\mathcal{L}}\mathcal{L}_0^{-1}$ is bounded from $L^2_0(\psi_0)$ to $L^2(\psi_0)$. For C^∞ and compactly supported functions φ,

$$\frac{\gamma}{\beta} \|\nabla_p \varphi\|^2_{L^2(\psi_0)} = -\langle \mathcal{L}_0 \varphi, \varphi \rangle_{L^2(\psi_0)} \leqslant \|\mathcal{L}_0 \varphi\|_{L^2(\psi_0)} \|\varphi\|_{L^2(\psi_0)},$$

so that

$$\|\widetilde{\mathcal{L}}\varphi\|^2_{L^2(\psi_0)} \leqslant \|\nabla_p \varphi\|^2_{L^2(\psi_0)} \leqslant \frac{\beta}{\gamma} \|\mathcal{L}_0 \varphi\|_{L^2(\psi_0)} \|\varphi\|_{L^2(\psi_0)}.$$

Since $\mathcal{L}_0$ is invertible on $L^2_0(\psi_0)$ by Corollary 2.21, it follows that, for any $\varphi \in L^2_0(\psi_0)$,

$$\|\widetilde{\mathcal{L}}\mathcal{L}_0^{-1}\varphi\|^2_{L^2(\psi_0)} \leqslant \frac{\beta}{\gamma} \|\varphi\|_{L^2(\psi_0)} \|\mathcal{L}_0^{-1}\varphi\|_{L^2(\psi_0)}.$$

The operator $\widetilde{\mathcal{L}}\mathcal{L}_0^{-1}$ is therefore bounded from $L^2_0(\psi_0)$ to $L^2(\psi_0)$, with operator norm bounded from above by $\beta \|\mathcal{L}_0^{-1}\|_{\mathcal{B}(L^2(\psi_0))}/\gamma$. Denoting by Π_0 the projection operator

$$\Pi_0 \varphi = \varphi - \int_{\mathcal{E}} \varphi \, \psi_0,$$

it follows that the operator $\Pi_0 \widetilde{\mathcal{L}}\mathcal{L}_0^{-1}$ is bounded on $L^2_0(\psi_0)$, so that its adjoint $(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^* \Pi_0$ is also bounded on $L^2_0(\psi_0)$. In fact, $(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^* \Pi_0 = (\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*$ on $L^2_0(\psi_0)$, which proves that $(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*$ is bounded on $L^2_0(\psi_0)$ with operator

norm bounded from above by $\beta\|\mathcal{L}_0^{-1}\|_{\mathcal{B}(L^2(\psi_0))}/\gamma$. As a consequence the spectral radius r satisfies $r \leqslant \beta\|\mathcal{L}_0^{-1}\|_{\mathcal{B}(L^2(\psi_0))}/\gamma$.

It is then easy to check that (5.11) is a convergent series in $L^2(\psi_0)$ when $|\eta|r < 1$, since the series

$$\sum_{n=1}^{+\infty}(-\eta)^n[(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*]^n$$

converges in $\mathcal{B}(L_0^2(\psi_0))$. Therefore, the function f_η defined by (5.11) is well defined in $L^2(\psi_0)$. Henceforth we denote $\widetilde{\psi}_\eta = f_\eta\psi_0$. Note that $\int_{\mathcal{E}}\widetilde{\psi}_\eta = 1$ since $f_\eta - \mathbf{1} \in L_0^2(\psi_0)$.

Our aim is to prove that $\psi_\eta = \widetilde{\psi}_\eta$. To this end, we consider the following characterization of the invariant probability measure: for any C^∞ and compactly supported function φ,

$$\int_{\mathcal{E}}(\mathrm{e}^{t\mathcal{L}_\eta}\varphi)\psi_\eta = \int_{\mathcal{E}}\varphi\,\psi_\eta, \quad \text{for all } t \geqslant 0. \tag{5.14}$$

The same equality holds with ψ_η replaced by $\widetilde{\psi}_\eta$. Indeed, a straightforward computation shows that, by definition of f_η,

$$\mathcal{L}_\eta^* f_\eta = \mathcal{L}_0^*\big(1 + \eta(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*\big)f_\eta = \mathcal{L}_0^*\mathbf{1} = 0.$$

Therefore, $\mathrm{e}^{t\mathcal{L}_\eta^*}f_\eta = f_\eta$ for all $t \geqslant 0$. The measure $\widetilde{\psi}_\eta = f_\eta\psi_0$ is therefore such that, for all $\varphi \in L^2(\psi_0)$ and all $t \geqslant 0$,

$$\int_{\mathcal{E}}(\mathrm{e}^{t\mathcal{L}_\eta}\varphi)\widetilde{\psi}_\eta = \int_{\mathcal{E}}\varphi\,(\mathrm{e}^{t\mathcal{L}_\eta^*}f_\eta)\psi_0 = \int_{\mathcal{E}}\varphi\,f_\eta\psi_0 = \int_{\mathcal{E}}\varphi\,\widetilde{\psi}_\eta.$$

This equality is true in particular for bounded, measurable functions.

However, it is not possible to conclude at this stage that $\psi_\eta = \widetilde{\psi}_\eta$, because (5.14) characterizes invariant probability measures, whereas it is not clear that $\widetilde{\psi}_\eta$ is a probability measure: the non-negativity of the density is not guaranteed (but see Remark 5.3 below, which provides an alternative way to conclude the proof presented here). To prove non-negativity, we rely on the ergodicity of the continuous dynamics: for any bounded measurable function φ and any initial condition $(q_0, p_0) \in \mathcal{E}$, ergodicity of the Langevin dynamics[16] implies

$$\mathbb{E}^{(q_0,p_0)}\left(\frac{1}{t}\int_0^t\varphi(q_s,p_s)\right) = \frac{1}{t}\int_0^t(\mathrm{e}^{s\mathcal{L}_\eta}\varphi)(q_0,p_0)\,\mathrm{d}s \xrightarrow[t\to+\infty]{} \int_{\mathcal{E}}\varphi\,\psi_\eta,$$

[16] The proof of the almost-sure convergence of averages over a trajectory, for all initial conditions, is obtained as in Section 3.1, via the results by Kliemann (1987). The first step is to prove that ψ_η is positive, for which we refer to the discussion after (2.76).

so that, using Fubini's theorem to justify the first equality and the dominated convergence theorem for the limit,

$$\frac{1}{t}\int_0^t \left(\int_{\mathcal{E}} (\mathrm{e}^{s\mathcal{L}_\eta}\varphi)\, \widetilde{\psi}_\eta \right) \mathrm{d}s = \int_{\mathcal{E}} \left(\frac{1}{t}\int_0^t \mathrm{e}^{s\mathcal{L}_\eta}\varphi \, \mathrm{d}s \right) \widetilde{\psi}_\eta \xrightarrow[t\to+\infty]{} \int_{\mathcal{E}} \varphi\, \psi_\eta.$$

Since, for any $t > 0$,

$$\frac{1}{t}\int_0^t \left(\int_{\mathcal{E}} (\mathrm{e}^{s\mathcal{L}_\eta}\varphi)\, \widetilde{\psi}_\eta \right) \mathrm{d}s = \int_{\mathcal{E}} \varphi\, \widetilde{\psi}_\eta,$$

we conclude that

$$\int_{\mathcal{E}} \varphi\, \widetilde{\psi}_\eta = \int_{\mathcal{E}} \varphi\, \psi_\eta.$$

This shows that $\widetilde{\psi}_\eta = \psi_\eta$. In particular, $\widetilde{\psi}_\eta \geqslant 0$. □

Remark 5.3 (positivity of the invariant measure). It is possible to prove directly the positivity of the invariant density $f_\eta\psi_0$ by introducing the following Poisson equation:

$$\mathcal{L}_\eta \Phi_\eta = \mathbf{1}_{\{f_\eta \leqslant 0\}} - \int_{\mathcal{E}} \mathbf{1}_{\{f_\eta \leqslant 0\}}\, \psi_\eta, \quad \int_{\mathcal{E}} \Phi_\eta\, \psi_\eta = 0, \tag{5.15}$$

where f_η is defined by (5.11). This equation is well posed since $\mathcal{L}_\eta$ is invertible on weighted L^∞-spaces of functions with average 0 with respect to ψ_η (see the discussion after Proposition 5.1). Now, by the definition (5.11) of f_η,

$$\int_{\mathcal{E}} (\mathcal{L}_\eta \Phi_\eta) f_\eta \psi_0 = 0. \tag{5.16}$$

On the other hand, by definition of Φ_η,

$$\begin{aligned} \int_{\mathcal{E}} (\mathcal{L}_\eta \Phi_\eta) f_\eta \psi_0 &= \int_{\mathcal{E}} \mathbf{1}_{\{f_\eta \leqslant 0\}} f_\eta\, \psi_0 - \left(\int_{\mathcal{E}} \mathbf{1}_{\{f_\eta \leqslant 0\}}\, \psi_\eta \right) \int_{\mathcal{E}} f_\eta \psi_0 \\ &= \int_{\mathcal{E}} \mathbf{1}_{\{f_\eta \leqslant 0\}} f_\eta\, \psi_0 - \int_{\mathcal{E}} \mathbf{1}_{\{f_\eta \leqslant 0\}}\, \psi_\eta. \end{aligned}$$

The first integral on the right-hand side of the previous equality is non-positive, while the second one is non-negative; hence the quantity on the right-hand side is non-positive. Since the right-hand side is equal to 0 by (5.16), each term must be 0. This allows us to conclude that $f_\eta \geqslant 0$ almost everywhere (using the first integral, since ψ_0 has a positive density with respect to the Lebesgue measure).

5.2.3. Reformulating the linear response as an integrated correlation

A very useful corollary of (5.12)–(5.13) is the following reformulation of the linear response definition (5.10) of the transport coefficient through the

celebrated Green–Kubo formula. To state it, we introduce the conjugate response function, formally defined as $S = \widetilde{\mathcal{L}}^* \mathbf{1}$. Its expression is found in practice by integration by parts, as follows: for any C^∞ compactly supported function φ,

$$\int_{\mathcal{E}} \widetilde{\mathcal{L}} \varphi \, \psi_0 = \int_{\mathcal{E}} \varphi \, S \, \psi_0. \tag{5.17}$$

Note that the expression of S is determined by the applied perturbation $\widetilde{\mathcal{L}}$, and not by the response function R chosen in (5.10). For the non-equilibrium Langevin dynamics (5.1), a simple computation shows that

$$S(q,p) = \beta R(q,p) = \beta F^T M^{-1} p.$$

Note that $S \in L^2_0(\psi_0)$ in this case.

Proposition 5.4 (Green–Kubo formula). Consider the non-equilibrium Langevin dynamics (5.1) and the definition (5.17) of the conjugate function S. Then, for any $R \in L^2_0(\psi_0)$,

$$\lim_{\eta \to 0} \frac{\mathbb{E}_\eta(R)}{\eta} = \int_0^{+\infty} \mathbb{E}_0(R(q_t,p_t) S(q_0,p_0)) \, \mathrm{d}t, \tag{5.18}$$

where $\mathbb{E}_\eta$ is an average with respect to the invariant measure $\psi_\eta(q,p) \, \mathrm{d}q \, \mathrm{d}p$ of the non-equilibrium dynamics (5.1), while the expectation $\mathbb{E}_0$ on the right-hand side is taken over all initial conditions distributed according to the canonical measure $\psi_0(q,p) \, \mathrm{d}q \, \mathrm{d}p$, and over all realizations of the reference equilibrium dynamics with generator $\mathcal{L}_0$.

The Green–Kubo formula thus shows that a non-equilibrium property (namely the transport coefficient $\rho_F = \lim_{\eta \to 0} \mathbb{E}_\eta(R)/\eta$ of the left-hand side of (5.18)) can be obtained using simulations at equilibrium, namely for $\eta = 0$ (see the right-hand side of (5.18)). This result can easily be generalized to other dynamics when the perturbation $\widetilde{\mathcal{L}}$ is such that $S \in L^2(\psi_0)$ and the linear response result (5.12) holds (see Section 5.2.4 for possible assumptions on $\widetilde{\mathcal{L}}$).

Proof. In view of (5.10), (5.12) and (5.13), as well as the equality

$$-\mathcal{L}_0^{-1} = \int_0^{+\infty} \mathrm{e}^{t\mathcal{L}_0} \, \mathrm{d}t$$

as operators on $L^2_0(\psi_0)$ (as given by Corollary 2.21), we can write, for $R \in L^2_0(\psi_0)$,

$$\begin{aligned} \lim_{\eta \to 0} \frac{\mathbb{E}_\eta(R)}{\eta} &= \int_{\mathcal{E}} R \, \mathfrak{f}_1 \, \psi_0 = - \int_{\mathcal{E}} [\mathcal{L}_0^{-1} R] [\widetilde{\mathcal{L}}^* \mathbf{1}] \, \psi_0 \\ &= \int_0^{+\infty} \mathbb{E}(R(q_t,p_t) S(q_0,p_0)) \, \mathrm{d}t, \end{aligned}$$

which gives the claimed result. □

Let us rewrite equation (5.18) more precisely in the context of (5.1). The mobility ρ_F in the direction F, defined in (5.10), is equal to β times the integrated velocity autocorrelation:

$$\rho_F = \lim_{\eta \to 0} \frac{\mathbb{E}_\eta(F \cdot M^{-1} p)}{\eta} = \beta \int_0^{+\infty} \mathbb{E}_0((F \cdot M^{-1} p_t)(F \cdot M^{-1} p_0)) \, \mathrm{d}t. \quad (5.19)$$

5.2.4. Generalization to other dynamics

It is of course possible to generalize linear response results and Green–Kubo formulae to dynamics other than (5.1), either by generalizing the assumptions ensuring that Theorem 5.2 and Proposition 5.4 hold (as we do below), or by working with a different set of assumptions as in Assaraf, Jourdain, Lelièvre and Roux (2015).

An inspection of the proof of Theorem 5.2 shows that the result can be generalized to other perturbations, and in fact to other reference equilibrium dynamics $(x_t^\eta)_{t \geqslant 0}$ with generator $\mathcal{L}_\eta = \mathcal{L}_0 + \eta \widetilde{\mathcal{L}}$, under the following conditions.

(a) *Existence of a unique invariant measure.* For any $\eta \in \mathbb{R}$, the perturbed dynamics admits a unique invariant measure with C^∞ density $\psi_\eta(x)$ with respect to the Lebesgue measure $\mathrm{d}x$ on $\mathcal{X}$. Such results can be proved with the techniques from Section 2.4.

(b) *Ergodicity of the perturbed dynamics.* The perturbed dynamics is ergodic in the following sense: for any bounded measurable function φ and almost all initial condition x_0,

$$\frac{1}{t} \int_0^t \varphi(x_s) \, \mathrm{d}s \xrightarrow[t \to +\infty]{} \int_{\mathcal{X}} \varphi \, \psi_\eta \quad \text{almost surely.}$$

See Section 3.1 for techniques to obtain such convergence results.

Alternatively, as discussed in Remark 5.3, the ergodicity condition can be replaced with solvability conditions for the Poisson equation (5.15).

(c) *Properties of the equilibrium dynamics.* $\mathrm{Ker}(\mathcal{L}_0^*) = \mathbf{1}$ and $\mathcal{L}_0^*$ is invertible on $L_0^2(\psi_0)$. See Corollary 2.4 for overdamped Langevin dynamics and Corollary 2.21 for Langevin dynamics.

(d) *Properties of the perturbation.* $\mathrm{Ran}(\widetilde{\mathcal{L}}^*) \subset L_0^2(\psi_0)$ and $(\widetilde{\mathcal{L}} \mathcal{L}_0^{-1})^*$ is bounded on $L_0^2(\psi_0)$. Note that, if $\widetilde{\mathcal{L}} \mathbf{1} = 0$, then the function $\widetilde{\mathcal{L}}^* \varphi$ has average 0 with respect to ψ_0:

$$\int_{\mathcal{X}} \widetilde{\mathcal{L}}^* \varphi \, \psi_0 = \int_{\mathcal{X}} \varphi \, \widetilde{\mathcal{L}} \mathbf{1} \, \psi_0 = 0.$$

The last condition, namely that $(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*$ is bounded on $L^2_0(\psi_0)$, expresses the fact that the perturbation $\widetilde{\mathcal{L}}$ is sufficiently small. A typical way of proving that $(\widetilde{\mathcal{L}}\mathcal{L}_0^{-1})^*$ is bounded on $L^2_0(\psi_0)$ is, as at the beginning of the proof of Theorem 5.2, to show that $\widetilde{\mathcal{L}}$ is $\mathcal{L}_0$-bounded, namely that there exist $a, b > 0$ such that, for all C^∞ and compactly supported functions φ,

$$\|\widetilde{\mathcal{L}}\varphi\|_{L^2(\psi_0)} \leqslant a\|\mathcal{L}_0\varphi\|_{L^2(\psi_0)} + b\|\varphi\|_{L^2(\psi_0)}. \tag{5.20}$$

Then, for all $\varphi \in L^2_0(\psi_0)$,

$$\|\widetilde{\mathcal{L}}\mathcal{L}_0^{-1}\varphi\|_{L^2(\psi_0)} \leqslant a\|\varphi\|_{L^2(\psi_0)} + b\|\mathcal{L}_0^{-1}\varphi\|_{L^2(\psi_0)},$$

so that $\widetilde{\mathcal{L}}\mathcal{L}_0^{-1}$, and its adjoint, are bounded operators on $L^2_0(\psi_0)$ with operator norm bounded by $a + b\|\mathcal{L}_0^{-1}\|_{\mathcal{B}(L^2_0(\psi_0))}$.

Remark 5.5 (stronger perturbations). There are situations for which the condition (5.20) is not satisfied. This is the case for generalizations of the dynamics (5.2), for which the perturbation operator $\widetilde{\mathcal{L}}$ involves second derivatives of the momenta p, whereas $\mathcal{L}_0$ may control only first-order derivatives. In this case, it is not possible to write an expansion of the invariant measure as a power series in η. Nonetheless, linear response results on average properties can still be stated by following the same strategy as in the proof of Theorem 3.3. To this end, we assume that Assumption 3.2 is satisfied for $\mathcal{A}_1 = \mathcal{L}_0$ and $\mathcal{A}_1 = \mathcal{L}_0^*$. First, introduce $\mathfrak{f}_1$ such that

$$\int_{\mathcal{X}} [(\mathcal{L}_0 + \eta\widetilde{\mathcal{L}})\varphi](1 + \eta\mathfrak{f}_1)\psi_0 = \mathrm{O}(\eta^2).$$

A simple computation shows that $\mathfrak{f}_1$ is formally given by (5.12). In fact, $\mathfrak{f}_1 = -(\mathcal{L}_0^*)^{-1}S$, where S is defined in (5.17). We next replace φ by $Q_\eta\varphi$, where Q_η is the approximate inverse

$$Q_\eta = \Pi_0\mathcal{L}_0^{-1}\Pi_0 - \eta\Pi_0\mathcal{L}_0^{-1}\Pi_0\widetilde{\mathcal{L}}\Pi_0\mathcal{L}_0^{-1}\Pi_0.$$

The operator Q_η is a well-defined operator acting on C^∞ functions with zero average with respect to ψ_0. It can therefore be shown that, under moment conditions on the invariant measure (that is, ψ_η integrates all the scale functions appearing in Definition 3.1),

$$\int_{\mathcal{X}} \varphi\,\psi_\eta = \int_{\mathcal{X}} \varphi\,\psi_0 + \eta\int_{\mathcal{X}} \varphi\,\mathfrak{f}_1\,\psi_0 + \eta^2 r_{\varphi,\eta},$$

where $|r_{\varphi,\eta}| \leqslant K$ for $|\eta|$ sufficiently small.

5.3. Error estimates on the computation of transport coefficients

The results of the previous section show that there are two ways to compute transport coefficients, either by discretizing the integrated correlation

function based on the Green–Kubo formula (5.18), or by approximating the derivative of equilibrium averages with respect to the magnitude of the external forcing as in (5.10). In this section we provide, for both approaches, error estimates on the bias generated by the use of finite time steps to integrate the dynamics. The results quoted here were first presented in Leimkuhler *et al.* (2015).

5.3.1. Green–Kubo formulae

We first provide error estimates on linear responses computed using Green–Kubo formulae such as (5.18). We state the result on a general space $\mathcal{X}$ for a general dynamics (3.1) with generator $\mathcal{L}$ (instead of the notation $\mathcal{L}_0$ used in Section 5.2). The precise result is formulated for smooth functions in the sense of Definition 3.1. We let π denote the reference invariant measure, and still suppose that Assumption 3.2 holds with $\mathcal{A}_1 = \mathcal{L}$ (*i.e.*, the space $\mathcal{S}$ introduced in Definition 3.1 is dense in $L^2(\pi)$ and the operator $\mathcal{L}^{-1} : \mathcal{S}_0 \to \mathcal{S}_0$ is well defined, where $\mathcal{S}_0$ is defined in (3.16)).

As in Section 3.2, we consider a discretization of the continuous dynamics under consideration, with transition operator $P_{\Delta t}$, and let $\pi_{\Delta t}$ denote an invariant measure of the numerical scheme (see Section 3.2.2 for a discussion of possible conditions ensuring the existence of an invariant measure). We also introduce the projection operator

$$\Pi_{\Delta t}\varphi = \varphi - \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi_{\Delta t},$$

as well as

$$L^\infty_{\mathcal{K}_s,\Delta t}(\mathcal{X}) = \Pi_{\Delta t} L^\infty_{\mathcal{K}_s}(\mathcal{X}) = \left\{ \varphi \in L^\infty_{\mathcal{K}_s}(\mathcal{X}) \;\middle|\; \int_{\mathcal{X}} \varphi \, \mathrm{d}\pi_{\Delta t} = 0 \right\}.$$

The range of $\Pi_{\Delta t}$ is contained in the set of functions with average zero with respect to the invariant measure $\pi_{\Delta t}$ of the numerical scheme.

Theorem 5.6 (error estimates for Green–Kubo). Consider a numerical method with an invariant measure $\pi_{\Delta t}$ which integrates any scale function $\mathcal{K}_n$ introduced in Definition 3.1. Assume that there exists an integer $\alpha \geqslant 1$, such that, for any observable $\phi \in \mathcal{S}$, there is $K, \Delta t^* > 0$ for which the following conditions hold.

(i) *Error on the invariant measure*:

$$\int_{\mathcal{X}} \phi \, \mathrm{d}\pi_{\Delta t} = \int_{\mathcal{X}} \phi \, \mathrm{d}\pi + \Delta t^\alpha r_{\phi,\Delta t}, \tag{5.21}$$

with $|r_{\phi,\Delta t}| \leqslant K$ for $0 < \Delta t \leqslant \Delta t^*$.

(ii) *Expansion of* $P_{\Delta t}$:

$$-\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\phi = \mathcal{L}\phi + \Delta t S_1 \phi + \cdots + \Delta t^{\alpha-1} S_{\alpha-1}\phi + \Delta t^\alpha \widetilde{R}_{\alpha,\Delta t}\phi, \tag{5.22}$$

where the operators $S_1, \dots, S_{\alpha-1}, \widetilde{R}_{\alpha,\Delta t}$ (which are defined independently of ϕ) are well defined on $\mathcal{S}$ with values in $\mathcal{S}$; and there exists $s_0 \in \mathbb{N}$ (depending on α and ϕ) such that $\|\widetilde{R}_{\alpha,\Delta t}\phi\|_{L^\infty_{\mathcal{K}_{s_0}}} \leqslant K$ for $0 < \Delta t \leqslant \Delta t^*$.

Moreover, we assume the following.

(iii) *Uniform-in-Δt exponential convergence of $P_{\Delta t}$.* For any $s \geqslant 0$, there exist $C_s, \lambda_s > 0$ such that, for all $0 < \Delta t \leqslant \Delta t^*$,

$$\|P^n_{\Delta t}\|_{\mathcal{B}(L^\infty_{\mathcal{K}_s,\Delta t})} \leqslant C_s\, \mathrm{e}^{-\lambda_s n \Delta t}, \quad \text{for all } n \in \mathbb{N}. \tag{5.23}$$

Then, the integrated correlation of two observables $\psi, \varphi \in \mathcal{S}_0$ can be approximated by a Riemann sum up to an error of order Δt^α: there exists $R > 0$ such that

$$\int_0^{+\infty} \mathbb{E}_\pi(\psi(x_t)\varphi(x_0))\,\mathrm{d}t = \Delta t \sum_{n=0}^{+\infty} \mathbb{E}_{\pi_{\Delta t}}(\widetilde{\psi}_{\Delta t,\alpha}(x^n)\varphi(x^0)) + \Delta t^\alpha r^{\psi,\varphi}_{\Delta t}, \tag{5.24}$$

with $|r^{\psi,\varphi}_{\Delta t}| \leqslant R$ for $0 < \Delta t \leqslant \Delta t^*$. In this expression, the expectation $\mathbb{E}_\pi$ is over all initial conditions $(q_0, p_0) \sim \pi$ and over all realizations of the continuous dynamics with generator $\mathcal{L}$, while the expectation $\mathbb{E}_{\pi_{\Delta t}}$ is over all initial conditions $(q_0, p_0) \sim \pi_{\Delta t}$ and over all realizations of the Markov chain induced by $P_{\Delta t}$. Moreover, the modified observable $\widetilde{\psi}_{\Delta t,\alpha} \in \mathcal{S}$ is defined as $\widetilde{\psi}_{\Delta t,\alpha} = \Pi_{\Delta t}\psi_{\Delta t,\alpha}$ with

$$\psi_{\Delta t,\alpha} = (\mathrm{Id} + \Delta t\, S_1\mathcal{L}^{-1} + \cdots + \Delta t^{\alpha-1} S_{\alpha-1}\mathcal{L}^{-1})\psi. \tag{5.25}$$

This result deserves several comments, both on the three main assumptions (5.21), (5.22) and (5.23), as well as on the error estimate (5.24) itself. The proof is presented after this discussion.

Let us start with the assumptions of Theorem 5.6. Equation (5.21) can be proved by following the general strategy presented in Theorem 3.3. Note that the remainder term in (5.21) can vanish, for example when a Metropolis procedure is superimposed on the numerical scheme. In this case, α is determined by (5.22). Condition (5.22) has already been encountered when proving error estimates on the invariant measure (see (3.15)). Finally, the uniform-in-Δt convergence (5.23) is given, for the Euler–Maruyama discretization of overdamped Langevin dynamics, by (3.29), which is a consequence of the uniform minorization condition of Lemma 3.4; similar results exist for certain discretization of Langevin dynamics (see Leimkuhler *et al.* 2015).

Let us now comment on the error estimate (5.24). First, the result shows that the error is of order Δt^α, upon modifying the observable ψ as $\widetilde{\psi}_{\Delta t,\alpha}$. Therefore, a first limitation to the reduction of the error in Green–Kubo formulae arises from the error on the invariant measure itself. However, in practice, especially when α is large, the error is actually determined by

the approximation of the corrected observable $\widetilde{\psi}_{\Delta t,\alpha}$. When the operators S_k are powers of $\mathcal{L}$, the correction terms $S_k\mathcal{L}^{-1}\psi$ in (5.25) can be easily computed. Let k_0 denote the last index for which $S_k\mathcal{L}^{-1}\psi$ can be evaluated. For example (see the discussion after (3.15)), if the discretization method is of weak order k_0, then $S_k = \mathcal{L}^k/k!$ for all $k = 1, \dots, k_0$, so that $S_k\mathcal{L}^{-1}\psi$ is easy to evaluate up to $k = k_0$. If $k_0 < \alpha$, the error estimate which can be used in practice is

$$\int_0^{+\infty} \mathbb{E}_\pi(\psi(x_t)\varphi(x_0))\,\mathrm{d}t = \Delta t \sum_{n=0}^{+\infty} \mathbb{E}_{\pi_{\Delta t}}(\widetilde{\psi}_{\Delta t,k_0}(x^n)\varphi(x^0)) + \Delta t^{k_0+1} r_{\Delta t}^{\psi,\varphi}.$$

However, even if $\psi_{\Delta t,k_0}$ can be evaluated, it still remains to approximate $\widetilde{\psi}_{\Delta t,k_0} = \Pi_{\Delta t}\psi_{\Delta t,k_0}$, typically by approximating the average of $\psi_{\Delta t,k_0}$ with respect to $\pi_{\Delta t}$ by trajectory averages.

Remark 5.7 (error estimates for Metropolis–Hastings dynamics). For discretizations of the continuous dynamics stabilized by a Metropolis–Hastings procedure, such as MALA for overdamped Langevin, the invariant measure of the numerical scheme is exact by construction. However, since the quantity $S_1\mathcal{L}^{-1}\psi$ cannot be evaluated in general, the resulting approximation of the integrated correlation is based on $\widetilde{\psi}_{\Delta t,0} = \Pi_{\Delta t}\psi$, which leads to an approximation of order Δt of the Green–Kubo integral. An error estimate of order $\Delta t^{3/2}$ can nevertheless be obtained by modifying the Metropolis–Hastings proposal. It is even possible to obtain errors of order Δt^2 by modifying the Metropolis acceptance rule. See Fathi *et al.* (2014) and Fathi and Stoltz (2015) for further details.

Let us now present the proof of Theorem 5.6.

Proof. Fix two observables $\varphi, \psi \in \mathcal{S}_0$. Note that

$$\int_0^{+\infty} \mathbb{E}_\pi(\psi(x_t)\varphi(x_0))\,\mathrm{d}t = \int_{\mathcal{X}} (-\mathcal{L}^{-1}\psi)\varphi\,\mathrm{d}\pi.$$

In order to introduce the correlation functions of the numerical scheme, we would like, in view of (5.22), to replace the measure π by $\pi_{\Delta t}$ and the operator $-\mathcal{L}^{-1}$ by

$$\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)^{-1} = \Delta t \sum_{n=0}^{+\infty} P_{\Delta t}^n,$$

However, this is not possible as such, for two reasons. First, as indicated in (5.23), the above sum is only convergent when the operators under consideration are restricted to subspaces of functions with average 0 with respect to $\pi_{\Delta t}$. We therefore need to introduce the projections operators $\Pi_{\Delta t}$ to restrict $\mathcal{L}^{-1}$ to the range of $\Pi_{\Delta t}$. Second, it is not possible to consider the inverse of the right-hand side of (5.22) directly, so we will introduce

$(\mathrm{Id} - P_{\Delta t})^{-1}(\mathrm{Id} - P_{\Delta t})$ instead, in order to retrieve some operator $\mathcal{L}$ to dominant order in Δt, so as to cancel the inverse operator $\mathcal{L}^{-1}$.

Let us first introduce the projection operators $\Pi_{\Delta t}$ and the invariant measure $\pi_{\Delta t}$ of the numerical scheme, using the fact that $-\mathcal{L}^{-1}\psi$ has zero average with respect to π:

$$\begin{aligned}\int_{\mathcal{X}} (-\mathcal{L}^{-1}\psi)\varphi \,\mathrm{d}\pi &= \int_{\mathcal{X}} (-\mathcal{L}^{-1}\psi)\Pi_{\Delta t}\varphi \,\mathrm{d}\pi \\ &= \int_{\mathcal{X}} (-\mathcal{L}^{-1}\psi)\Pi_{\Delta t}\varphi \,\mathrm{d}\pi_{\Delta t} + \Delta t^{\alpha}\widetilde{r}_{\Delta t}^{\mathcal{L}^{-1}\psi,\varphi}, \\ &= \int_{\mathcal{X}} \Pi_{\Delta t}(-\mathcal{L}^{-1}\psi)\Pi_{\Delta t}\varphi \,\mathrm{d}\pi_{\Delta t} + \Delta t^{\alpha}\widetilde{r}_{\Delta t}^{\mathcal{L}^{-1}\psi,\varphi}, \qquad (5.26)\end{aligned}$$

where $|\widetilde{r}_{\Delta t}^{\mathcal{L}^{-1}\psi,\varphi}| \leqslant K/2$ for $0 < \Delta t \leqslant \Delta t^*$ by (5.21) (possibly upon increasing the value of K and decreasing Δt^*).

The next step is to approximate $\Pi_{\Delta t}(-\mathcal{L}^{-1}\psi)$ in terms of powers of $P_{\Delta t}$. We use the fact that $\Pi_{\Delta t}P_{\Delta t} = P_{\Delta t}\Pi_{\Delta t}$, and that

$$\begin{aligned}\Pi_{\Delta t} &= \Pi_{\Delta t}\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)^{-1}\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right) \\ &= \Pi_{\Delta t}\left(\Delta t\sum_{n=0}^{+\infty} P_{\Delta t}^n\right)\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right).\end{aligned}$$

Note that the above sum is convergent in $\mathcal{B}(L^{\infty}_{\mathcal{K}_s,\Delta t})$ in view of (5.23). Relying on (5.22),

$$\begin{aligned}-\Pi_{\Delta t}\mathcal{L}^{-1}\psi &= -\Pi_{\Delta t}\left(\Delta t\sum_{n=0}^{+\infty} P_{\Delta t}^n\right)\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)\mathcal{L}^{-1}\psi \\ &= \Delta t\left(\sum_{n=0}^{+\infty} P_{\Delta t}^n\right)\Pi_{\Delta t}(\mathcal{L} + \cdots + \Delta t^{\alpha-1}S_{\alpha-1} + \Delta t^{\alpha}\widetilde{R}_{\alpha,\Delta t})\mathcal{L}^{-1}\psi, \\ &= \Delta t\sum_{n=0}^{+\infty} P_{\Delta t}^n\widetilde{\psi}_{\Delta t,\alpha} + \Delta t^{\alpha}\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)^{-1}\Pi_{\Delta t}\widetilde{R}_{\alpha,\Delta t}\mathcal{L}^{-1}\psi.\end{aligned}$$

Note that the sums on the right-hand side are well defined in view of the decay estimates (5.23). Substituting the above equality in (5.26) leads to

$$\begin{aligned}\int_{\mathcal{X}} (-\mathcal{L}^{-1}\psi)\varphi \,\mathrm{d}\pi &= \Delta t\sum_{n=0}^{+\infty}\int_{\mathcal{X}} P_{\Delta t}^n\widetilde{\psi}_{\Delta t,\alpha}(\Pi_{\Delta t}\varphi)\,\mathrm{d}\pi_{\Delta t} \\ &+ \Delta t^{\alpha}\int_{\mathcal{X}}\left(\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)^{-1}\Pi_{\Delta t}\widetilde{R}_{\alpha,\Delta t}\mathcal{L}^{-1}\psi\right)\Pi_{\Delta t}\varphi \,\mathrm{d}\pi_{\Delta t} + \Delta t^{\alpha}\widetilde{r}_{\Delta t}^{\mathcal{L}^{-1}\psi,\varphi}.\end{aligned}$$

To conclude the proof, we use the fact that, for a given smooth function f and upon increasing K and decreasing Δt^*, there exist an integer s and a constant $\widetilde{K} > 0$ (depending on f) such that $\|\widetilde{R}_{\alpha,\Delta t} f\|_{L^\infty_{\mathcal{K}_s}} \leqslant \widetilde{K}$ for any $\Delta t \in (0, \Delta t^*]$. In addition, the following resolvent bound is directly obtained from (5.23):

$$\left\| \left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t} \right)^{-1} \right\|_{\mathcal{B}(L^\infty_{\mathcal{K}_s,\Delta t})} \leqslant \frac{C_s}{\lambda_s}.$$

Finally, $\pi_{\Delta t}$ integrates all scale functions $\mathcal{K}_n$ by assumption. Therefore, upon increasing the value of K, the following inequality holds for any $\Delta t \in (0, \Delta t^*]$:

$$\left| \int_{\mathcal{X}} \left(\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t} \right)^{-1} \Pi_{\Delta t} \widetilde{R}_{\alpha,\Delta t} \mathcal{L}^{-1} \psi \right) \Pi_{\Delta t} \varphi \, \mathrm{d}\pi_{\Delta t} \right| \leqslant \frac{K}{2}.$$

Since

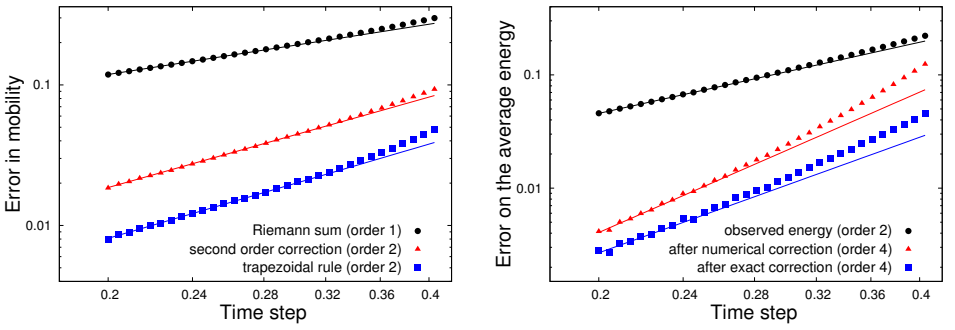
$$\begin{aligned} \sum_{n=0}^{+\infty} \int_{\mathcal{X}} P^n_{\Delta t} \widetilde{\psi}_{\Delta t,\alpha} (\Pi_{\Delta t} \varphi) \, \mathrm{d}\pi_{\Delta t} &= \sum_{n=0}^{+\infty} \int_{\mathcal{X}} P^n_{\Delta t} \widetilde{\psi}_{\Delta t,\alpha} \, \varphi \, \mathrm{d}\pi_{\Delta t} \\ &= \sum_{n=0}^{+\infty} \mathbb{E}_{\pi_{\Delta t}} (\widetilde{\psi}_{\Delta t,\alpha}(x^n) \varphi(x^0)), \end{aligned}$$

equation (5.24) finally follows. □

Remark 5.8 (Green–Kubo formulae for second-order schemes). In the particular case when $\alpha = 2$ in (5.21), which is very relevant in practice (*e.g.* for second-order splittings of Langevin dynamics, or the geometric Langevin algorithms discussed in Section 3.3), it is possible to avoid modifying the observable ψ when S_1 is proportional to $\mathcal{L}^2$, by appropriately changing the quadrature rule. For schemes of weak order 2, for which $S_1 = \mathcal{L}^2/2$, this amounts to discretizing the time integral with a trapezoidal rule instead of a Riemann sum. See Corollary 2.20 in Leimkuhler *et al.* (2015) and Theorem 6 in Fathi and Stoltz (2015) for further details.

To illustrate Theorem 5.6, we consider a simple two-dimensional system with position $q = (x, y) \in \mathcal{M} = (2\pi\mathbb{T})^2$, a potential energy function $V(q) = 2\cos(2x) + \cos(y)$, an inverse temperature $\beta = 1$, a friction $\gamma = 1$ and an identity mass matrix. The external force is set to $F = (1, 0)$. Figure 5.1 displays simulation results obtained for the scheme associated with $P^{\gamma C,B,A,B,\gamma C}_{\Delta t}$ (see the discussion after (3.31) and Leimkuhler *et al.* 2015) when approximating the mobility

$$\rho_F = \beta \int_0^{+\infty} \mathbb{E}_\mu (p_t \cdot p_0) \, \mathrm{d}t.$$



(a) (b)

Figure 5.1. (a) Error in the mobility as a function of the time step Δt when the integrated velocity autocorrelation function is computed using a Riemann sum or with the corrected formula (5.24) in the case $\alpha = 2$. The result from computing the integral using the trapezoidal rule is also shown. (b) Error in the computed average of total energy, with the correction term computed using the same step size by a discretization of (3.32) with the discrete Green–Kubo formula (5.24). Averages obtained with a reference correction computed more accurately at a smaller time step in a separate simulation are labelled as the 'exact correction'. In both cases, 'order α' means that the error scales as $C\Delta t^\alpha$.

For this second-order scheme, an approximation of the time integral based on a simple Riemann sum leads to errors of order Δt, while a second-order convergence is obtained upon modifying the observables according to (5.25) or using a trapezoidal rule. In addition, we show how to improve the computation of the average value of the Hamiltonian $H(q,p) = V(q) + |p|^2/2$ based on an estimation of the correction term (3.32) reformulated as some integrated correlation function. In the case of the total energy, the corresponding correction is proportional to the mobility ρ_F. As predicted by theoretical results similar to Theorem 3.3, errors in average properties can be reduced from Δt^2 to Δt^3 when the correction term is correctly estimated; in fact, the errors can be shown to be reduced to order Δt^4 (see Theorem 2.16 in Leimkuhler *et al.* 2015).

5.3.2. Linear response approaches

In this part, we discuss how to approximate a transport coefficient by estimating numerically the derivative of equilibrium averages with respect to the forcing parameter; see (5.10). This requires expansions of the invariant measures of numerical approximations of non-equilibrium dynamics with respect to two small parameters: the magnitude η of the forcing and the time step Δt. To be more concrete, we illustrate the approach on the computation of the mobility with the perturbed Langevin dynamics (5.1). We first present numerical schemes for this dynamics, before stating error estimates

on the computation of average properties and linear responses. The discussion here follows the presentation of Leimkuhler *et al.* (2015, Section 3), to which we refer for further details.

We still consider splitting schemes which reduce to the schemes presented in Section 3.3 when $\eta = 0$. Recall that $\mathcal{L}_0 = A + B + \gamma C$, where the elementary operators A, B, C are introduced in (3.30):

$$A = M^{-1}p \cdot \nabla_q, \quad B = -\nabla V(q) \cdot \nabla_p, \quad C = -M^{-1}p \cdot \nabla_p + \frac{1}{\beta}\Delta_p.$$

Since the aim is to decompose the evolution generated by $\mathcal{L}_\eta = \mathcal{L}_0 + \eta\widetilde{\mathcal{L}}$ into analytically integrable parts, there are two principal options: either replace B by

$$B_\eta = B + \eta\widetilde{\mathcal{L}}$$

or replace γC by $\gamma C + \eta\widetilde{\mathcal{L}}$. However, the schemes built on the latter option do not perform correctly in the overdamped limit, since their invariant measures are not consistent with the invariant measures of non-equilibrium overdamped Langevin dynamics, that is,

$$\mathrm{d}q_t = (-\nabla V(q_t) + \eta F)\,\mathrm{d}t + \sqrt{\frac{2}{\beta}}\,\mathrm{d}W_t. \tag{5.27}$$

The latter dynamics is obtained from (5.1) in the limit $\gamma \to +\infty$ upon rescaling the time as γt. To illustrate this point, consider for instance the first-order splitting scheme associated with the evolution operator

$$P_{\Delta t}^{A,B,\gamma C+\eta\widetilde{\mathcal{L}}} = \mathrm{e}^{\Delta t\, A}\, \mathrm{e}^{\Delta t\, B}\, \mathrm{e}^{\Delta t(\gamma C+\eta\widetilde{\mathcal{L}})},$$

that is,

$$\begin{aligned}
q^{n+1} &= q^n + \Delta t\, M^{-1}p^n,\\
\widetilde{p}^{n+1} &= p^n - \Delta t\, \nabla V(q^{n+1}),\\
p^{n+1} &= \alpha_{\Delta t}\widetilde{p}^{n+1} + \frac{1-\alpha_{\Delta t}}{\gamma}\,\eta F + \sqrt{\frac{1-\alpha_{\Delta t}^2}{\beta}M}\,G^n,
\end{aligned}$$

where $\alpha_{\Delta t} = \exp(-\gamma M^{-1}\Delta t)$ is defined after (3.31), and (G^n) is a sequence of i.i.d. Gaussian random vectors with identity covariance. When $M = \mathrm{Id}$ and $\gamma \to +\infty$, a standard Euler–Maruyama discretization of the *equilibrium* overdamped Langevin dynamics (*i.e.* $\eta = 0$) is obtained; whereas we would like to obtain a consistent discretization of non-equilibrium overdamped Langevin dynamics (5.27). This suggests that numerical methods based on the integration of $\gamma C + \eta\widetilde{\mathcal{L}}$ will not estimate the mobility for large γ correctly.

We therefore prefer to consider schemes obtained by replacing B with $B + \eta\widetilde{\mathcal{L}}$, such as the first-order splitting

$$P_{\Delta t}^{A,B+\eta\widetilde{\mathcal{L}},\gamma C} = \mathrm{e}^{\Delta t\, A}\, \mathrm{e}^{\Delta t (B+\eta\widetilde{\mathcal{L}})}\, \mathrm{e}^{\gamma \Delta t\, C}, \tag{5.28}$$

or the second-order splitting

$$P_{\Delta t}^{\gamma C,B+\eta\widetilde{\mathcal{L}},A,B+\eta\widetilde{\mathcal{L}},\gamma C} = \mathrm{e}^{\gamma \Delta t\, C/2}\, \mathrm{e}^{\Delta t (B+\eta\widetilde{\mathcal{L}})/2}\, \mathrm{e}^{\Delta t\, A}\, \mathrm{e}^{\Delta t (B+\eta\widetilde{\mathcal{L}})/2}\, \mathrm{e}^{\gamma \Delta t\, C/2}. \tag{5.29}$$

For example, the numerical scheme associated with $P_{\Delta t}^{A,B+\eta\widetilde{\mathcal{L}},\gamma C}$,

$$\begin{aligned} q^{n+1} &= q^n + \Delta t\, M^{-1} p^n, \\ \widetilde{p}^{n+1} &= p^n + \Delta t(-\nabla V(q^{n+1}) + \eta F), \\ p^{n+1} &= \alpha_{\Delta t} \widetilde{p}^{n+1} + \sqrt{\frac{1-\alpha_{\Delta t}^2}{\beta} M}\, G^n, \end{aligned}$$

when $M = \mathrm{Id}$ and in the limit as $\gamma \to +\infty$, is a consistent discretization of non-equilibrium Langevin dynamics (5.27). Henceforth we let $P_{\eta,\Delta t}$ denote the evolution operator associated with one of these schemes for a fixed value of the friction γ (see Remark 5.11 for a discussion of the limiting regime $\gamma \to +\infty$).

It can be shown that there exists a unique invariant measure $\mu_{\eta,\Delta t}$ for the Markov chains induced by $P_{\eta,\Delta t}$ by following the strategy outlined in Section 2.4. The crucial point is that the gradient structure of the force term is never used explicitly in the proofs of the Lyapunov and minorization conditions (see Assumptions 2.23 and 2.24) since we rely solely on the boundedness of the force. This in fact allows us to obtain exponential convergence results similar to (2.65) (Leimkuhler *et al.* 2015, Proposition 3.3): for any $\eta_* > 0$ and for any $s \geqslant 0$, there exist $C, \lambda, \Delta t^* > 0$ (which depend on s and η_*) such that, for all $\Delta t \in (0, \Delta t^*]$ and for all $\eta \in [-\eta_*, \eta_*]$,

$$\|P_{\eta,\Delta t}^n\|_{\mathcal{B}(L^\infty_{\mathcal{K}_s,\Delta t,\eta})} \leqslant C\, \mathrm{e}^{-\lambda n \Delta t}, \quad \text{for all } n \in \mathbb{N},$$

where

$$L^\infty_{\mathcal{K}_s,\Delta t,\eta}(\mathcal{E}) = \left\{ \varphi \in L^\infty_{\mathcal{K}_s}(\mathcal{E}) \,\middle|\, \int_{\mathcal{E}} \varphi \,\mathrm{d}\mu_{\eta,\Delta t} = 0 \right\}.$$

Let us emphasize that we do not have any control over the convergence rate λ in terms of η^*, and it could well be that λ goes to 0 as η^* increases.

The following result provides error estimates for the invariant measure of splitting schemes such as (5.28) and (5.29) (see Theorem 3.4 in Leimkuhler *et al.* (2015)). The proof follows the same lines as the proof of Theorem 3.3, except that expansions are performed with respect to the two small parameters Δt and η.

Theorem 5.9 (error estimates on the invariant measure for $\eta \neq 0$). Set $\alpha = 1$ for first-order splitting schemes such as (5.28) and $\alpha = 2$ for second-order splitting schemes such as (5.29). Then there exist functions $f_{\alpha,0}, f_{\alpha,1} \in \mathcal{S}_0$ such that, for any smooth function $\varphi \in \mathcal{S}$, there is $\Delta t^*, \eta^*, K > 0$ (depending on φ) for which, for all $\eta \in [-\eta^*, \eta^*]$ and $\Delta t \in (0, \Delta t^*]$,

$$\int_{\mathcal{E}} \varphi \, \mathrm{d}\mu_{\eta,\Delta t} = \int_{\mathcal{E}} \varphi(1 + \eta f_{0,1} + \Delta t^\alpha f_{\alpha,0} + \eta \Delta t^\alpha f_{\alpha,1}) \, \mathrm{d}\mu + r_{\varphi,\eta,\Delta t}, \quad (5.30)$$

where $f_{0,1}$ is the unique solution of the Poisson equation

$$\mathcal{L}_0^* f_{0,1} = -\widetilde{\mathcal{L}}^* \mathbf{1} = -\beta F^T M^{-1} p,$$

and

$$|r_{\varphi,\eta,\Delta t}| \leqslant K(\eta^2 + \Delta t^{\alpha+1}), \quad \frac{|r_{\varphi,\eta,\Delta t} - r_{\varphi,0,\Delta t}|}{\eta} \leqslant K(\eta + \Delta t^{\alpha+1}).$$

Let us now comment on (5.30). In this formula, the function $f_{0,1}$ encodes the linear response of the invariant measure when the perturbation is turned on for the continuous dynamics (see (5.12)), while $f_{\alpha,0}$, to leading order, accounts for the perturbation induced by the use of finite time steps. As shown below in Corollary 5.10, the errors on transport coefficients are determined at leading order by the cross-term of order $\eta \Delta t^\alpha$, which involves the correction function $f_{\alpha,1}$. Note that the remainder term $r_{\varphi,\eta,\Delta t}$ now collects higher-order terms both as powers of the time step Δt and the non-equilibrium parameter η. The estimates we obtain on the remainder, however, allow us to take the linear response limit $\eta \to 0$, as made precise by the following error estimate on the transport coefficient (which is an immediate consequence of Theorem 5.9). In order to state the result, we introduce the reference linear response for an observable $\varphi \in \mathcal{S}$, namely

$$\mathscr{R}_{\varphi,0} = \lim_{\eta \to 0} \frac{1}{\eta} \left(\int_{\mathcal{E}} \varphi \, \mathrm{d}\mu_\eta - \int_{\mathcal{E}} \varphi \, \mathrm{d}\mu \right),$$

and its numerical approximation,

$$\mathscr{R}_{\varphi,\Delta t} = \lim_{\eta \to 0} \frac{1}{\eta} \left(\int_{\mathcal{E}} \varphi \, \mathrm{d}\mu_{\eta,\Delta t} - \int_{\mathcal{E}} \varphi \, \mathrm{d}\mu_{0,\Delta t} \right). \quad (5.31)$$

It is often the case that the observable φ of interest has a vanishing average with respect to μ, as for the function $\widetilde{\mathcal{L}}^* \mathbf{1} = \beta F^T M^{-1} p$ used to compute the mobility in (5.10). Even in such cases, φ generally has a non-zero average with respect to the invariant measure $\mu_{0,\Delta t}$ of the numerical scheme associated with a discretization of the equilibrium dynamics, so that it is indeed important to subtract the average obtained with $\eta = 0$ in (5.31).

Corollary 5.10 (error estimates on linear responses). Under the assumptions of Theorem 5.9, and for any $\varphi \in \mathcal{S}$, there exist $\Delta t^* > 0$ and a constant $K > 0$ such that

$$\mathscr{R}_{\varphi,\Delta t} = \mathscr{R}_{\varphi,0} + \Delta t^\alpha \int_{\mathcal{E}} \varphi \, f_{\alpha,1} \, \mathrm{d}\mu + \Delta t^{\alpha+1} r_{\varphi,\Delta t},$$

with $|r_{\varphi,\Delta t}| \leqslant K$ when $0 < \Delta t \leqslant \Delta t^*$.

Note that, in contrast with the error estimates provided by the Green–Kubo formulae, the error is of order Δt^α without any need to modify the observable. This makes the linear response approach more attractive than Green–Kubo techniques when α is large and the correction function (5.25) is difficult to compute.

As an application, we obtain the following estimate on the numerically computed mobility:

$$\begin{aligned} \rho_{F,\Delta t} &= \lim_{\eta \to 0} \frac{1}{\eta} \left(\int_{\mathcal{E}} F^T M^{-1} p \, \mu_{\eta,\Delta t}(\mathrm{d}q \, \mathrm{d}p) - \int_{\mathcal{E}} F^T M^{-1} p \, \mu_{0,\Delta t}(\mathrm{d}q \, \mathrm{d}p) \right) \\ &= \rho_F + \Delta t^\alpha \int_{\mathcal{E}} F^T M^{-1} p \, f_{\alpha,1} \, \mathrm{d}\mu + \Delta t^{\alpha+1} r_{\Delta t}, \end{aligned} \tag{5.32}$$

where ρ_F is defined in (5.10). This error estimate is illustrated in Figures 5.2 and 5.3 for the same system as in Figure 5.1. More precisely, we check in Figure 5.2 that, for a given time step Δt, the average velocity in the direction F is indeed linear with respect to η for η sufficiently small. The corresponding slope gives an estimate of $\rho_{F,\Delta t}$. These estimates are then reported as a function of Δt in Figure 5.3. They extrapolate to the same value at $\Delta t = 0$, with errors of order Δt for first-order splitting schemes, and Δt^2 for second-order splitting schemes, as expected from Corollary 5.10.

Remark 5.11 (overdamped limits). As explained in Leimkuhler *et al.* (2015, Section 3.4), it is possible to study the overdamped limit $\gamma \to +\infty$ in the above results, and in particular to obtain error estimates on the invariant measure and on the linear response which hold uniformly for $\gamma \geqslant 1$.

5.4. Variance reduction for non-equilibrium systems

One of the difficulties with the computation of average properties of non-equilibrium systems is that standard variance reduction techniques, such as those described in Section 3.4, cannot be used as such, as we illustrate below. We believe that finding appropriate variance reduction techniques for non-reversible dynamics is a challenging and interesting open problem.

5.4.1. Importance sampling

If the drift term $-\nabla V$ of equilibrium (overdamped) Langevin dynamics is modified to $-\nabla(V + \widetilde{V})$, the changes in the invariant probability measure

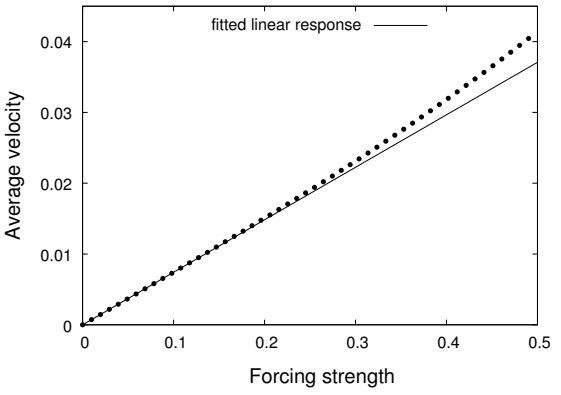


Figure 5.2. Linear response of the numerical approximation of the average velocity $\mathbb{E}_\eta(F^T M^{-1} p)$ as a function of η for the scheme associated with $P_{\Delta t}^{\gamma C, B+\eta\widetilde{\mathcal{L}}, A, B+\eta\widetilde{\mathcal{L}}, \gamma C}$, for $\Delta t = 0.01$. A linear fit on the first ten values gives a slope of $\rho_{F,\Delta t} \simeq 0.07416$.

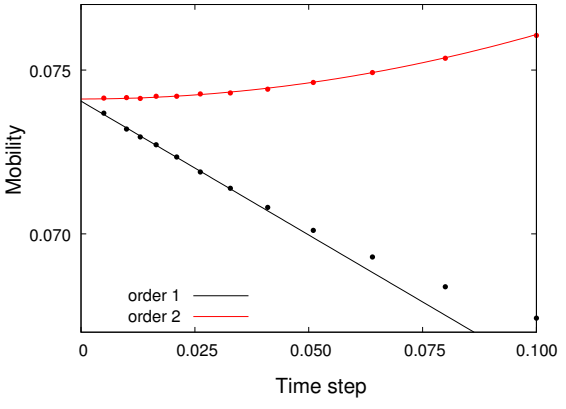


Figure 5.3. Estimated mobility $\rho_{F,\Delta t}$ for the first-order scheme $P_{\Delta t}^{A, B+\eta\widetilde{\mathcal{L}}, \gamma C}$ and second-order scheme $P_{\Delta t}^{\gamma C, B+\eta\widetilde{\mathcal{L}}, A, B+\eta\widetilde{\mathcal{L}}, \gamma C}$ as a function of the time step Δt. The fits give $\rho_{F,\Delta t} \simeq 0.0740 + 0.0817\Delta t$ and $\rho_{F,\Delta t} \simeq 0.0741 + 0.197\Delta t^2$, respectively.

can be explicitly written down: this measure now reads $\widetilde{Z}^{-1}\mathrm{e}^{-\beta(V+\widetilde{V})(q)}\mathrm{d}q$. On the other hand, consider non-equilibrium dynamics

$$\mathrm{d}q_t = b(q_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t,$$

with invariant measure $\psi_\infty(q)\,\mathrm{d}q$, perturbed by a gradient term, such as

$$\mathrm{d}q_t = (b(q_t) + \nabla A(q_t))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t.$$

We let $\psi_\infty^A(q)\,\mathrm{d}q$ denote the invariant measure of this process, assuming it exists. In general, $\psi_\infty^A(q)\,\mathrm{d}q$ is different from $Z^{-1}\psi_\infty(q)\mathrm{e}^{A(q)}\,\mathrm{d}q$ (consider for instance the simple example (5.5), for which the unique invariant probability measure is the uniform measure on $\mathcal{D} = \mathbb{T}$ when $V = 0$, which transforms into (5.7) when $V \neq 0$). The expression of ψ_∞^A is not known, and generally has no simple relationship with the expression of ψ_∞. It is therefore unclear how to use importance sampling strategies for non-equilibrium systems.

5.4.2. Stratification

For equilibrium systems, it is easy to construct constrained dynamics to sample the restriction of the invariant measure of the unconstrained dynamics to some submanifold. This is the principle of thermodynamic integration, for example (see Section 4.5). The invariant measure of constrained non-equilibrium dynamics may, on the other hand, have no relationship whatsoever with the invariant measure of the unconstrained non-equilibrium dynamics. Let us illustrate this point with a simple example. Consider the dynamics

$$\begin{aligned}\mathrm{d}q_{1,t} &= \partial_{q_2}U(q_{1,t}, q_{2,t}) + \sqrt{2}\,\mathrm{d}W_{1,t},\\ \mathrm{d}q_{2,t} &= -\partial_{q_1}U(q_{1,t}, q_{2,t}) + \sqrt{2}\,\mathrm{d}W_{2,t},\end{aligned}$$

on the state space $\mathbb{T}^2$, for a given C^∞ periodic function U. A simple computation shows that $\psi_\infty = \mathbf{1}_{\mathbb{T}^2}$ is an invariant probability measure. In addition, this is the unique invariant probability measure since the process is irreducible. Consider now the constraint $\xi(q) = 0$ for the choice $\xi(q) = q_2$. On the one hand, the restriction of ψ_∞ to the space $\{q \in \mathbb{T}^2 \mid \xi(q) = 0\} = \mathbb{T} \times \{0\}$ is $\overline{\psi}_\infty(q_1) = \mathbf{1}_{\mathbb{T}}$. On the other hand, the process constrained using a constraining force in the direction of $\nabla\xi$ reads

$$\mathrm{d}q_{1,t} = f(q_{1,t})\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_{1,t}, \quad f(q_1) = \partial_{q_2}U(q_1, 0).$$

In general, the invariant measure for this process is different from $\mathbf{1}_{\mathbb{T}}$. Indeed, introduce

$$F = \int_0^1 f, \quad V(q_1) = -\int_0^{q_1} (f(s) - F)\,\mathrm{d}s.$$

Note that V is a periodic function ($V(0) = V(1) = 0$), and that $f(q_1) = -V'(q_1)$. The computations performed in Section 5.1 show that the unique

invariant probability measure of the constrained process is

$$\psi_\infty(q_1) = Z^{-1} \int_0^1 \mathrm{e}^{V(q_1+y)-V(q_1)-Fy}\,\mathrm{d}y,$$

which is different from $\mathbf{1}_{\mathbb{T}}$ in general.

5.4.3. Control variate method

This method has been applied to non-equilibrium systems, but only for stochastic dynamics, for which the coupling between two configurations driven by the same random noise is strong – for example one-dimensional lattice gas systems (Goodman and Lin 2009). The general idea of the control variate method in this context is to simulate a system at equilibrium and a system subjected to a small external forcing, and to monitor the difference between the flux of interest in the non-equilibrium system, and the flux in the equilibrium one (which, up to statistical errors, should vanish). When the coupling is sufficiently strong, the variance of the difference of the fluxes is much smaller than the variance of the flux of the non-equilibrium system alone. On the other hand, this approach is very difficult to use in systems where the coupling is too weak, such as perturbations of Hamiltonian dynamics, or overdamped Langevin dynamics in regions where the Hessian of the potential is not positive definite. With some care, however, it is possible to estimate finite time correlation functions, as done by Ciccotti and Jacucci (1975).

An interesting question is: Is there a way to modify the dynamics in order to increase its coupling properties, while keeping the value of the linear response of the observable of interest fixed? This is, to the best of our knowledge, an open problem.

Remark 5.12 (adding non-reversible drifts). Non-reversible dynamics can also be obtained by perturbing reversible dynamics with a non-gradient force, which is divergence-free with respect to the equilibrium measure, however, and hence does not modify the invariant measure under consideration. At variance with the situations discussed above, where the non-reversible drift in the dynamics is fixed by the physical problem at hand (the computation of a given transport coefficient, induced for instance by the constant force ηF in (5.1)) and the non-equilibrium steady state is unknown, here the non-equilibrium steady state is fixed and equal to the equilibrium measure of the reversible dynamics under consideration. Moreover, the non-reversible drift is *chosen* by the user in order to accelerate the sampling procedure. For overdamped Langevin dynamics, there are several results confirming that the addition of non-gradient but divergence-free drifts improves the sampling, measured either in terms of convergence of the law of the process to the invariant distribution or in terms of asymptotic variance of observables of interest. In fact, both criteria are related to

spectral gap estimates. For further precision we refer to Hwang, Hwang-Ma and Sheu (1993, 2005) Lelièvre, Nier and Pavliotis (2013), Duncan *et al.* (2016) and Rey-Bellet and Spiliopoulos (2015, 2016), for example.

6. Computing dynamical quantities: path sampling and accelerated dynamics

In this section we discuss the computation of dynamical quantities, namely average properties over trajectories of Langevin or overdamped Langevin dynamics (1.9) and (1.8). As already explained above (see in particular Section 1.2), the main difficulty when computing average properties with such dynamics is metastability: the trajectories remain trapped for very long times in some metastable states. This means that behind the continuous state-space dynamics (1.9) or (1.8), there is a discrete state-space Markov process encoding the jumps from one metastable state to another. This has already been mentioned in Section 2.5 and is discussed at length in Section 6.3. Discrete state-space Markov dynamics are also very prevalent in molecular dynamics: these are called kinetic Monte Carlo or Markov state models (see for example Voter 2007). Continuous state-space models are then typically used to parametrize these Markovian models, that is, to compute the jump rates between metastable states, using Arrhenius (or Eyring–Kramers) formulae, for example (see (2.83)).

We will present some techniques to efficiently sample the exit event from a metastable state, or the reactive path between metastable states. In Section 6.1 we first introduce a preliminary tool, to be used many times thereafter, namely the Feynman–Kac formula, which we immediately illustrate with a simple application: the Jarzynski identity (which we already mentioned briefly in Section 4.5). We then present two ideas which have been followed to efficiently sample metastable paths: importance sampling approaches on path space, using control problems to build good importance functions (see Section 6.2), and the notion of quasi-stationary distributions (QSDs), to analyse the exit event from a metastable state (see Section 6.3). Finally, in Section 6.4, we present accelerated dynamics techniques, which were introduced by A. F. Voter in the late 1990s, and which can be analysed using the framework introduced in Section 6.3.

Let us finally make two preliminary remarks before embarking on the details. First, as explained in Sections 3 and 4, computing thermodynamic quantities (*i.e.* sampling the canonical measure) is a fairly well understood problem, with many techniques available that are used in many other scientific fields (*e.g.* computational statistics and uncertainty quantification). In contrast, the computation of dynamical properties is a numerical challenge that has been less explored by applied mathematicians, probably because it is more specific to molecular dynamic simulations. Many techniques

are available, and this is a very lively subject with new methods appearing at a rapid pace. We will devote Section 6.5 to an overview of such techniques, to the best of our current knowledge. The second remark is more technical. In this section we mainly focus on overdamped Langevin dynamics (1.8) when presenting mathematical results, but the reader should bear in mind that the algorithms presented apply equally to Langevin dynamics (1.9). We will comment below on the mathematical questions raised by the extensions of the proofs to Langevin dynamics.

6.1. Feynman–Kac formula and the Jarzynski identity

6.1.1. Feynman–Kac formula

In this subsection we introduce the Feynman–Kac formula which, together with the Fokker–Planck equation presented in Section 2.1, is the second natural way to associate a partial differential equation with a stochastic differential equation. It is a generalization of formula (2.8).

Let us consider a general stochastic process $(x_t)_{t\geqslant 0}$ in $\mathbb{R}^d$ satisfying the stochastic differential equation (1.11). Let us recall that we let $\mathcal{L}$ denote the associated infinitesimal generator (see (1.12)).

Proposition 6.1. Let S be a C^∞ bounded domain of $\mathbb{R}^d$, and let $f : S \to \mathbb{R}^d$, $v_0 : S \to \mathbb{R}^d$ and $\varphi : \partial S \to \mathbb{R}$ be three C^∞ functions. Let $v(t,y)$ be a smooth solution (C^1 in t and C^2 in y is sufficient) to the boundary value problem:

$$\begin{aligned} \partial_t v &= \mathcal{L}v + fv \quad \text{for } t \geqslant 0,\ y \in S, \\ v &= \varphi \qquad\quad\ \ \text{on } \partial S, \\ v(0,y) &= v_0(y). \end{aligned}$$

Then,

$$\begin{aligned} v(t,y) = \mathbb{E}^y\left(1_{\tau_S<t}\,\varphi(x_{\tau_S})\exp\left(\int_0^{\tau_S} f(x_s)\,\mathrm{d}s\right)\right) \\ + \mathbb{E}^y\left(1_{\tau_S\geqslant t}\,v_0(x_t)\exp\left(\int_0^{t} f(x_s)\,\mathrm{d}s\right)\right), \end{aligned} \tag{6.1}$$

where $(x_t)_{t\geqslant 0}$ is the process satisfying (1.11), τ_S is the first exit time of $(x_t)_{t\geqslant 0}$ from S,

$$\tau_S = \inf\{t \geqslant 0, x_t \notin S\},$$

and the superscript y in the notation $\mathbb{E}^y$ indicates that the initial condition is $x_0 = y \in \mathbb{R}^d$.

Let us emphasize that (6.1) is a representation formula: it gives a probabilistic representation to a smooth solution of the PDE, but does not say anything as such about the existence or the regularity of this solution.

Proof. Fix a time $t > 0$ and consider $u(s, y) = v(t - s, y)$ for $s \in [0, t]$. The function u satisfies

$$\begin{aligned} \partial_s u + \mathcal{L}u + fu &= 0 \qquad && \text{for } s \in [0,t],\ y \in S, \\ u &= \varphi \qquad && \text{on } \partial S, \\ u(t, y) &= v_0(y). \end{aligned}$$

Using Itô calculus, for all $s \in [0, \min(\tau_S, t)]$ we obtain

$$\begin{aligned} & u(s, x_s) \exp\left(\int_0^s f(x_r)\, \mathrm{d}r \right) \\ &\quad = u(0, x_0) + \int_0^s (\partial_r u + \mathcal{L}u + fu)(r, x_r) \exp\left(\int_0^r f(x_\rho)\, \mathrm{d}\rho \right) \mathrm{d}r \\ &\qquad + \sqrt{2\beta^{-1}} \int_0^s \exp\left(\int_0^r f(x_\rho)\, \mathrm{d}\rho \right) \nabla u(r, x_r) \cdot \sigma(x_r)\, \mathrm{d}W_r \\ &\quad = u(0, y) + M_s, \end{aligned}$$

where

$$M_s = \sqrt{2\beta^{-1}} \int_0^s \exp\left(\int_0^r f(x_\rho)\, \mathrm{d}\rho \right) \nabla u(r, x_r) \cdot \sigma(x_r)\, \mathrm{d}W_r$$

is a local martingale. Since u and f are assumed to be C^∞, and x_r lives in the bounded domain S up to time $\min(\tau_S, t)$, we conclude that

$$\begin{aligned} v(t, y) = u(0, y) &= \mathbb{E}^y \left(u\big(\min(t, \tau_S), x_{\min(t, \tau_S)}\big) \exp\left(\int_0^{\min(t, \tau_S)} f(x_r)\, \mathrm{d}r \right) \right) \\ &= \mathbb{E}^y \left(1_{\tau_S < t}\, u(\tau_S, x_{\tau_S}) \exp\left(\int_0^{\tau_S} f(x_r)\, \mathrm{d}r \right) \right) \\ &\qquad + \mathbb{E}^y \left(1_{\tau_S \geqslant t}\, u(t, x_t) \exp\left(\int_0^t f(x_r)\, \mathrm{d}r \right) \right) \\ &= \mathbb{E} \left(1_{\tau_S < t}\, \varphi(x_{\tau_S}) \exp\left(\int_0^{\tau_S} f(x_r)\, \mathrm{d}r \right) \right) \\ &\qquad + \mathbb{E} \left(1_{\tau_S \geqslant t}\, v_0(x_t) \exp\left(\int_0^t f(x_r)\, \mathrm{d}r \right) \right). \end{aligned}$$

This concludes the proof of (6.1). □

We have presented the Feynman–Kac formula in a specific setting and under strong regularity assumptions. Many extensions exist; see for example Friedman (1975, Section 6) for a general introduction and the recent work by Le Bris and Lions (2008) for generalizations to non-smooth data.

6.1.2. Jarzynski identity

Let us now present the Jarzynski identity (Jarzynski 1997), which is not a method for computing dynamical quantities but a free energy computation technique, like the adaptive importance sampling methods presented in Section 4. This method is described in this section since it is a simple application of the Feynman–Kac formula. For simplicity, we present it in the so-called alchemical setting, where the free energy is indexed by an external parameter λ rather than by the value of the reaction coordinate (as in Section 4.1). For extensions to the reaction coordinate case, we refer for example to Lelièvre *et al.* (2010, Chapter 5) and to Lelièvre *et al.* (2012).

Let us assume that the potential energy V is indexed by a parameter $\lambda \in [0,1]$, and that $(\lambda, q) \mapsto V_\lambda(q)$ is C^1 in λ and C^∞ in q. Typically, one could think of the situation $V_\lambda(q) = \lambda V_1(q) + (1-\lambda)V_0(q)$, where V_0 and V_1 are associated with two different models. For example, V_0 is a potential for a homogeneous system with N particles, and V_1 is the equivalent potential but for $N+1$ particles. Going from $\lambda = 0$ to $\lambda = 1$ is thus equivalent to inserting one particle in a bath of N particles – a process known in the physics and chemistry literature as Widom insertion (Widom 1963). Since this process needs not be physical, we can simulate extraordinary chemical operations such as the change of lead into gold, which explains the name 'alchemical setting'.

The free energy change associated with the transformation from V_0 to V_λ is defined as

$$F(\lambda) - F(0) = -\beta^{-1} \ln\left(\frac{Z_\lambda}{Z_0}\right)$$

(similar to the formula (4.5) in the reaction coordinate case), where for any $\lambda \in [0,1]$,

$$Z_\lambda = \int_{\mathcal{D}} \exp(-\beta V_\lambda).$$

Let us explain how the Jarzynski identity can be used to compute this quantity. The idea is to introduce a C^1 schedule $\Lambda : [0,T] \to [0,1]$ such that

$$\Lambda(0) = 0,$$

and then to simulate the non-homogeneous-in-time stochastic process

$$\mathrm{d}q_t = -\nabla V_{\Lambda(t)}(q_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t \tag{6.2}$$

with an initial condition q_0 distributed according to the probability measure $Z_0^{-1}\exp(-\beta V_0(q))\,\mathrm{d}q$. We associate with each trajectory a weight defined by

$$\mathcal{W}_t = \int_0^t \partial_\lambda V_{\Lambda(s)}(q_s)\dot{\Lambda}(s)\,\mathrm{d}s, \tag{6.3}$$

where $\dot\Lambda$ is the time derivative of Λ. Intuitively, the weight $\mathcal{W}_t$ represents the work associated with the alchemical process up to time t. Then we have the following interesting result.

Proposition 6.2. Assume that $(q_t)_{t\geqslant 0}$ satisfies (6.2) with q_0 distributed according to the probability measure $Z_0^{-1}\exp(-\beta V_0(q))\,\mathrm{d}q$. Then, for any time $t\in[0,T]$,

$$\mathbb{E}[\exp(-\beta\mathcal{W}_t)] = \exp[-\beta(F(\Lambda(t)) - F(0))], \tag{6.4}$$

where $\mathcal{W}_t$ is defined by (6.3).

This formula is called the Jarzynski identity. It shows that one can recover an equilibrium quantity (the free energy difference $F(\Lambda(t)) - F(0)$ from non-equilibrium simulations (*i.e.* the non-homogeneous-in-time dynamics (6.2)) started at equilibrium. In many cases of interest, only the free energy difference $F(1) - F(0)$ associated with the transformation from V_0 to V_1 is of interest: one then chooses Λ such that $\Lambda(T) = 1$, and the free energy difference is given by $F(1) - F(0) = -\beta^{-1}\ln(\mathbb{E}[\exp(-\beta\mathcal{W}_T)])$.

Proof. The proof is based on a Feynman–Kac formula in the spirit of (6.1) but adapted to the non-homogeneous-in-time setting. Indeed, for a fixed time t, let us consider a C^∞ solution $v : [0,t]\times\mathcal{D}\to\mathbb{R}$ to the backward-in-time partial differential equation

$$\begin{aligned} \partial_s v(s,x) &= -\mathcal{L}_s v(s,x) + \beta\partial_\lambda V_{\Lambda(s)}(x)\dot\Lambda(s)v(s,x) && \text{for } (s,x)\in[0,t]\times\mathcal{D},\\ v(t,x) &= f(x) && \text{for } x\in\mathcal{D}, \end{aligned}$$

where $\mathcal{L}_s = -\nabla V_{\Lambda(s)}\cdot\nabla + \beta^{-1}\Delta$ is the infinitesimal generator of (6.2). This is a standard parabolic equation which admits a unique C^∞ solution under standard regularity assumptions on the coefficients, the domain $\mathcal{D}$ and the function f. The Feynman–Kac formula in this context is given by

$$v(s,x) = \mathbb{E}^{s,x}(f(q_t)\,\mathrm{e}^{-\beta(\mathcal{W}_t-\mathcal{W}_s)}), \tag{6.5}$$

where the superscript s,x indicates that one considers the solution to (6.2) for $t\geqslant s$ with initial condition $q_s = x$. The proof of (6.5) is very similar to the proof of (6.1). Indeed, using Itô calculus over the time interval $[s,t]$, one obtains

$$\begin{aligned} v(t,q_t)\exp\left(-\beta\int_s^t \partial_\lambda V_{\Lambda(\rho)}(q_\rho)\dot\Lambda(\rho)\,\mathrm{d}\rho\right) &= v(s,q_s)\\ &+ \sqrt{2\beta^{-1}}\int_s^t \exp\left(-\beta\int_s^r \partial_\lambda V_{\Lambda(\rho)}(q_\rho)\dot\Lambda(\rho)\,\mathrm{d}\rho\right)\nabla v(r,q_r)\cdot\,\mathrm{d}W_r \end{aligned}$$

where we have used the partial differential equation satisfied by v to cancel

the bounded variation part in the Itô formula. Equation (6.5) then immediately follows by taking the expectation $\mathbb{E}^{s,x}$.

Now, it is easy to check that for any $s \in [0,t]$,

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(\int_{\mathcal{D}} v(s,x)\exp(-\beta V_{\Lambda(s)}(x))\,\mathrm{d}x\right) = 0. \tag{6.6}$$

Indeed, using the partial differential equation satisfied by v,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}s}\left(\int_{\mathcal{D}} v(s,x)\exp(-\beta V_{\Lambda(s)}(x))\,\mathrm{d}x\right)\\
&\quad= \int_{\mathcal{D}} (\partial_s v(s,x) - \beta\partial_\lambda V_{\Lambda(s)}(x)\dot{\Lambda}(s)v(s,x))\exp(-\beta V_{\Lambda(s)}(x))\,\mathrm{d}x\\
&\quad= -\int_{\mathcal{D}} \mathcal{L}_s v(s,x)\exp(-\beta V_{\Lambda(s)}(x))\,\mathrm{d}x = 0,
\end{aligned}$$

where the last equality is a consequence of the invariance of the canonical measure (see (1.17)). As a consequence of (6.6), we have

$$\int_{\mathcal{D}} v(t,x)\exp(-\beta V_{\Lambda(t)}(x))\,\mathrm{d}x = \int_{\mathcal{D}} v(0,x)\exp(-\beta V_{\Lambda(0)}(x))\,\mathrm{d}x.$$

Using (6.5), this equality can be rewritten as

$$\begin{aligned}
&\frac{Z_{\Lambda(t)}}{Z_0}\int_{\mathcal{D}} f(x)Z_{\Lambda(t)}^{-1}\exp(-\beta V_{\Lambda(t)}(x))\,\mathrm{d}x\\
&\qquad= \int_{\mathcal{D}} \mathbb{E}^{0,x}[f(q_t)\exp(-\beta\mathcal{W}_t)]Z_0^{-1}\exp(-\beta V_{\Lambda(0)}(x))\,\mathrm{d}x.
\end{aligned}$$

Taking $f = 1$, the left-hand side is $\exp(-\beta(F(\Lambda(t)) - F(0)))$ and the right-hand side is (using a conditioning argument on initial conditions) $\mathbb{E}(\exp(-\beta\mathcal{W}_t))$, where the weight $\mathcal{W}_t$ is computed using the solution $(q_t)_{t\geqslant 0}$ to (6.2) with initial conditions q_0 distributed according to the probability measure $Z_0^{-1}\exp(-\beta V_{\Lambda(0)}(x))\,\mathrm{d}x$. This concludes the proof of (6.4). □

The Jarzynski identity is mainly useful for extracting free energy differences from experimental measurements, by performing many non-equilibrium externally driven experiments between two states of interest (Hummer and Szabo 2001).

There are many extensions to this identity: generalization to other dynamics (Langevin, Hamiltonian, Nosé–Hoover dynamics), computation of free energy differences associated with a reaction coordinate (with driven constrained processes), and combination of forward-in-time and backward-in-time schedules (Jarzynski–Crooks identity). We refer to Lelièvre *et al.* (2010, Chapter 5) for a more comprehensive presentation.

6.2. Importance sampling on path space and Hamilton–Jacobi equations

In this section we present ideas that were developed in order to construct optimal importance sampling functions for some observables which are functions of the path of the stochastic process. This can be seen as an extension of importance sampling for canonical averages, as presented in Section 3.4.3. In this setting, partial differential equations (and in particular Hamilton–Jacobi equations) play a crucial role.

6.2.1. General setting

We consider overdamped Langevin dynamics (1.8), which we recall for convenience:

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

In this section we assume for simplicity that the initial condition is deterministic: $q_0 = x_0$ for a fixed $x_0 \in \mathcal{D}$. Assume that the observable of interest is

$$I = \mathbb{E}\left[\exp\left(g(q_\tau) + \int_0^\tau f(q_s)\,\mathrm{d}s\right)\right] \tag{6.7}$$

for given functions $f : \mathcal{D} \to \mathbb{R}$ and $g : \mathcal{D} \to \mathbb{R}$ and a stopping time τ, defined by

$$\tau = \inf\{t \geqslant 0,\, q_t \in D\},$$

where $D \subset \mathcal{D}$ is a given domain. Let us give two examples.

- One could think of $D = S^c$, the complement of a bounded set S. Then, when $f = 0$, (6.7) is the average of the test function $\exp(g)$ with respect to the first exit point probability on ∂S (which thus characterizes the exit point distribution by varying g). When $g = 0$ and $f = \lambda$, (6.7) is $\mathbb{E}(\exp(\lambda\tau))$, where τ is the exit time from S, which thus gives the moment generating function of τ. Computing (6.7) is thus important in understanding the exit event from S (see Section 6.3 for more discussion of the exit event from a state S).
- Another example[17] is $D = A \cup B$, where A and B are two metastable states, $f = 0$ and $\exp(g) = 1_B$, in which case (6.7) is the probability of reaching B before A. Trajectories starting from x_0 close to A and which reach B before A are called reactive trajectories (Hummer 2004, E and Vanden-Eijnden 2004, Lu and Nolen 2015) and are important to sample in order to identify the transition states between A and B, for example.

[17] The reasoning here is partially formal since it requires us to choose $g = 0$ on B and $g = -\infty$ on B^c. One should think of $g = -M$ on B^c, with a large M. See for example Dupuis, Spiliopoulos and Wang (2012, Proposition 4.2) for a rigorous treatment.

Typically, approximating I by an empirical average

$$\frac{1}{N}\sum_{n=1}^{N}\hat{I}_n,$$

where $\hat{I}_n$ are independent realizations of the random variable

$$\hat{I} = \exp\left(g(q_\tau) + \int_0^\tau f(q_s)\,\mathrm{d}s\right), \tag{6.8}$$

yields a large statistical error since

$$\sqrt{\frac{\mathrm{Var}(\hat{I})}{I^2}}$$

is large. For instance, in the second example above, I is the probability of reaching B before A, which is very small, so the relative error

$$\sqrt{\frac{\mathrm{Var}(\hat{I})}{I^2}} = \sqrt{\frac{I(1-I)}{I^2}}$$

is large. We would therefore like to apply an importance sampling procedure to reduce the variance in the estimation of I.

6.2.2. Biased dynamics

Let us assume that we consider the biased dynamics

$$\mathrm{d}\tilde{q}_t = -\nabla\big(V + \widetilde{V}\big)(\tilde{q}_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t, \tag{6.9}$$

as for classical importance sampling (see (3.36)). Here $\widetilde{V} : \mathcal{D} \to \mathbb{R}$ is a C^∞ biasing potential. We assume that $\tilde{q}_0 = q_0 = x_0$. Using the Girsanov theorem (see for example Revuz and Yor 1999, Chapter VIII), one can show that[18]

$$I = \mathbb{E}[\hat{I}_{\widetilde{V}}],$$

[18] The Girsanov theorem holds under some assumptions on $\widetilde{V}$ and on the stopping time τ. One way to make the argument fully rigorous is to consider from the beginning the observable $I = \mathbb{E}[\exp(g(q_\tau) + \int_0^\tau f(q_s)\,\mathrm{d}s)1_{\tau<T}]$ for a fixed time $T > 0$; see for example Dupuis, Spiliopoulos and Zhou (2015). Another approach to make this reasoning rigorous is to directly prove that $\mathbb{E}^x(\hat{I}_{\widetilde{V}}) = \mathbb{E}^x(\hat{I})$ for all $x \in D^c$, by using the Feynman–Kac formula (6.1) to show that $\mathbb{E}^x(\hat{I}_{\widetilde{V}})$ and $\mathbb{E}^x(\hat{I})$ are solutions of the same partial differential equation. Here we stick to a formal and hopefully intuitive presentation.

where

$$\hat{I}_{\widetilde{V}} = \exp\left(g(\tilde{q}_{\tilde{\tau}}) + \int_0^{\tilde{\tau}} f(\tilde{q}_s)\,\mathrm{d}s \right) \times \exp\left(\int_0^{\tilde{\tau}} \frac{1}{\sqrt{2\beta^{-1}}} \nabla \widetilde{V}(\tilde{q}_s) \cdot \mathrm{d}W_s - \frac{1}{2} \int_0^{\tilde{\tau}} \frac{1}{2\beta^{-1}} |\nabla \widetilde{V}|^2(\tilde{q}_s)\,\mathrm{d}s \right), \tag{6.10}$$

where $\tilde{\tau}$ is now the first hitting time of D for $(\tilde{q}_t)_{t\geqslant 0}$. This is the equivalent of the unbiasing formula (3.37) on path space: the Girsanov formula gives the change of measure so that averages with respect to the law $(q_t)_{0\leqslant t\leqslant \tau}$ can be obtained from weighted averages with respect to the law of $(\tilde{q}_t)_{0\leqslant t\leqslant \tilde{\tau}}$. Using Itô calculus, we have

$$\widetilde{V}(\tilde{q}_t) - \widetilde{V}(x_0) = \int_0^t (-\nabla V \cdot \nabla \widetilde{V} - |\nabla \widetilde{V}|^2 + \beta^{-1} \Delta \widetilde{V})(\tilde{q}_s)\,\mathrm{d}s + \sqrt{2\beta^{-1}} \int_0^t \nabla \widetilde{V}(\tilde{q}_s) \cdot \mathrm{d}W_s, \tag{6.11}$$

which, once inserted in (6.10), yields

$$\hat{I}_{\widetilde{V}} = \exp\left[g(\tilde{q}_{\tilde{\tau}}) + \int_0^{\tilde{\tau}} f(\tilde{q}_s)\,\mathrm{d}s + \frac{1}{2\beta^{-1}} \big(\widetilde{V}(\tilde{q}_{\tilde{\tau}}) - \widetilde{V}(x_0) \big) + \frac{1}{2\beta^{-1}} \int_0^{\tilde{\tau}} \left(\nabla V \cdot \nabla \widetilde{V} + \frac{1}{2} |\nabla \widetilde{V}|^2 - \beta^{-1} \Delta \widetilde{V} \right)(\tilde{q}_s)\,\mathrm{d}s \right]. \tag{6.12}$$

The aim is now to build a good biasing potential $\widetilde{V}$ such that $\mathrm{Var}(\hat{I}_{\widetilde{V}}) \ll \mathrm{Var}(\hat{I})$. Let us consider

$$\widetilde{V} = -2\beta^{-1} \ln u, \tag{6.13}$$

where, for $x \in D^c$,

$$u(x) = \mathbb{E}^x \left[\exp\left(g(q_\tau) + \int_0^\tau f(q_s)\,\mathrm{d}s \right) \right].$$

We recall that $(q_t)_{t\geqslant 0}$ satisfies overdamped Langevin dynamics (1.8) and that the superscript x in $\mathbb{E}^x$ indicates that $q_0 = x$. Notice in particular that $I = u(x_0)$. From Proposition 6.1 (formally consider the limit $t \to \infty$ in (6.1)), the function $u : D^c \to \mathbb{R}$ satisfies the partial differential equation

$$\begin{aligned} \mathcal{L}u + fu &= 0 \quad && \text{in } D^c, \\ u &= \exp(g) \quad && \text{on } \partial D, \end{aligned} \tag{6.14}$$

where $\mathcal{L} = -\nabla V \cdot \nabla + \beta^{-1} \Delta$ is the infinitesimal generator of overdamped

Langevin dynamics introduced in (1.13). Therefore $\widetilde{V} = -2\beta^{-1} \ln u$ satisfies

$$\begin{aligned} &\nabla V \cdot \nabla \widetilde{V} - \beta^{-1} \Delta \widetilde{V} + \frac{1}{2} |\nabla \widetilde{V}|^2 + 2\beta^{-1} f = 0 \quad \text{in } D^c, \\ &\widetilde{V} = -2\beta^{-1} g \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{on } \partial D \end{aligned} \tag{6.15}$$

(this is the celebrated Cole–Hopf transformation). Using $\widetilde{V} = -2\beta^{-1} \ln u$ as a biasing potential, one thus obtains from (6.12)

$$\begin{aligned} \hat{I}_{\widetilde{V}} &= \exp\left(g(\tilde{q}_{\tilde{\tau}}) + \frac{1}{2\beta^{-1}} (\widetilde{V}(\tilde{q}_{\tilde{\tau}}) - \widetilde{V}(x_0)) \right) \\ &= \exp\left(-\frac{1}{2\beta^{-1}} \widetilde{V}(x_0) \right) = u(x_0) = I. \end{aligned} \tag{6.16}$$

Let us summarize this discussion in a lemma.

Lemma 6.3. Consider the random variable $\hat{I}_{\widetilde{V}}$ defined by (6.12), built using the biased dynamics (6.9). This random variable is an unbiased estimator of I (defined by (6.7)):

$$\mathbb{E}(\hat{I}_{\widetilde{V}}) = I.$$

Moreover, let us assume that the biasing potential $\widetilde{V}$ is defined by (6.13). Then, $\hat{I}_{\widetilde{V}}$ is a zero variance estimator of I: almost surely,

$$\hat{I}_{\widetilde{V}} = I.$$

Therefore, we see that using $\widetilde{V} = -2\beta^{-1} \ln u$ as a biasing potential, $\hat{I}_{\widetilde{V}}$ is a zero variance estimator of I: the biasing potential $\widetilde{V} = -2\beta^{-1} \ln u$ is optimal in terms of variance. Of course, this result is not surprising, since being able to compute $\widetilde{V}$ requires us to know the quantity we would like to estimate from the very beginning:

$$I = u(x_0) = \exp(-2\beta^{-1} \widetilde{V}(x_0)). \tag{6.17}$$

This is very much reminiscent of the discussion of the optimal importance functions for classical importance sampling approaches in Section 3.4.3. This result also explains why we restrict ourselves *ab initio* to biasing drifts which are gradients: this is sufficient to get zero variance estimators. Let us mention that this approach can be easily generalized to non-homogeneous-in-time stochastic differential equations. We refer to Vaikuntanathan and Jarzynski (2008) for examples related to the dynamics (6.2).

6.2.3. Approximating $\widetilde{V}$ in practice

The above computations are nonetheless interesting for the following reason: if an approximation $\widetilde{V}$ of the optimal bias $-2\beta^{-1} \ln u$ is available, it can be used as a bias in the dynamics, to sample the estimator $\hat{I}_{\widetilde{V}}$. The closer $\widetilde{V}$

is to $-2\beta^{-1}\ln u$, the better is the associated Monte Carlo procedure. Now the question is how one should approximate the solution to (6.14), or equivalently, the solution to (6.15). One idea that has been explored in the literature is to interpret the solution to (6.15) as an optimal control problem, using the link relating Hamilton–Jacobi–Bellman equations to control problems. Indeed, (6.15) can be rewritten as

$$\begin{aligned} &\min_{u\in\mathcal{D}}\left(\frac{1}{2}|u|^2-(\nabla V+u)\cdot\nabla\widetilde{V}+\beta^{-1}\Delta\widetilde{V}-2\beta^{-1}f\right)=0 \quad \text{in } D^c,\\ &\widetilde{V}=-2\beta^{-1}g \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{on } \partial D. \end{aligned} \tag{6.18}$$

It is also well known that, under some regularity assumptions on the coefficients that we do not make precise (see *e.g.* Fleming 1977, Fleming and Soner 2006), the solution to this equation satisfies

$$\widetilde{V}(x)=\inf_{(u_s)_{s\geqslant 0}\in\mathcal{U}}\mathbb{E}^x\left(\frac{1}{2}\int_0^{\tau^u}|u_t|^2\,\mathrm{d}t-2\beta^{-1}\int_0^{\tau^u}f(q_t^u)\,\mathrm{d}t-2\beta^{-1}g(q_{\tau^u}^u)\right), \tag{6.19}$$

where $(q_t^u)_{t\geqslant 0}$ satisfies the stochastic differential equation

$$\mathrm{d}q_t^u=-(\nabla V(q_t^u)+u_t)\,\mathrm{d}t+\sqrt{2\beta^{-1}}\,\mathrm{d}W_t$$

and τ^u is the first entrance time in D for $(q_t^u)_{t\geqslant 0}$. We recall that the superscript x in (6.19) means that the stochastic process starts from x: $q_0^u=x$. Furthermore, $\mathcal{U}$ denotes the set of adapted stochastic processes with values in $\mathcal{D}$. Let us recall quickly why (6.19) holds. The proof is in two steps. For $\widetilde{V}$ solution to (6.15) (or equivalently (6.18)) and any stochastic process $(u_s)_{s\geqslant 0}\in\mathcal{U}$, we first prove that

$$\widetilde{V}(x)\leqslant\mathbb{E}^x\left(\frac{1}{2}\int_0^{\tau^u}|u_t|^2\,\mathrm{d}t-2\beta^{-1}\int_0^{\tau^u}f(q_t^u)\,\mathrm{d}t-2\beta^{-1}g(q_{\tau^u}^u)\right). \tag{6.20}$$

Indeed, by Itô calculus, we have

$$\begin{aligned}\widetilde{V}(q_{\tau^u}^u)=\widetilde{V}(q_0^u)&+\int_0^{\tau^u}(-(\nabla V(q_t^u)+u_t)\cdot\nabla\widetilde{V}(q_t^u)+\beta^{-1}\Delta\widetilde{V}(q_t^u))\,\mathrm{d}t\\&+\sqrt{2\beta^{-1}}\int_0^{\tau^u}\nabla\widetilde{V}(q_t^u)\cdot\mathrm{d}W_t.\end{aligned}$$

Taking the expectation and applying (6.18) to deduce

$$-(\nabla V(q)+u)\cdot\nabla\widetilde{V}(q)+\beta^{-1}\Delta\widetilde{V}(q)\geqslant 2\beta^{-1}f(q)-\frac{1}{2}|u|^2,$$

we obtain (6.20). Now, to conclude the proof of (6.19), it is sufficient to exhibit a process $(u_s^*)_{s\geqslant 0}$ such that

$$\widetilde{V}(x) = \mathbb{E}^x\left(\frac{1}{2}\int_0^{\tau^{u^*}} |u_t^*|^2\,\mathrm{d}t - 2\beta^{-1}\int_0^{\tau^{u^*}} f(q_t^{u^*})\,\mathrm{d}t - 2\beta^{-1}g(q_{\tau^{u^*}}^{u^*})\right), \quad (6.21)$$

where $\widetilde{V}$ is again the solution to (6.15). This is the second step of the proof, which considers the *feedback control*

$$u_t^* = \nabla\widetilde{V}(\tilde{q}_t), \quad (6.22)$$

where $(\tilde{q}_t)_{t\geqslant 0}$ satisfies (6.9), so that, with the previous notation, $(\tilde{q}_t)_{t\geqslant 0} = (q_t^{u^*})_{t\geqslant 0}$ and $\tilde{\tau} = \tau^{u^*}$. Using (6.15) and Itô calculus in succession, we obtain

$$\begin{aligned}
&\mathbb{E}^x\left(\frac{1}{2}\int_0^{\tilde{\tau}} |\nabla\widetilde{V}(\tilde{q}_t)|^2\,\mathrm{d}t - 2\beta^{-1}\int_0^{\tilde{\tau}} f(\tilde{q}_t)\,\mathrm{d}t - 2\beta^{-1}g(\tilde{q}_{\tilde{\tau}})\right)\\
&\quad = \mathbb{E}^x\left(\int_0^{\tilde{\tau}} (|\nabla\widetilde{V}(\tilde{q}_t)|^2 + \nabla V\cdot\nabla\widetilde{V}(\tilde{q}_t) - \beta^{-1}\Delta\widetilde{V}(\tilde{q}_t))\,\mathrm{d}t - 2\beta^{-1}g(\tilde{q}_{\tilde{\tau}})\right)\\
&\quad = \mathbb{E}^x(\widetilde{V}(\tilde{q}_0) - \widetilde{V}(\tilde{q}_{\tilde{\tau}}) - 2\beta^{-1}g(\tilde{q}_{\tilde{\tau}})) = \widetilde{V}(x).
\end{aligned}$$

This is very close to the computations leading to (6.11) above. This concludes the proof of (6.21), and thus of (6.19). In particular, it shows that the infimum in the right-hand side of (6.19) is attained for the feedback control (6.22).

Using the characterization (6.19) of the optimal biasing potential $\widetilde{V}$ as a solution to a control problem, many ideas have been explored to approximate $\widetilde{V}$. For example, Hartmann and Schütte (2012) have proposed solving the control problem by discretizing the minimization set over feedback controls of the form $u_t = \big(\sum_{j=1}^m a_j\nabla\varphi_j\big)(q_t^u)$ for fixed basis functions $\varphi_j : \mathcal{D}\to\mathbb{R}$ (Galerkin procedure), or using coarse-grained models to approximate the solution to the control problem (see Schütte, Winkelmann and Hartmann 2012, Zhang *et al.* 2014, Hartmann, Schütte, Weber and Zhang 2015, Banisch and Hartmann 2016). Another approach, followed in Dupuis *et al.* (2012, 2015) and Vanden-Eijnden and Weare (2012), for example, consists in sending the temperature to zero ($\beta\to\infty$) while scaling f and g such that $2\beta^{-1}f = \tilde{f}$ and $2\beta^{-1}g = \tilde{g}$, in order to replace the stochastic control problem with a deterministic control problem. Formally, (6.15) becomes

$$\begin{aligned}
\nabla V\cdot\nabla\widetilde{V} + \frac{1}{2}|\nabla\widetilde{V}|^2 + \tilde{f} = 0 \quad &\text{in } D^c,\\
\widetilde{V} = -\tilde{g} \quad &\text{on } \partial D,
\end{aligned} \quad (6.23)$$

and (6.19) can be rewritten as

$$\widetilde{V}(x) = \inf_{(u_s)_{s\geqslant 0}\in\mathcal{U}} \left(\frac{1}{2}\int_0^{\tau^u} |u_t|^2 \,\mathrm{d}t - \int_0^{\tau^u} \tilde{f}(q_t^u)\,\mathrm{d}t - \tilde{g}(q_{\tau^u}^u) \right), \tag{6.24}$$

where $(q_t^u)_{t\geqslant 0}$ satisfies the ordinary differential equation

$$q_t^u = x + \int_0^t (-\nabla V(q_t^u) - u_t)\,\mathrm{d}t.$$

From the definition (6.13) of the optimal bias $\widetilde{V}$ at positive temperature, one observes that the solution to the problem (6.24) should be the limit

$$\lim_{\beta\to\infty} -2\beta^{-1} \ln\left(\mathbb{E}^x \left[\exp\left(\frac{\beta}{2} \left(\int_0^\tau \tilde{f}(q_s)\,\mathrm{d}s + \tilde{g}(q_\tau) \right) \right) \right] \right).$$

This is related to large deviation results: see Fleming (1977), Fleming and Soner (2006), Freidlin and Wentzell (1984) or Dupuis and Wang (2004). The idea is thus to use the rate function (here $\widetilde{V}$) associated with a large deviation principle (here in the limit $\beta \to \infty$) to build good importance sampling functions (see also Siegmund 1976, Bucklew 2004). Let us emphasize that our presentation here is very informal, and that proving rigorously that efficient estimators are obtained by using a large deviation principle requires some work: see for example Dupuis *et al.* (2012, 2015). In particular, many works by Dupuis *et al.* actually rely on smooth *subsolutions* of (6.23), which is sufficient to get asymptotically efficient importance sampling schemes. Other related works studying rare events via large deviations and control theory are those of Jack and Sollich (2010) and Chetrite and Touchette (2014).

Let us also mention the work by Kundu, Sabhapandit and Dhar (2011), where an importance sampling procedure based on a Girsanov transform is used for Langevin dynamics for efficient sampling of the energy flux in a system linking two thermal baths at different temperatures (for further references on this model, see (5.2) and the subsequent discussion).

6.2.4. Concluding remarks

Let us conclude this section with two comments.

First, in the second example given at the beginning of Section 6.2.1 ($D = A \cup B$, where A and B are two metastable states, $f = 0$ and $\exp(g) = 1_B$), the function $u(x) = \mathbb{P}^x(\tau_B < \tau_A)$ is the so-called committor function, which is the solution to (see (6.14))

$$\begin{aligned} \mathcal{L}u &= 0 \quad \text{in } (A\cup B)^c, \\ u &= 0 \quad \text{on } A, \\ u &= 1 \quad \text{on } B. \end{aligned}$$

This function plays an important role in importance sampling (as explained here) but also in optimal splitting techniques (Vanden-Eijnden, Venturoli, Ciccotti and Elber 2008, Guyader 2015). It is also very useful for theoretical purposes, to compute transition rates from A to B (E and Vanden-Eijnden 2004, Hummer 2004, Lu and Nolen 2015).

Second, the function u solution to (6.14), used here to build an optimal importance sampling scheme, can also be used to build an optimal control variate (see Section 3.4.2 for an introduction to variance reduction by control variates). Indeed, by Itô calculus, it is easy to check that if u satisfies (6.14) and $(q_t)_{t\geqslant 0}$ follows overdamped Langevin dynamics (1.8), then

$$u(q_t)\exp\left(\int_0^t f(q_s)\,\mathrm{d}s\right) = u(q_0) + \int_0^t (\mathcal{L}u + fu)(q_s)\exp\left(\int_0^s f(q_r)\,\mathrm{d}r\right)\mathrm{d}s + \sqrt{2\beta^{-1}}\int_0^t \exp\left(\int_0^s f(q_r)\,\mathrm{d}r\right)\nabla u(q_s)\cdot \mathrm{d}W_s$$

so that, by choosing $t=\tau$ (see (6.17), with $x_0 = q_0$),

$$\exp\left(g(q_\tau)+\int_0^\tau f(q_s)\,\mathrm{d}s\right) = I + \sqrt{2\beta^{-1}}\int_0^\tau \exp\left(\int_0^s f(q_r)\,\mathrm{d}r\right)\nabla u(q_s)\cdot \mathrm{d}W_s.$$

On the left-hand side, one recognizes the unbiased estimator $\hat{I}$ (see (6.8)) of I, and from this equality, one can see that

$$\hat{Z} = \sqrt{2\beta^{-1}}\int_0^\tau \exp\left(\int_0^s f(q_r)\,\mathrm{d}r\right)\nabla u(q_s)\cdot \mathrm{d}W_s$$

is an optimal control variate for $\hat{I}$, since $\mathbb{E}(\hat{Z}) = 0$ and

$$\mathrm{Var}(\hat{I} - \hat{Z}) = 0.$$

This shows that a good approximation of the solution to (6.14) could also be used to build a good control variate.

6.3. The quasi-stationary distribution and the exit event from a metastable state

As mentioned in the introduction of Section 6, the difficulty when sampling paths of overdamped Langevin or Langevin dynamics is that the system remains trapped for very long times in metastable states. In this section we analyse the exit event from such a metastable state, denoted by $S \subset \mathcal{D}$ in the following.

An interesting mathematical tool to study the exit event is the quasi-stationary distribution (QSD). The reason why a partial differential equation is involved in this problem is that the QSD attached to a state S is simply the first eigenvector of the Fokker–Planck operator with homogeneous Dirichlet boundary conditions on ∂S.

6.3.1. The quasi-stationary distribution

Let us consider a fixed state S (*i.e.* an open bounded[19] and smooth subset of the configurational space $\mathcal{D}$) and let us focus for simplicity on overdamped Langevin dynamics (1.8). Consider the first exit time from S:

$$\tau_S = \inf\{t \geqslant 0,\, q_t \notin S\},$$

where $(q_t)_{t\geqslant 0}$ follows overdamped Langevin dynamics (1.8).

Let us start with the definition of the QSD.

Definition 6.4. A probability measure ν_S with support in S is called a QSD for the Markov process $(q_t)_{t\geqslant 0}$ if and only if

$$\nu_S(A) = \frac{\int_S \mathbb{P}^x(q_t \in A, t < \tau_S)\, \nu_S(\mathrm{d}x)}{\int_S \mathbb{P}^x(t < \tau_S)\, \nu_S(\mathrm{d}x)}, \quad \text{for all } t > 0,\ A \subset S. \tag{6.25}$$

We recall that $\mathbb{P}^x$ denotes the probability measure under which $q_0 = x$. In other words, ν_S is a QSD if, when q_0 is distributed according to ν_S, the law of q_t, conditional on $(q_s)_{0\leqslant s\leqslant t}$ remaining in the state S, is still ν_S, for all positive t.

The QSD satisfies three properties which will be crucial in the following. We refer for example to Le Bris, Lelièvre, Luskin and Perez (2012) for detailed proofs of these results and to Collet, Martínez and San Martín (2013) for more general results on QSDs.

Proposition 6.5. Let $(q_t)_{t\geqslant 0}$ follow the dynamics (1.8) with an initial condition q_0 distributed according to a distribution ν_{init} with support in S. Then there exists a probability distribution ν_S with support in S such that, for any initial distribution ν_{init} with support in S,

$$\lim_{t\to\infty} \mathrm{Law}(q_t | \tau_S > t) = \nu_S. \tag{6.26}$$

The distribution ν_S is the QSD associated with S.

A consequence of this proposition is the existence and uniqueness of the QSD. The QSD is the long-time limit of the law of the (time marginal of the) process conditioned to stay in the state S: it can be seen as a 'local ergodic measure' for the stochastic process in S. This proposition can be useful for properly defining a metastable state. A metastable state is a state such that the typical exit time is much larger than the local equilibration time, namely the time to observe the convergence to the QSD in (6.26); see Section 6.4.1 for a precise discussion of this point.

Let us now give a second property of the QSD.

[19] Some of the results presented below can be generalized to an unbounded subset S. Boundedness implies the compactness of the embedding $H^1(S) \subset L^2(S)$ which is used in the proof of Proposition 6.6 to show the existence and uniqueness of the QSD.

Proposition 6.6. Let $\mathcal{L} = -\nabla V \cdot \nabla + \beta^{-1}\Delta$ be the infinitesimal generator of $(q_t)_{t\geqslant 0}$ (satisfying (1.8)). Let us consider the first eigenvalue and eigenfunction associated with the adjoint operator $\mathcal{L}^\dagger = \operatorname{div}(\nabla V + \beta^{-1}\nabla)$ with a homogeneous Dirichlet boundary condition:

$$\begin{aligned} \mathcal{L}^\dagger u_1 &= -\lambda_1 u_1 \quad \text{on } S, \\ u_1 &= 0 \qquad\quad\ \ \text{on } \partial S. \end{aligned} \tag{6.27}$$

Then, the QSD ν_S associated with S satisfies

$$\mathrm{d}\nu_S = \frac{u_1(x)\,\mathrm{d}x}{\int_S u_1(x)\,\mathrm{d}x},$$

where $\mathrm{d}x$ denotes the Lebesgue measure on S.

The QSD thus has a density with respect to Lebesgue measure, which is simply the ground state of the Fokker–Planck operator $\mathcal{L}^\dagger$ associated with the dynamics with absorbing boundary conditions. In (6.27), it can be shown that u_1 has a constant sign and hence can be chosen non-negative. In addition, $\lambda_1 > 0$. Let us give the main arguments of the proof of this proposition (see Le Bris *et al.* 2012 for more details).

Proof. Let us first notice the equivalence between two properties: $(-\lambda_1, u_1)$ is an eigenvalue–eigenfunction pair for $\mathcal{L}^\dagger$ (with zero Dirichlet boundary condition on ∂S) and $(-\lambda_1, u_1 \exp(\beta V))$ is an eigenvalue–eigenfunction pair for $\mathcal{L}$ (with zero Dirichlet boundary condition on ∂S). This is a simple consequence of the following identity: for any C^∞ test function $\varphi : \mathcal{D} \to \mathbb{R}$,

$$\mathcal{L}^\dagger(\varphi \exp(-\beta V)) = (\mathcal{L}\varphi)\exp(-\beta V).$$

Now, the operator $\mathcal{L}$ with homogeneous Dirichlet boundary condition is symmetric for the weighted $L^2(\nu)$ scalar product (where, we recall, $\nu(\mathrm{d}q) = Z_\nu^{-1}\exp(-\beta V(q))\,\mathrm{d}q$ is the canonical measure: see (1.5)). For any two C^∞ test functions φ and ψ which vanish on S^c,

$$\int_S \varphi(\mathcal{L}\psi)\,\mathrm{d}\nu = \int_S \psi(\mathcal{L}\varphi)\,\mathrm{d}\nu = -\beta^{-1}\int_S \nabla\varphi\cdot\nabla\psi\,\mathrm{d}\nu.$$

Since V is assumed to be C^∞ and S bounded, the inverse of the operator $\mathcal{L}$ from $L^2(\nu)$ to $L^2(\nu)$ is compact and symmetric. Thus $\mathcal{L}$ admits a discrete spectrum, and in particular a smallest eigenvalue $-\lambda_1$ associated with an eigenvector φ_1. Using standard arguments, one can show that $\varphi_1 > 0$ and $-\lambda_1$ is a non-degenerate eigenvalue.

Let us now introduce the probability density $u_1 = Z_S^{-1}\varphi_1\exp(-\beta V)$ with $Z_S = \int_S \varphi_1 \exp(-\beta V)$. As explained at the beginning of the proof, (λ_1, u_1) satisfies (6.27). Let us prove that $u_1(x)\,\mathrm{d}x$ is a QSD, by checking (6.25). Fix a subset A of S and introduce $v(t,x) = \mathbb{P}^x(q_t \in A, t < \tau_S)$ as well as

$\overline{v}(t,x) = \mathbb{P}^x(t < \tau_S)$. From Proposition 6.1, the function v satisfies

$$\begin{aligned} \partial_t v &= \mathcal{L} v \qquad && \text{for } t \geqslant 0,\ y \in S, \\ v &= 0 && \text{on } \partial S, \\ v(0,y) &= 1_A(y). \end{aligned}$$

Likewise, the function $\overline{v}$ satisfies the same equation, with initial condition $\overline{v}(0,y) = 1$. Therefore,

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_S \mathbb{P}^x(q_t \in A, t < \tau_S) u_1(x)\,\mathrm{d}x\right) \\ &\quad = Z_S^{-1} \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_S v(t,\cdot)\varphi_1 \exp(-\beta V)\right) \\ &\quad = Z_S^{-1} \int_S (\mathcal{L} v)\varphi_1 \exp(-\beta V) \\ &\quad = Z_S^{-1} \int_S v(\mathcal{L}\varphi_1) \exp(-\beta V) \\ &\quad = -\lambda_1 Z_S^{-1} \int_S v\varphi_1 \exp(-\beta V) \\ &\quad = -\lambda_1 \int_S \mathbb{P}^x(q_t \in A, t < \tau_S) u_1(x)\,\mathrm{d}x, \end{aligned}$$

which yields

$$\int_S \mathbb{P}^x(q_t \in A, t < \tau_S) u_1(x)\,\mathrm{d}x = \left(\int_A u_1(x)\,\mathrm{d}x\right) \exp(-\lambda_1 t). \tag{6.28}$$

Likewise, considering $\overline{v}$, it is easy to check that (taking $A = S$ in (6.28))

$$\int_S \mathbb{P}^x(t < \tau_S) u_1(x)\,\mathrm{d}x = \exp(-\lambda_1 t). \tag{6.29}$$

This concludes the proof in view of (6.25). □

Finally, the third property of the QSD concerns the exit event, when q_0 is distributed according to ν_S.

Proposition 6.7. Let us assume that q_0 is distributed according to the QSD ν_S in S. Then the law of the pair (τ_S, q_{τ_S}) (*i.e.* the first exit time and the first exit point) is fully characterized by the following properties.

- τ_S is exponentially distributed with parameter λ_1 (defined in (6.27)).
- τ_S is independent of q_{τ_S}.
- The law of q_{τ_S} is given as follows: for any bounded measurable function $\varphi : \partial S \to \mathbb{R}$,

$$\mathbb{E}^{\nu_S}(\varphi(q_{\tau_S})) = -\frac{\int_{\partial S} \varphi\, \partial_n u_1\,\mathrm{d}\sigma}{\beta\lambda_1 \int_S u_1(x)\,\mathrm{d}x}, \tag{6.30}$$

where σ denotes the Lebesgue measure on ∂S induced by the Lebesgue measure on $\mathcal{D}$ and the Euclidean scalar product, and $\partial_n u_1 = \nabla u_1 \cdot n$ denotes the outward normal derivative of u_1 (defined in (6.27)) on ∂S.

In (6.30), the superscript ν_S indicates that the initial condition q_0 is assumed to be distributed according to ν_S. Let us give a brief proof of this important result (see again Le Bris *et al.* 2012 for more details).

Proof. The fact that τ_S is exponentially distributed with parameter λ_1 is a direct consequence of (6.29). To prove the two remaining items, let us consider the solution v to the following PDE:

$$\begin{aligned} \partial_t v &= \mathcal{L} v \quad \text{for } t \geqslant 0,\ x \in S, \\ v &= \varphi \quad \text{on } \partial S, \\ v(0,x) &= 0 \quad \text{for } x \in S. \end{aligned}$$

From Proposition 6.1, we have the probabilistic representation formula

$$v(t,x) = \mathbb{E}^x(1_{\tau_S < t}\, \varphi(q_{\tau_S})).$$

Let us introduce

$$f(t) = \mathbb{E}^{\nu_S}(1_{\tau_S<t}\, \varphi(q_{\tau_S})) = \int_S \mathbb{E}^x(1_{\tau_S<t}\, \varphi(q_{\tau_S}))\nu_S(\mathrm{d}x) = Z_S^{-1} \int_S v(t,\cdot)u_1,$$

where $Z_S = \int_S u_1$. Using integration by parts and the fact that $u_1 = 0$ on ∂S, we obtain

$$\begin{aligned} f' &= Z_S^{-1} \int_S \partial_t v\, u_1 \\ &= Z_S^{-1} \int_S (\mathcal{L} v) u_1 = Z_S^{-1} \int_S (-\nabla V \cdot \nabla v + \beta^{-1} \Delta v) u_1 \\ &= Z_S^{-1} \int_S v\big(\mathrm{div}(u_1 \nabla V) + \beta^{-1} \Delta u_1\big) - Z_S^{-1} \beta^{-1} \int_{\partial S} v \partial_n u_1 \,\mathrm{d}\sigma \\ &= -\lambda_1 f - Z_S^{-1} \beta^{-1} \int_{\partial S} \varphi \partial_n u_1 \,\mathrm{d}\sigma. \end{aligned}$$

By solving this ordinary differential equation in f, we thus get

$$f(t) = \mathbb{E}^{\nu_S}(1_{\tau_S<t}\, \varphi(q_{\tau_S})) = \big(1 - \exp(-\lambda_1 t)\big) \int_{\partial S} \varphi \,\mathrm{d}\rho, \tag{6.31}$$

where ρ denotes the probability measure

$$\mathrm{d}\rho = -\frac{\beta^{-1}}{\lambda_1 Z_S} \partial_n u_1 \,\mathrm{d}\sigma.$$

Note that $\partial_n u_1 \leqslant 0$ and $Z_S \geqslant 0$ when u_1 is chosen positive in S, hence ρ indeed has a non-negative density with respect to σ. Since $1 - \exp(-\lambda_1 t) = \mathbb{P}^\nu(\tau_S < t)$, equation (6.31) shows that, if $q_0 \sim \nu_S$, the random variables

τ_S and q_{τ_S} are independent, and q_{τ_S} is distributed according to ρ. This concludes the proof of Proposition 6.7. □

Remark 6.8. The existence of the QSD and the convergence of the conditioned process towards the QSD for the Langevin process (1.9) require extra work compared to the overdamped Langevin process (1.8). For results in that direction, we refer to the recent paper by Nier (2014).

6.3.2. Exit event and the state-to-state dynamics

Proposition 6.7 explains the role of the QSD in relating the continuous-in-time process $(q_t)_{t\geqslant 0}$ to a jump process between metastable states.

Indeed, if the process remains for a long time in S, then, according to Proposition 6.5, it is approximately distributed according to the QSD. Therefore, to study the exit from S, it is relevant to consider a process starting from ν_S and to look at the pair (τ_S, q_{τ_S}) which fully characterizes the exit event from S. Proposition 6.7 shows that the exit event from the state S is Markovian. It can be rewritten as one step of a Markov jump process, thanks to two important properties: τ_S is exponentially distributed and independent of q_{τ_S}. This is reminiscent of what is assumed to build kinetic Monte Carlo (Voter 2007) or Markov state models (Sarich and Schütte 2013, Bowman, Pande and Noé 2014).

Let us make this more precise. The aim is to build from the continuous state-space dynamics $(q_t)_{t\geqslant 0}$ a discrete state-space dynamics, which jumps from state to state. Let us assume that we are given a mapping

$$\mathcal{S} : \mathcal{D} \to \mathbb{N}, \tag{6.32}$$

which to any given $x \in \mathcal{D}$ associates $\mathcal{S}(x)$, the label of the state in which x lies. The mapping $\mathcal{S}$ thus defines a partition of the state space. In the simple two-dimensional case of Figure 1.1, for example, one could use just two states: $\mathcal{S}(x, y)$ is 0 if $x < 0$ and 1 if $x > 0$. The state-to-state dynamics is then simply the dynamics

$$(\mathcal{S}(q_t))_{t\geqslant 0}, \tag{6.33}$$

where $(q_t)_{t\geqslant 0}$ is obtained by the Langevin (1.9) or overdamped Langevin (1.8) dynamics. Of course this is *not* a Markovian dynamics. If $\mathcal{S}$ is well chosen, however, it is *close* to a Markovian dynamics. In the next section we will come back to one way to correct for non-Markovian effects. We concentrate here on the parametrization of the Markov dynamics associated with the partition and the original dynamics. Let us assume that the process $(q_t)_{t\geqslant 0}$ remains in one of the components of the partition for a sufficiently long time, so that the limit in (6.26) is reached (see Proposition 6.9 for an estimate of this convergence time). Let S denote this state: $S = \mathcal{S}^{-1}(\{n\})$ for some $n \in \mathbb{N}$.

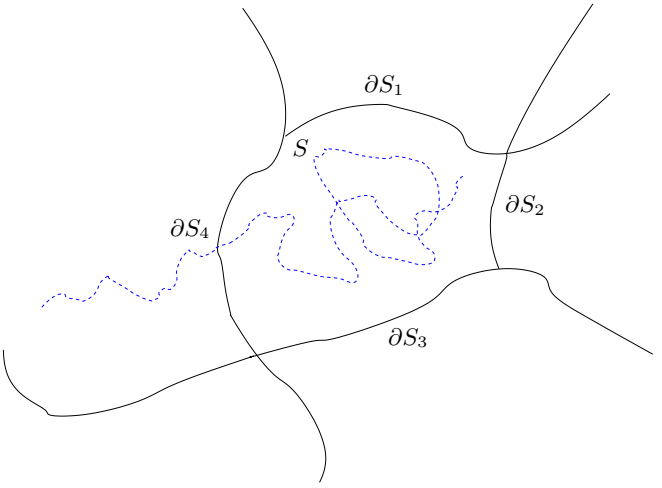


Figure 6.1. The state S with the partition of the boundary ∂S into four subsets $(\partial S_i)_{i=1,\dots,4}$. The dashed line represents the trajectory of a dynamics entering S.

Let us now assume that the state S is surrounded by I neighbouring states. The boundary ∂S is then divided into I disjoint subsets $(\partial S_i)_{i=1,\dots,I}$, each of them associated with an exit towards one of the neighbouring states (see Figure 6.1 for a situation where $I = 4$). The exit event from S is characterized by the pair $(\tau_S, \mathcal{I})$, where $\mathcal{I}$ is a random variable which gives the next visited state:

$$\text{for } i = 1, \dots, I, \quad \{\mathcal{I} = i\} = \{q_{\tau_S} \in \partial S_i\}.$$

Notice that τ_S and $\mathcal{I}$ are independent random variables. The kinetic Monte Carlo model is then parametrized as follows. Introduce

$$p(i) = \mathbb{P}(q_{\tau_S} \in \partial S_i) = -\frac{\int_{\partial S_i} \partial_n u_1 \, \mathrm{d}\sigma}{\beta \lambda_1 \int_S u_1 \, \mathrm{d}x}, \quad \text{for } i = 1, \dots, I.$$

For each exit region ∂S_i there is a corresponding rate $k(i) = \lambda_1 p(i)$. Now, if $(T_i)_{i=1,\dots,I}$ are independent random variables, where T_i is exponentially distributed with parameter $k(i)$, one can check that

$$(\tau_S, \mathcal{I}) \stackrel{\mathcal{L}}{=} \left(\min_{i \in \{1,\dots,I\}} (T_i), \arg\min_{i \in \{1,\dots,I\}} (T_i) \right),$$

where $\stackrel{\mathcal{L}}{=}$ denotes equality in law. A natural way to parametrize a kinetic Monte Carlo model associated with the partition $(\mathcal{S}^{-1}(\{n\}))_{n \in \mathbb{N}}$ is thus to consider the stochastic process starting from the QSD associated with each state, and to use the random variables (τ_S, q_{τ_S}) to parametrize the jump process. In practice, computing the rates $k(i)$ associated with S may be complicated. Most often, rates are evaluated using an approximation (Voter

2007, Berglund 2013): for instance the harmonic transition state theory, the Arrhenius law, or the Eyring–Kramers law (see formula (6.39) below).

We have discussed one natural way to relate a continuous state-space dynamics to a discrete state-space dynamics: introduce a partition of the state space and consider the position of the process within this partition. Another natural idea is to introduce an ensemble of disjoint subsets $(M_i)_{i\geqslant 1}$ and to project the dynamics $(q_t)_{t\geqslant 0}$ onto a discrete state-space dynamics by considering *the last visited milestone* (Buchete and Hummer 2008, Schütte *et al.* 2011). Notice that these subsets do not create a partition of the state space. They are sometimes called milestones (Faradjian and Elber 2004), target sets or core sets (Schütte *et al.* 2011) in the literature. The natural parametrization of the underlying jump process is then to consider, starting from a milestone (say M_1), the time to reach any of the other milestones $(M_j)_{j\neq 1}$ and the index of the next visited milestone. This requires us to study the reactive paths among the milestones, for which many techniques have been proposed (Dellago and Bolhuis 2009): for example, transition interface sampling (van Erp, Moroni and Bolhuis 2003, van Erp and Bolhuis 2005), forward flux sampling (Allen, Warren and ten Wolde 2005, Allen, Valeriani and ten Wolde 2009), the string method (E and Vanden-Eijnden 2004, E, Ren and Vanden-Eijnden 2002, E, Ren and Vanden-Eijnden 2005), and adaptive multilevel splitting techniques (Cérou and Guyader 2007, Cérou, Guyader, Lelièvre and Pommier 2011). Reactive paths are the pieces of a trajectory at equilibrium between metastable states: see Hummer (2004), E and Vanden-Eijnden (2004) and Lu and Nolen (2015). Let us finally discuss the error introduced by this approach, namely the Markovianity of the projected dynamics. On the one hand, in the limit of very small milestones,[20] the sequence of visited states (*i.e.* the skeleton of the projected process) is naturally Markovian (even though the transition time is not necessarily exponential), but the description of the underlying continuous state-space dynamics is very poor (since the information of the last visited milestone is not very informative about the actual state of the system). On the other hand, taking larger milestones, the projected process is close to a Markov process under some metastability assumptions with respect to these milestones. We refer to Sarich, Noé and Schütte (2010), Sarich and Schütte (2013) and Bovier and den Hollander (2015) for a mathematical analysis.

6.4. Accelerated dynamics

The idea of the numerical methods presented in this section is to take advantage of metastability in order to simulate paths over very long time scales

[20] One could think of one-dimensional overdamped Langevin dynamics, with milestones defined as points: in this case the sequence of visited points is Markovian.

more efficiently. In the spirit of a predictor–corrector method, the principle is indeed to use the underlying Markov jump process to more efficiently simulate the original dynamics, or more precisely the state-to-state dynamics which encodes the ambient metastable state at a given time (see equation (6.33)). These techniques, called *accelerated dynamics*, were introduced by Voter and his co-workers in the late 1990s.

Using the notation of the previous section, let us consider a function $\mathcal{S} : \mathcal{D} \to \mathbb{N}$ which defines a partition of the state space. The aim of accelerated dynamics techniques is to efficiently generate paths of the dynamics $(\mathcal{S}(q_t))_{t \geqslant 0}$. As explained above, this is not a Markovian dynamics but, if $\mathcal{S}$ is well chosen, it is close to a Markovian dynamics. The main idea of *accelerated dynamics* is to take advantage of this fact, by checking the Markovianity on-the-fly (see Section 6.4.1 below). However – and this is very important – we show below that for any mapping $\mathcal{S}$, the accelerated dynamics are consistent, in the sense that they simulate the state-to-state dynamics (6.33) up to an error term which goes to zero when certain numerical parameters go to zero. Somehow, the mapping $\mathcal{S}$ is very similar to the reaction coordinate ξ introduced in Section 4.1, except that it takes values in a discrete state space (which can be chosen to be $\mathbb{N}$ without loss of generality: see (6.32)). For further comments on the practical way to choose $\mathcal{S}$, we refer to Remark 6.11 below.

Given the mapping $\mathcal{S}$, the principal idea of accelerated dynamics algorithms is as follows: if the stochastic process remains trapped for a very long time in a given state $S = \mathcal{S}^{-1}(\{n\})$ (for some n), then it has 'lost its memory', and there is a way to generate the exit event from this state much more efficiently than by running the original dynamics until the exit time. This is formalized mathematically below. These algorithms thus efficiently simulate the *exit event* from S, which means that they efficiently sample the pair of random variables (τ_S, q_{τ_S}). Thanks to these techniques, one efficiently samples the state-to-state dynamics $(\mathcal{S}(q_t))_{t \geqslant 0}$. Let us emphasize that the aim is to get the correct *law on the paths* (in order to be able to compute dynamical quantities, such as the typical time to go from one state to another). This is much more demanding than, for example, simply computing the relative likelihood of the states $(\mathcal{S}^{-1}(\{n\}))_{n \in \mathbb{N}}$ (which is equivalent to computing the free energy associated with $\mathcal{S}$: see Section 4.1).

Three algorithms have been proposed by Voter and co-workers. Before going into the details, let us briefly describe the ideas behind these three algorithms.

In the *Parallel Replica* method (Voter 1998), the principle is to simulate in parallel many trajectories following the original dynamics (1.9) or (1.8), to consider the first exit event among the replicas, and to generate from this first exit event a consistent exit time and exit point. The gain is thus obtained in terms of wall clock time. This algorithm can be seen as a way to

parallelize a computation in time, which is not an easy problem in general due to the sequential nature of the time evolution.

In the *Hyperdynamics* method (Voter 1997), the idea is to modify the potential V within the state S in order to accelerate the exit from the state for the original dynamics (1.9) or (1.8). Again, using an appropriate time rescaling, it is possible to generate from the observed exit event on the biased potential an exit event which is consistent with what would have been observed on the original unbiased potential V.

The *Temperature Accelerated Dynamics* (TAD) method (Sorensen and Voter 2000) considers the original dynamics (1.9) or (1.8) at a higher temperature than the original one. The idea is then that, under appropriate assumptions, there is a way to infer from the exit events observed at high temperature the exit event which would have been observed at the original lower temperature.

This section is organized as follows. In Section 6.4.1 we present the decorrelation step: this is used in practice to assess whether the process has remained sufficiently long in a state that one can assume it is distributed according to the QSD. In Sections 6.4.2–6.4.4 we successively introduce the three algorithms: Parallel Replica, Hyperdynamics, and Temperature Accelerated Dynamics. Section 6.4.5 is devoted to some concluding remarks.

6.4.1. The decorrelation step

The accelerated dynamics algorithms are based on the assumption that the process remains sufficiently long in the state S that it can be assumed to be distributed according to the QSD ν_S.

A natural preliminary question is therefore how to assess in practice that the limit has been reached in (6.26). This is done in the so-called decorrelation step, which consists in waiting for a given time $t_{\rm corr}$ (a so-called decorrelation time) before assuming that the local equilibrium ν_S has been reached. This correlation time can be state-dependent, and is typically assumed to be known *a priori*. However, let us mention the recent work of Binder, Simpson and Lelièvre (2015), where a numerical method is proposed in order to approximate the time to reach the QSD using a Fleming–Viot particle process together with convergence diagnostics.

From a mathematical viewpoint, $t_{\rm corr}$ should be chosen sufficiently large that the distance between the law of $q_{t_{\rm corr}}$ conditioned to $\tau_S \geqslant t_{\rm corr}$ and the QSD ν_S is small. Le Bris *et al.* (2012) prove the following result.

Proposition 6.9. Let $(q_t)_{t\geqslant 0}$ satisfy (1.8) with $q_0 \in S$. Introduce the first two eigenvalues $-\lambda_2 < -\lambda_1 < 0$ of the operator $\mathcal{L}^\dagger$ on S with homogeneous Dirichlet boundary conditions on ∂S (see Proposition 6.6 for the definition of $\mathcal{L}^\dagger$). Then there exists a constant $C > 0$ (which depends on the law

of q_0), such that, for all $t \geqslant C/(\lambda_2 - \lambda_1)$,

$$\sup_{f,\, \|f\|_{L^\infty} \leqslant 1} |\mathbb{E}(f(\tau_S - t, q_{\tau_S}) | \tau_S \geqslant t) - \mathbb{E}^{\nu_S}(f(\tau_S, q_{\tau_S}))| \leqslant C \exp(-(\lambda_2 - \lambda_1)t).$$

The supremum is taken over all bounded functions $f : \mathbb{R}_+ \times \partial S \to \mathbb{R}$, with L^∞-norm smaller than one.

In other words, the total variation norm (2.26) between the law of $(\tau_S - t, q_{\tau_S})$ conditioned to $\tau_S \geqslant t$ (for any initial condition $q_0 \in S$), and the law of (τ_S, q_{τ_S}) when q_0 is distributed according to ν_S, decreases exponentially fast with rate $\lambda_2 - \lambda_1$. This means that $t_{\rm corr}$ should be chosen of order $1/(\lambda_2 - \lambda_1)$. Of course, this is not a very practical result since computing these eigenvalues is in general impossible. From a theoretical viewpoint, this result tells us that the local equilibration time is of order $1/(\lambda_2 - \lambda_1)$, so the state S is metastable if this time is much smaller than the exit time (which is typically of order $1/\lambda_1$: see Proposition 6.6).

From now on, we assume that the decorrelation step has been successful (the process remained in S for a time $t_{\rm corr}$), and we look for efficient techniques to generate the exit event (*i.e.* a sample of the random variables (τ_S, q_{τ_S})) starting from the QSD. Let us describe in succession the three algorithms proposed by Voter and his co-workers.

6.4.2. Parallel Replica

Consider an initial condition $q_0 \in S$ such that q_0 is distributed according to the QSD in S. Assume that we are given a computer with many CPUs in parallel. The idea of the Parallel Replica method is to distribute N independent initial conditions $(q_0^i)_{1 \leqslant i \leqslant N}$ in S according to the QSD ν_S, to let them evolve according to (1.8) driven by independent Brownian motions (so that the replicas remain independent), and then to consider the first exit event among the replicas:

$$I_0 = \operatorname*{arg\,min}_{i \in \{1,\ldots,N\}} \tau_S^i \quad \text{where } \tau_S^i = \inf\{t \geqslant 0 \ : \ q_t^i \notin S\}. \tag{6.34}$$

The integer $I_0 \in \{1, \ldots, N\}$ is the index of the first replica to leave S, and $\min(\tau_S^1, \ldots, \tau_S^N) = \tau_S^{I_0}$. The effective exit time is set as N times the first exit time, and the effective exit point is simply the exit point for the first exit event.

The consistency of the method is a corollary of Proposition 6.7. Indeed, using the fact that, starting from the QSD, the exit time is exponentially distributed and independent of the exit point, we easily obtain that

$$N\tau_S^{I_0} = N \min(\tau_S^1, \ldots, \tau_S^N) \overset{\mathcal{L}}{=} \tau_S^1, \tag{6.35}$$

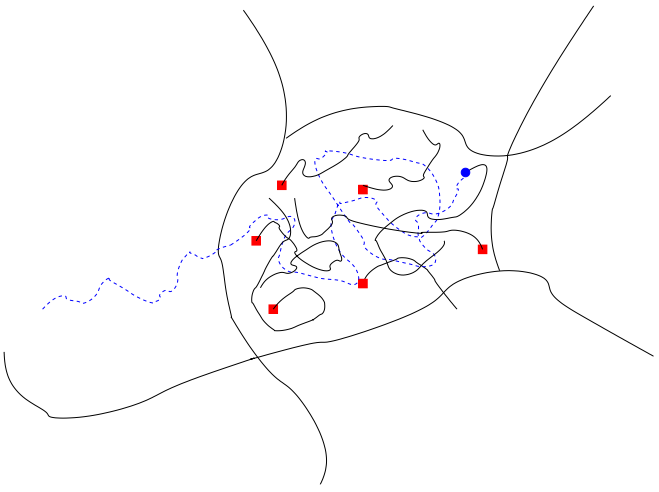


Figure 6.2. *Parallel Replica.* Many exit events are simulated in parallel, all starting from the QSD in S. The dashed line trajectory represents the reference walker, which stays sufficiently long within S that we can assume the end point (circle) is distributed according to the QSD. The other dots (squares) represent i.i.d. initial conditions distributed according to the QSD. The solid line trajectories are simulated in parallel.

which means that the effective exit time has the correct law, and

$$X^{I_0}_{\tau_S^{I_0}} \overset{\mathcal{L}}{=} X^1_{\tau_S^1},$$

which means that the first exit point of the replica I_0 (the first one to exit among N) has the same law as the first exit point of any of them. Moreover

$$N\tau_S^{I_0} = N\min(\tau_S^1, \ldots, \tau_S^N) \quad \text{and} \quad X^{I_0}_{\tau_S^{I_0}}$$

are independent, so we have proved the following lemma.

Lemma 6.10. Let I_0 be the index of the first replica exiting S, defined by (6.34). Then we have the equality in law

$$\left(N\tau_S^{I_0}, X^{I_0}_{\tau_S^{I_0}}\right) \overset{\mathcal{L}}{=} \left(\tau_S^1, X^1_{\tau_S^1}\right).$$

This lemma shows that Parallel Replica is exact: the law of the effective exit time and exit point is exactly the law of the exit time and exit point which would have been observed for only one replica.

Let us make a few remarks on this algorithm. First, the full algorithm actually iterates the following three steps; see Figure 6.2 for a schematic illustration of the method.

- The *decorrelation step* (see Section 6.4.1), where a reference walker is run following the dynamics (1.8) until it remains trapped for such a long time in one of the sets $\mathcal{S}^{-1}(\{n\})$ that it can be assumed to be distributed according to the QSD ν_S associated with this set. During this step, the algorithm thus simply integrates the original dynamics. No error is introduced and there is no computational gain.
- The *dephasing step*, which is a preparation step during which N independent initial conditions distributed according to ν_S are sampled, each one on a different CPU. This is done in parallel. During this step the simulation clock is stopped. This step is thus purely overhead. It requires appropriate algorithms to sample the QSD such as the rejection algorithm or Fleming–Viot particle systems (see Le Bris *et al.* 2012). For example, the rejection algorithm consists in running independent walkers following the dynamics (1.8) (starting from a point within S) and to consider the final point of the trajectory conditional on the walker remaining in the state, for a sufficiently long trajectory (typically for the time $t_{\rm corr}$ introduced in the previous paragraph).
- The *parallel step*, just described above, which consists in running the N replicas independently in parallel, and waiting for the first exit event among the N replicas. The simulation clock is then updated by adding the effective exit time $N\tau_S^{I_0}$. The exit point $X_{\tau_S^{I_0}}^{I_0}$ is used as the initial condition of the reference walker for the next decorrelation step. The computational gain of the whole algorithm comes from this step, which divides the wall clock time to sample the exit event by the number of replicas N.

In practice, if the rejection algorithm is used in the dephasing step, one does not actually need to wait for the N replicas to be dephased to proceed to the parallel step: see Voter (1998), Le Bris *et al.* (2012) and Binder *et al.* (2015).

The Parallel Replica algorithm is very versatile; see Aristoff, Lelièvre and Simpson (2014) for an extension to Markov chains. In particular, it applies to both energetic and entropic barriers. The only errors introduced in the algorithm are related to the rate of convergence of the law of the conditioned process to the QSD. For an analysis of the errors introduced by the algorithm and a discussion of the parallel efficiency, we refer to Le Bris *et al.* (2012) and Binder *et al.* (2015). The algorithm is more efficient when the convergence time to the QSD is small compared to the exit time (*i.e.*, the states are metastable): in this case, the speed-up in terms of wall clock time is linear as a function of N.

Remark 6.11 (on the mapping $\mathcal{S}$). Let us comment on the mapping $\mathcal{S}$ which defines the states. These comments are actually also relevant to

the two other algorithms (Hyperdynamics and TAD). First, from a practical viewpoint, it is important to notice that the numbering of the states need not be defined *a priori*: the states are numbered as the simulation proceeds, when they are successively discovered by the stochastic process.

Now, an important question is how to choose the partition $(\mathcal{S}^{-1}(\{n\}))_{n\in\mathbb{N}}$ of the state space. A useful output of the mathematical analysis is that, thanks to the decorrelation step, the Parallel Replica algorithm is consistent whatever the choice of the partition, in the limit of a large $t_{\rm corr}$. However, the efficiency of the algorithm highly depends on the choice of the partition: the states should be metastable regions, in the sense that the law of the stochastic process typically reaches the QSD (a local equilibrium) before the exit event occurs. In this case, $t_{\rm corr}$ can be chosen small so that the decorrelation and dephasing steps are quickly completed, and most of the computational time is spent in the parallel steps, which yields an important speed-up in terms of wall clock time. The design of a good partition is a difficult question. For a system with high energy barriers (this is often the case for applications in materials science, for example), the original idea of Voter and co-workers is to define the states as the basins of attraction of the local minima of ∇V for the simple gradient dynamics $\dot{x} = -\nabla V(x)$. For a system with more diffusive or entropic barriers (this is typically the case for biological applications), one could think of defining the states using relevant reaction coordinates; see for example Kum *et al.* (2004), where the states are defined in terms of the molecular topology of the molecule of interest. Notice that choosing a good partition also implies being able to estimate the correlation time within each state, either from some *a priori* knowledge or from on-the-fly estimates (Binder *et al.* 2015).

Finally, let us mention that the Parallel Replica method (and the other two algorithms below) can be seen as a way of efficiently simulating the exit event from the state S, once the QSD in S has been reached. From this point of view, these techniques can be applied even if no partition of the state space is available but only an ensemble of disjoint subsets of the configuration space, in the spirit of the milestoning technique mentioned at the end of Section 6.3. In this case, they can be used to efficiently simulate exit events from these states, if the system happens to be trapped in one of them. In between these states, standard molecular dynamics is run. If the union of the disjoint subsets contains most of the metastable states, this could yield a very efficient simulation technique.

6.4.3. Hyperdynamics

Let us again assume that we are given an initial condition $q_0 \in S$ such that q_0 is distributed according to the QSD in S. In other words, assume that we are at the end of the decorrelation step: the reference walker stayed sufficiently long in S.

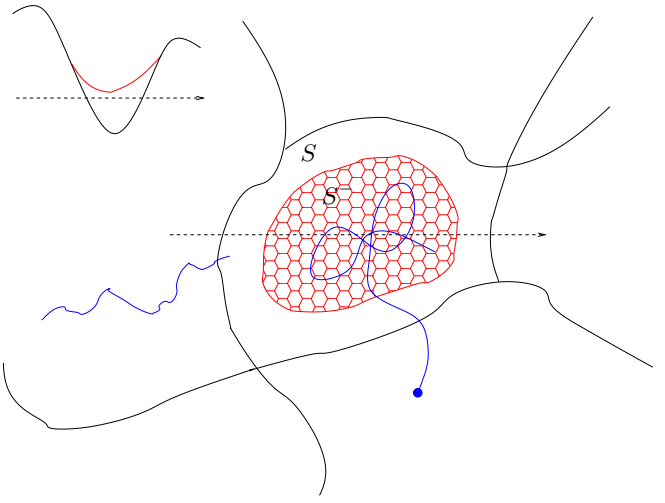


Figure 6.3. *Hyperdynamics.* The exit event is simulated on a biased potential in S. The honeycomb region S^- represents the support of the biasing potential δV. The inset represents the original potential V (black line) and the biased potential $V + \delta V$ along the axis represented by the dashed line.

The principle of the Hyperdynamics algorithm is to raise the potential inside the state in order to accelerate the exit from S. The algorithm thus requires a biasing potential $\delta V : S \to \mathbb{R}$, which satisfies appropriate assumptions detailed below. The algorithm then proceeds as follows.

- Equilibrate the dynamics on the biased potential $V + \delta V$. In other words, run the dynamics (1.8) on the process $(q_t^{\delta V})_{t \geqslant 0}$ over the biased potential conditional on staying in the well, up to the time the random variable $q_t^{\delta V}$ has a distribution close to the QSD $\nu_S^{\delta V}$ associated with the biased potential. This first step is a preparation step, which is purely overhead. The end point $q_t^{\delta V}$ is used as the initial condition for the next step.

- Run the dynamics (1.8) over the biased potential $V + \delta V$ up to the exit time $\tau_S^{\delta V}$ from the state S. The simulation clock is updated by adding the effective exit time $B\tau_S^{\delta V}$, where B is the so-called boost factor defined by

$$B = \frac{1}{\tau_S^{\delta V}} \int_0^{\tau_S^{\delta V}} \exp\bigl[\beta\, \delta V(q_t^{\delta V})\bigr]\, \mathrm{d}t. \tag{6.36}$$

  The exit point is then used as the starting point for a new decorrelation step.

See Figure 6.3 for a schematic illustration of the method.

Roughly speaking, the assumptions required on δV in the original paper of Voter (1997) are twofold.

- δV is sufficiently small that the exit event from the state S still satisfies the standard assumptions used for kinetic Monte Carlo models and transition state theory: the exit time is exponentially distributed and independent of the exit point.

- δV is zero on (a neighbourhood of) the boundary ∂S.

The original derivation of the method relies on explicit formulae for the laws of the exit time and exit point, using the transition state theory: see Voter (1997). The aim of the mathematical analysis presented below is to give a rigorous set of assumptions to make this algorithm consistent.

The algorithm we present here is actually slightly different from the way it is introduced in the original paper by Voter (1997). Indeed, in the original version, the local equilibration steps (decorrelation step and equilibration step on the biased potential) are omitted: it is assumed that the states are sufficiently metastable (for both the original potential and the biased potential), so these local equilibrations are immediate. It would be interesting to check if the modifications proposed here improve the accuracy of the method.

Let us now discuss the mathematical foundations of this technique, and in particular a way to understand the formula (6.36) for the boost factor. We actually need to compare two exit events. The first one is the exit event for the original process $(q_t)_{t\geqslant 0}$ following the dynamics (1.8), starting from the QSD ν_S associated with the state S and the dynamics with potential V. The second is the exit event for the process $(q_t^{\delta V})_{t\geqslant 0}$ following the dynamics (1.8) on the biased potential $V + \delta V$, starting from the QSD $\nu_S^{\delta V}$ associated with the state S and the dynamics with potential $V + \delta V$. Referring to Proposition 6.7, comparing the two exit events amounts to understanding how the first eigenvalue λ_1 and the normal derivative of the first eigenvector $\partial_n u_1$ are modified when changing the potential from V to $V + \delta V$. Let $\lambda_1(V)$ (resp. $\lambda_1(V + \delta V)$) and $\partial_n u_1(V)$ (resp. $\partial_n u_1(V + \delta V)$) denote the first eigenvalue and the normal derivative when considering the original potential V (resp. the biased potential $V + \delta V$). In Lelièvre and Nier (2015), the following result is proved.

Theorem 6.12. We assume there exists an open set S^- such that $\overline{S^-} \subset S$, and the following.

- *Regularity.* V and $V|_{\partial S}$ are Morse functions, which means that the critical points are non-degenerate (the Hessian at any critical point is invertible).

- *Localization in S^- of the eigenvectors associated with small eigenvalues*:
  (i) $|\nabla V| \neq 0$ in $\overline{S} \setminus S^-$,
  (ii) $\partial_n V > 0$ on ∂S^-,
  (iii) $\min_{\partial S} V \geqslant \min_{\partial S^-} V$,
  (iv) $\min_{\partial S^-} V - V^* > V^* - \min_{S^-} V$, where
  $$V^* = \max\{V(x), x \text{ s.t. } |\nabla V(x)| = 0\}.$$
- *Non-degeneracy of exponentially small eigenvalues.* The critical values of V in S^- are all distinct and the differences $V(y) - V(x)$ are all distinct, where x ranges over the set $\mathcal{U}^{(0)}$ of local minima of $V|_{S^-}$ and y ranges over the set $\mathcal{U}^{(1)}$ of critical points of $V|_{S^-}$ with index 1.

Let us also assume that the biasing potential δV is such that (see Figure 6.3 for an illustration of the two domains S^- and S):

- $V + \delta V$ satisfies the same assumptions as those on V above,
- $\delta V = 0$ on $\overline{S} \setminus S^-$.

Then, there exists $c > 0$ such that, in the limit $\beta \to \infty$,

$$\frac{\lambda_1(V + \delta V)}{\lambda_1(V)} = \frac{\int_S e^{-\beta V}}{\int_S e^{-\beta(V+\delta V)}} \big(1 + \mathrm{O}(e^{-\beta c})\big), \tag{6.37}$$

$$\frac{\partial_n[u_1(V + \delta V)]\big|_{\partial S}}{\|\partial_n[u_1(V + \delta V)]\|_{L^1(\partial S)}} = \frac{\partial_n[u_1(V)]\big|_{\partial S}}{\|\partial_n[u_1(V)]\|_{L^1(\partial S)}} + \mathrm{O}(e^{-\beta c}) \quad \text{in } L^1(\partial S). \tag{6.38}$$

The proof is based on results from semi-classical analysis for Witten Laplacians on bounded domains, which is an extension of the semi-classical analysis for Witten Laplacians on domains without boundaries, already discussed in Section 2.5.

Note that the formula (6.37) provides a justification of the formula (6.36) for the boost factor. Indeed, by assuming that $\tau_S^{\delta V}$ is sufficiently large, we have by an ergodic property

$$B = \frac{1}{\tau_S^{\delta V}} \int_0^{\tau_S^{\delta V}} \exp\big[\beta\, \delta V(X_t^{\delta V})\big]\, \mathrm{d}t \simeq \frac{\int_S \exp(\beta \delta V) \exp(-\beta(V + \delta V))}{\int_S \exp(-\beta(V + \delta V))}$$
$$= \frac{\int_S \exp(-\beta V)}{\int_S \exp(-\beta(V + \delta V))} \simeq \frac{\lambda_1(V + \delta V)}{\lambda_1(V)}.$$

By multiplying the exit time on the biased potential $V + \delta V$ by

$$\frac{\lambda_1(V + \delta V)}{\lambda_1(V)},$$

we indeed obtain (in law) the exit time on the original potential V. Moreover, the estimate (6.38) shows that (up to exponentially small errors in the limit of low temperature), the first exit point from S for the biased potential has the same distribution as the first exit point from S for the original potential (see equation (6.30) in Proposition 6.7).

A practical aspect we do not discuss here at all is the effective construction of the biasing potential δV. Voter (1997) proposes a technique based on the Hessian $\nabla^2 V$. A well-known alternative in the context of materials science is the bond–boost method introduced in Miron and Fichthorn (2003).

Notice that, contrary to the Parallel Replica method, Hyperdynamics is, at least for our mathematical analysis, limited to energetic barriers (see assumptions (iii) and (iv) in Theorem 6.12). On the other hand, for very high energetic barriers, Hyperdynamics is in principle much more efficient than Parallel Replica: the Parallel Replica method only divides the exit time by N (the number of replicas), while for deep wells, the boost factor B is very large.

6.4.4. Temperature Accelerated Dynamics

Let us finally introduce the Temperature Accelerated Dynamics (TAD) method proposed by Sorensen and Voter (2000). We again assume we are at the end of the decorrelation step: the reference walker stayed sufficiently long in S. The principle of TAD is to increase the temperature (*i.e.*, to increase β^{-1} in (1.8)) in order to accelerate the exit from S. The algorithm consists in

- simulating many exit events from S at high temperature, starting from the QSD at high temperature,
- extrapolating the high-temperature exit events to low-temperature exit events using the Arrhenius law.

As for the Hyperdynamics algorithm, in the original paper by Sorensen and Voter (2000) no equilibration step is used: it is assumed that, at both high and low temperatures, the states are sufficiently metastable that the convergence to the QSD is immediate. Let us now describe more precisely how the extrapolation procedure is made.

Consider the exit event from S, at a given temperature. Using the notation of Section 6.3, the set S is surrounded by I neighbouring states. Let ∂S_i denote the common boundary with the ith neighbouring state, $i \in \{1, \ldots, I\}$. The sets ∂S_i thus form a partition of the boundary ∂S. Introduce, for $i \in \{1, \ldots, I\}$, the saddle point $x_i \in \partial S_i$, which is the point of lowest energy on ∂S_i: in the low-temperature regime, the paths leaving the state S through ∂S_i leave through a neighbourhood of x_i; this can be inferred from results from large deviation theory, for example those of Freidlin and Wentzell (1984). Let us also denote by x_0 the global minimum

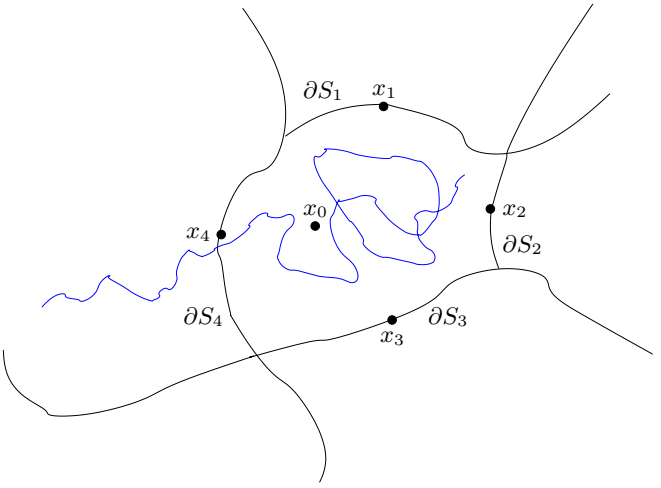


Figure 6.4. *Temperature Accelerated Dynamics.* Exit events are simulated at a higher temperature and then extrapolated to the original smaller temperature.

of V on S; see Figure 6.4 for a schematic representation of the geometry. As explained in Section 6.3, the interesting quantities to define the exit events are:

- the probability of exiting through ∂S_i, given by

$$p(i) = \mathbb{P}(q_{\tau_S} \in \partial S_i) = -\frac{\int_{\partial S_i} \partial_n u_1 \, \mathrm{d}\sigma}{\beta \lambda_1 \int_S u_1(x) \, \mathrm{d}x}$$

  (see Proposition 6.7),

- the parameter of the exponential random variable τ_S, that is,

$$\lambda_1 = 1/\mathbb{E}(\tau_S).$$

As explained in Section 6.3, one way to rewrite the exit event is to attach to each exit region ∂S_i (or to each saddle point x_i) a rate $k(i) = \lambda_1 p(i)$, and to consider I independent exponential random variables τ_i with parameter $k(i)$. The exit event is then given by

(i) the exit time $\min_{i \in \{1,\dots I\}}(\tau_i) \stackrel{\mathcal{L}}{=} \tau_S$, and
(ii) the exit region $\arg\min_{i \in \{1,\dots I\}}(\tau_i)$,

since, for $i \in \{1, \dots, I\}$,

$$\mathbb{P}\left(\underset{j \in \{1,\dots I\}}{\arg\min}(\tau_j) = i \right) = p(i).$$

The TAD algorithm requires as an input an approximate expression of the

rate $k(i)$, namely the Arrhenius law (see (2.83) and recall that the spectrum of $-\mathcal{L}$ is β times the formula given in (2.83))

$$k(i) = \lambda_1 p(i) \simeq \eta_i \exp(-\beta[V(x_i) - V(x_0)]), \tag{6.39}$$

where η_i is a prefactor independent of β. In the original paper by Sorensen and Voter (2000), the Arrhenius law is justified using the (harmonic) transition state theory.

Let us now go back to the TAD algorithm. Using the underlying kinetic Monte Carlo model presented above, and in view of the Arrhenius law (6.39), we obtain

$$\frac{k^{\mathrm{hi}}(i)}{k^{\mathrm{lo}}(i)} = \frac{\lambda_1^{\mathrm{hi}} p^{\mathrm{hi}}(i)}{\lambda_1^{\mathrm{lo}} p^{\mathrm{lo}}(i)} \simeq \exp\bigl(-(\beta^{\mathrm{hi}} - \beta^{\mathrm{lo}})(V(x_i) - V(x_0))\bigr) \tag{6.40}$$

where, with obvious notation, β^{lo} denotes the inverse low temperature, β^{hi} the inverse high temperature, and the superscripts 'lo' and 'hi' refer to the associated quantities at low and high temperature, respectively. The extrapolation formula (6.40) is used in TAD in order to infer the exit event at low temperature from the exit events observed at high temperature, by using the formula

$$(\tau_1^{\mathrm{lo}}, \dots, \tau_I^{\mathrm{lo}}) \overset{\mathcal{L}}{=} (\Theta^1 \tau_1^{\mathrm{hi}}, \dots, \Theta^I \tau_I^{\mathrm{hi}}), \tag{6.41}$$

where

$$\Theta^i = \frac{k^{\mathrm{hi}}(i)}{k^{\mathrm{lo}}(i)} \simeq \exp\bigl(-(\beta^{\mathrm{hi}} - \beta^{\mathrm{lo}})(V(x_i) - V(x_0))\bigr)$$

is a multiplicative factor constructed from the ratio of the rates (6.40). In the equality in law (6.41), the random variables $\tau_i^{\mathrm{hi/lo}}$ are, as described above, exponential random variables with parameter $k^{\mathrm{hi/lo}}(i)$. To have analytical formulae for the correction factors Θ_i and make the algorithm practical, the Arrhenius law is assumed to be exact, and in practice one uses

$$\Theta^i = \exp(-(\beta^{\mathrm{hi}} - \beta^{\mathrm{lo}})(V(x_i) - V(x_0))).$$

The TAD algorithm thus consists in running the dynamics at high temperature, observing the exit events through the saddle points on the boundary of the state,[21] and updating the exit time and exit region that would have been observed at low temperature. More precisely, if exits through

[21] In practice, to associate an exit path with a saddle point, the nudged elastic band method (Jónsson, Mills and Jacobsen 1998) can be used: it consists in computing the closest minimum energy path to the initial exit path using a gradient descent method, the end points of the trajectory being fixed. In addition, note that it is implicitly assumed that the exit point distribution through ∂S_i ($i \in \{1, \dots, I\}$) is the same for β^{hi} and β^{lo}: the exit points at β^{hi} are used to continue the dynamics, once the exit region has been selected.

the saddle points $\{s_1, \dots, s_n\} \subset \{1, \dots, I\}$ have been observed, one computes $\min(\Theta^{s_1}\tau^{\rm hi}_{s_1}, \dots, \Theta^{s_n}\tau^{\rm hi}_{s_n})$ and $\arg\min(\Theta^{s_1}\tau^{\rm hi}_{s_1}, \dots, \Theta^{s_n}\tau^{\rm hi}_{s_n})$ to get the exit time and the exit region at low temperature.

Let us now discuss how to determine the number n of exit events to be taken into account. The interesting feature of TAD compared to a brute force saddle point search is that it is not required to observe exits through all the saddle points in order to obtain a statistically correct exit event. Indeed, a stopping criterion is introduced to stop the calculations at high temperature when the extrapolation procedure will no longer modify the low-temperature exit event (*i.e.*, it will not modify $\min(\Theta^{s_1}\tau^{\rm hi}_{s_1}, \dots, \Theta^{s_n}\tau^{\rm hi}_{s_n})$, where $\{s_1, \dots, s_n\} \subset \{1, \dots, I\}$ are the saddle points discovered up to the time when the stopping criterion is fulfilled). This stopping criterion requires us to provide some *a priori* knowledge, typically a lower bound on the barriers $V(x_i) - V(x_0)$, for $i \in \{1, \dots, I\}$ (there is also a variant using a lower bound on the prefactors η_i in (6.39)). In some sense, TAD can be seen as a clever saddle point search, with a rigorous way to stop the searching procedure.

If the Arrhenius law (6.39) is exactly satisfied, one can check that the TAD algorithm simulates an exit event which is statistically exact. The mathematical question raised by this algorithm is thus to estimate the difference between the ratio of the rates

$$\frac{\lambda_1^{\rm hi} p^{\rm hi}(i)}{\lambda_1^{\rm lo} p^{\rm lo}(i)}$$

and the estimate deduced from the Arrhenius law

$$\exp(-(\beta^{\rm hi} - \beta^{\rm lo})(V(x_i) - V(x_0)))$$

(see the extrapolation formula (6.40) above). In Aristoff and Lelièvre (2014) we consider as a first step the case of a one-dimensional potential, where S is a single well, and we prove that in the limit $\beta^{\rm hi}, \beta^{\rm lo} \to \infty$ with $\beta^{\rm lo}/\beta^{\rm hi}$ fixed,

$$\frac{\lambda^{\rm hi} p_i^{\rm hi}}{\lambda^{\rm lo} p_i^{\rm lo}} = \mathrm{e}^{-(\beta^{\rm hi} - \beta^{\rm lo})(V(x_i) - V(x_0))} \left(1 + \mathrm{O}\left(\frac{1}{\beta^{\rm hi}} - \frac{1}{\beta^{\rm lo}}\right)\right). \tag{6.42}$$

The extension of this result to a high-dimensional setting is a work in progress.

Notice that, compared to Hyperdynamics, TAD is based on an additional assumption, namely the Arrhenius law. We expect that this implies larger errors for TAD than for Hyperdynamics: for Hyperdynamics the error in (6.37)–(6.38) is exponentially small in the limit $\beta \to \infty$, while for TAD, the error scales like $1/\beta$ in (6.42). However, the interest of TAD compared to Hyperdynamics is that it does not require a biasing potential, which may be complicated to build in some situations.

6.4.5. Concluding remarks and discussion

This section has given some elements on the analysis of three algorithms that were proposed by A. F. Voter and his co-workers in order to efficiently generate the state-to-state dynamics associated with a metastable stochastic process. The analysis is based on the notion of quasi-stationary distribution (QSD). As explained above, starting from the QSD within a well, the exit event is Markovian since the exit time is exponentially distributed and independent of the next visiting state. From a theoretical viewpoint, the QSD thus seems to be an interesting way to relate Markovian dynamics in continuous state spaces (such as Langevin dynamics (1.9) or overdamped Langevin dynamics (1.8)) with Markovian dynamics in discrete state spaces (kinetic Monte Carlo models or Markov state models). It also gives a natural definition of a metastable region (see Lelièvre 2013) as a region where the stochastic process reaches local equilibrium, namely the QSD, before exiting. The associated time scales can be used to quantify the metastability of the process (see after Theorem 4.6 for a related discussion of the quantification of metastability using an alternative approach, namely logarithmic Sobolev inequalities).

Going from Parallel Replica to Hyperdynamics to TAD, the assumptions required for the algorithm to consistently sample the state-to-state dynamics are ever more stringent. Indeed, for Parallel Replica, no assumptions are required beyond the fact that a good estimate of $t_{\rm corr}$ is available (to assess the convergence of the QSD). Hyperdynamics requires additional assumptions on the potential: the metastability of the state should come from energetic barriers (at least for our mathematical analysis to apply). Finally, the TAD algorithm also requires the Arrhenius law to be satisfied. Likewise, going from Parallel Replica to Hyperdynamics to TAD, the errors introduced by the algorithm are expected to be ever larger. On the other hand, the expected computational gain with Parallel Replica is typically smaller than for the two other methods. In addition, comparing Hyperdynamics with TAD, the advantage of TAD is that it does not require the construction of a biasing potential, which is a difficult task in general.

We would like to stress that all these algorithms are used in practice with Langevin dynamics (1.9). The mathematical results presented above assume that the dynamics is overdamped Langevin (1.8). Further work is required to extend these results to Langevin dynamics.

6.5. Conclusion: other algorithms and other mathematical frameworks for studying reactive trajectories

In the previous subsections we focused on importance sampling and control techniques to sample reactive paths (see Section 6.2) and on accelerated dynamics techniques to sample exit events from a metastable state (see

Section 6.4). As mentioned in the introduction to Section 6, there are in fact many other relevant methods for studying dynamical properties. For the sake of completeness, let us mention a few of them.

For the analysis and sampling of reactive paths and transitions between two metastable states that are *a priori* fixed, one can mention the following.

- *Local search techniques.* Starting from an initial guess, these methods look for transition paths through an iterative procedure on paths (in the spirit of local optimization methods). For example, let us mention the nudged elastic band (Jónsson *et al.* 1998), the string method (E *et al.* 2002, 2005), and the max flux approach (Zhao, Shen and Skeel 2010). These techniques aim to find one typical representative path (the minimum energy path or minimum free energy path). Let us also mention transition path sampling methods (Dellago, Bolhuis and Chandler 1999, Dellago and Bolhuis 2009), which sample transition paths starting from an initial guess using a Metropolis–Hastings algorithm in path space.

- *Global search, ensemble of trajectories.* Other approaches aim at sampling the ensemble of paths joining two metastable states, without any initial guess (in the spirit of global optimization methods). Examples include splitting techniques over ensembles of paths such as the adaptive multilevel splitting method (Cérou and Guyader 2007, Cérou *et al.* 2011, Bréhier *et al.* 2016), transition interface sampling (van Erp *et al.* 2003, van Erp and Bolhuis 2005), forward flux sampling (Allen *et al.* 2005, 2009), milestoning techniques (Faradjian and Elber 2004, Maragliano, Vanden-Eijnden and Roux 2009, Schütte *et al.* 2011), and the associated transition path theory, which gives a nice mathematical framework (E and Vanden-Eijnden 2004, Vanden-Eijnden *et al.* 2008, Lu and Nolen 2015).

There are also techniques which consist in looking for all the local minima and saddle points of the potential energy function, and then building from that a Markov state model using the Arrhenius law (6.39). Examples of saddle point search techniques are the dimer method (Henkelman and Jónsson 1999, Zhang and Du 2012), activation relaxation techniques (Barkema and Mousseau 1996, Mousseau and Barkema 1998), or gentlest ascent dynamics (Samanta and E 2012). This can be used to explore very complicated landscapes: see Wales (2003).

Acknowledgements

This work is supported by the European Research Council via the European Union's Seventh Framework Programme (FP/2007-2013) and ERC Grant

Agreement number 614492, and by the Agence Nationale de la Recherche under grant ANR-14-CE23-0012 (COSMOS). We have also benefited from the scientific environment of the Laboratoire International Associé linking the Centre National de la Recherche Scientifique and the University of Illinois at Urbana-Champaign. We would like to thank all our co-authors on these subjects, and those colleagues who read a preliminary version of the manuscript: Charles-Edouard Bréhier, Eric Cancès, Giacomo Di Gesu, Max Fathi, Carsten Hartmann, Claude Le Bris, Dorian Le Peutrec, Julien Reygner, Julien Roussel, Mathias Rousset, Marielle Simon, Zofia Trstanova, Jonathan Weare, Marcus Weber and Pierre-André Zitt.

REFERENCES[22]

A. Abdulle, G. Vilmart and K. C. Zygalakis (2015), 'Long time accuracy of Lie–Trotter splitting methods for Langevin dynamics', *SIAM J. Numer. Anal.* **53**, 1–16.

E. Akhmatskaya and S. Reich (2005), The targeted shadowing hybrid Monte Carlo (TSHMC) method. In *New Algorithms for Macromolecular Simulation*, Springer, pp. 141–153.

R. Allen, C. Valeriani and P. ten Wolde (2009), 'Forward flux sampling for rare event simulations', *J. Phys.-Condens. Mat.* **21**, 463102–463122.

R. Allen, P. Warren and P. ten Wolde (2005), 'Sampling rare switching events in biochemical networks', *Phys. Rev. Lett.* **94**, 018104.

H. Alrachid and T. Lelièvre (2015), 'Long-time convergence of an adaptive biasing force method: Variance reduction by Helmholtz projection', *SMAI J. Comput. Math.* **1**, 55–82.

L. Ambrosio, N. Fusco and D. Pallara (2000), *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Science Publications.

S. Andradóttir, D. P. Heyman and T. J. Ott (1993), 'Variance reduction through smoothing and control variates for Markov chain simulations', *ACM Trans. Model. Comput. Simul.* **3**, 167–189.

C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto and G. Scheffer (2000), *Sur les Inégalités de Sobolev Logarithmiques* (in French), Société Mathématique de France.

D. Aristoff and T. Lelièvre (2014), 'Mathematical analysis of temperature accelerated dynamics', *SIAM Multiscale Model. Simul.* **12**, 290–317.

D. Aristoff, T. Lelièvre and G. Simpson (2014), 'The parallel replica method for simulating long trajectories of Markov chains', *App. Math. Res. Express* **2**, 332–352.

A. Arnold, P. Markowich, G. Toscani and A. Unterreiter (2001), 'On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations', *Comm. Partial Diff. Equations* **26**, 43–100.

[22] The URLs cited in this work were correct at the time of going to press, but the publisher and the authors make no undertaking that the citations remain live or are accurate or appropriate.

R. Assaraf and M. Caffarel (1999), 'Zero-variance principle for Monte Carlo algorithms', *Phys. Rev. Lett.* **83**, 4682–4685.

R. Assaraf, B. Jourdain, T. Lelièvre and R. Roux (2015), Computation of sensitivities for the invariant measure of a parameter dependent diffusion. arXiv:1509.01348

D. Bakry and M. Emery (1985), Diffusions hypercontractives. In *Séminaire de Probabilités XIX*, Vol. 1123 of *Lecture Notes in Mathematics*, Springer, pp. 177–206.

D. Bakry, F. Barthe, P. Cattiaux and A. Guillin (2008), 'A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case', *Electron. Comm. Probab.* **13**, 60–66.

D. Bakry, I. Gentil and M. Ledoux (2014), *Analysis and Geometry of Markov Diffusion Operators*, Springer.

R. Balian (2007), *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics*, Vols I–II, Springer.

R. Banisch and C. Hartmann (2016), 'A sparse Markov chain approximation of LQ-type stochastic control problems', *Math. Control Rel. Fields*, to appear.

G. Barkema and N. Mousseau (1996), 'Event-based relaxation of continuous disordered systems', *Phys. Rev. Lett.* **77**, 4358–4361.

M. I. Baskes (1992), 'Modified embedded-atom potentials for cubic materials and impurities', *Phys. Rev. B* **46**, 2727–2742.

M. Benaïm and C.-E. Bréhier (2016), Convergence of adaptive biasing potential methods for diffusions. arXiv:1603.08088

N. Berglund (2013), 'Kramers' law: Validity, derivations and generalisations', *Markov Process. Rel. Fields* **19**, 459–490.

L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio and C. Landim (2015), 'Macroscopic fluctuation theory', *Rev. Mod. Phys.* **87**, 593–636.

R. N. Bhattacharya (1982), 'On the functional central limit theorem and the law of the iterated logarithm for Markov processes', *Z. Wahrsch. Verw. Gebiete* **60**, 185–201.

A. Binder, G. Simpson and T. Lelièvre (2015), 'A generalized parallel replica dynamics', *J. Comput. Phys.* **284**, 595–616.

G. Blower and F. Bolley (2006), 'Concentration of measure on product spaces with applications to Markov processes', *Studia Math.* **175**, 47–72.

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

S. D. Bond and B. J. Leimkuhler (2007), Molecular dynamics and the accuracy of numerically computed averages. In *Acta Numerica*, Vol. 16, Cambridge University Press, pp. 1–65.

F. Bonetto, J. L. Lebowitz and L. Rey-Bellet (2000), Fourier's law: A challenge to theorists. In *Mathematical Physics 2000* (A. Fokas *et al.*, eds), Imperial College Press, pp. 128–150.

N. Bou-Rabee and M. Hairer (2013), 'Nonasymptotic mixing of the MALA algorithm', *IMA J. Numer. Anal.* **33**, 80–110.

N. Bou-Rabee and H. Owhadi (2010), 'Long-run accuracy of variational integrators in the stochastic context', *SIAM J. Numer. Anal.* **48**, 278–297.

N. Bou-Rabee and E. Vanden-Eijnden (2012), 'A patch that imparts unconditional stability to explicit integrators for Langevin-like equations', *J. Comput. Phys.* **231**, 2565–2580.

F. Bouchet and J. Reygner (2015), Generalisation of the Eyring–Kramers transition rate formula to irreversible diffusion processes. arXiv:1507.02104

A. Bovier and F. den Hollander (2015), *Metastability: A Potential Theoretic Approach*, Springer. To appear.

A. Bovier, M. Eckhoff, V. Gayrard and M. Klein (2004*a*), 'Metastability in reversible diffusion processes I: Sharp asymptotics for capacities and exit times', *J. Eur. Math. Soc.* **6**, 399–424.

A. Bovier, V. Gayrard and M. Klein (2004*b*), 'Metastability in reversible diffusion processes II: Precise asymptotics for small eigenvalues', *J. Eur. Math. Soc.* **7**, 69–99.

G. Bowman, V. Pande and F. Noé (2014), *An Introduction to Markov State Models and their Application to Long Timescale Molecular Simulation*, Springer.

C.-E. Bréhier, M. Gazeau, L. Goudenège, T. Lelièvre and M. Rousset (2016), 'Unbiasedness of some generalized Adaptive Multilevel Splitting algorithms', *Ann. Appl. Probab.*, to appear.

A. Brünger, C. B. Brooks and M. Karplus (1984), 'Stochastic boundary conditions for molecular dynamics simulations of ST2 water', *Chem. Phys. Lett.* **105**, 495–500.

N. Buchete and G. Hummer (2008), 'Peptide folding kinetics from replica exchange molecular dynamics', *Phys. Rev. E* **77**, 030902.

J. Bucklew (2004), *Introduction to Rare Event Simulation*, Springer.

D. Burghelea (1997), 'Lectures on Witten–Helffer–Sjöstrand theory', *General Mathematics* **5**, 85–99.

G. Bussi and M. Parrinello (2007), 'Accurate sampling using Langevin dynamics', *Phys. Rev. E* **75**, 056707.

G. Bussi, A. Laio and M. Parrinello (2006), 'Equilibrium free energies from nonequilibrium metadynamics', *Phys. Rev. Lett.* **96**, 090601.

R. Caflisch (1998), Monte Carlo and quasi-Monte Carlo methods. In *Acta Numerica*, Vol. 7, Cambridge University Press, pp. 1–49.

E. Cancès, M. Defranceschi, W. Kutzelnigg, C. Le Bris and Y. Maday (2003), Computational quantum chemistry: A primer. In *Handbook of Numerical Analysis* (P. G. Ciarlet and C. Le Bris, eds), Vol. X (special volume on computational chemistry), Elsevier, pp. 3–270.

E. Carter, G. Ciccotti, J. Hynes and R. Kapral (1989), 'Constrained reaction coordinate dynamics for the simulation of rare events', *Chem. Phys. Lett.* **156**, 472–477.

P. Cattiaux (2014), Long time behavior of Markov processes. In *Journées MAS 2012*, Vol. 44 of *ESAIM Proc.*, EDP Sciences, pp. 110–128.

P. Cattiaux, A. Guillin and L.-M. Wu (2009), 'A note on Talagrand's transportation inequality and logarithmic Sobolev inequality', *Probab. Theory Rel. Fields* **148**, 285–304.

P. Cattiaux, A. Guillin and P.-A. Zitt (2013), 'Poincaré inequalities and hitting times', *Ann. Inst. Henri Poincaré Probab. Statist.* **49**, 95–118.

F. Cérou and A. Guyader (2007), 'Adaptive multilevel splitting for rare event analysis', *Stoch. Anal. Appl.* **25**, 417–443.

F. Cérou, A. Guyader, T. Lelièvre and D. Pommier (2011), 'A multiple replica approach to simulate reactive trajectories', *J. Chem. Phys.* **134**, 054108.

D. Chafaï (2015), 'From Boltzmann to random matrices and beyond', *Ann. Fac. Sci. Toulouse Math.* **6**, 641–689.

R. Chetrite and H. Touchette (2014), 'Nonequilibrium Markov processes conditioned on large deviations', *Ann. Henri Poincaré* **16**, 2005–2057.

C. Chipot and A. Pohorille, eds (2007), *Free Energy Calculations*, Vol. 86 of *Springer Series in Chemical Physics*, Springer.

N. Chopin, T. Lelièvre and G. Stoltz (2012), 'Free energy methods for Bayesian inference: Efficient exploration of univariate Gaussian mixture posteriors', *Statist. Comput.* **22**, 897–916.

G. Ciccotti and G. Jacucci (1975), 'Direct computation of dynamical response by molecular-dynamics: The mobility of a charged Lennard-Jones particle', *Phys. Rev. Lett.* **35**, 789–792.

G. Ciccotti, T. Lelièvre and E. Vanden-Eijnden (2008), 'Projection of diffusions on submanifolds: Application to mean force computation', *Commun. Pure Appl. Math.* **61**, 371–408.

P. Collet, S. Martínez and J. San Martín (2013), *Quasi-Stationary Distributions*, Springer.

H. Cycon, R. Froese, W. Kirsch and B. Simon (1987), *Schrödinger Operators with Application to Quantum Mechanics and Global Geometry*, Springer.

E. Darve and A. Porohille (2001), 'Calculating free energy using average forces', *J. Chem. Phys.* **115**, 9169–9183.

R. Dautray and J.-L. Lions (1990), *Mathematical Analysis and Numerical Methods for Science and Technology*, Vol. 3: *Spectral Theory and Applications*, Springer.

C. Dellago and P. Bolhuis (2009), Transition path sampling and other advanced simulation techniques for rare events. In *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, Vol. 221 of *Advances in Polymer Science*, Springer, pp. 167–233.

C. Dellago, P. Bolhuis and D. Chandler (1999), 'On the calculation of reaction rate constants in the transition path ensemble', *J. Chem. Phys* **110**, 6617–6625.

A. Dembo and O. Zeitouni (1998), *Large Deviations Techniques and Applications*, second edition, Springer.

F. den Hollander (2000), *Large Deviations*, Vol. 14 of *Fields Institute Monographs*, AMS.

W. K. den Otter and W. J. Briels (1998), 'The calculation of free-energy differences by constrained molecular-dynamics simulations', *J. Chem. Phys.* **109**, 4139–4146.

B. Derrida, J. L. Lebowitz and E. R. Speer (2002), 'Large deviation of the density profile in the steady state of the open symmetric simple exclusion process', *J. Statist. Phys.* **107**, 599–634.

J.-D. Deuschel and D. W. Stroock (1989), *Large Deviations*, Vol. 137 of *Pure and Applied Mathematics*, Academic.

A. Dhar (2008), 'Heat transport in low-dimensional systems', *Adv. Phys.* **57**, 457–537.
B. Dickson, F. Legoll, T. Lelièvre, G. Stoltz and P. Fleurat-Lessard (2010), 'Free energy calculations: An efficient adaptive biasing potential method', *J. Phys. Chem. B* **114**, 5823–5830.
M. D. Donsker and S. Varadhan (1975), 'Asymptotic evaluation of certain Markov process expectations for large times I', *Commun. Pure Appl. Math.* **28**, 1–47.
R. Douc, G. Fort and A. Guillin (2009), 'Subgeometric rates of convergence of f-ergodic strong Markov processes', *Stoch. Proc. Appl.* **119**, 897–923.
S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth (1987), 'Hybrid Monte-Carlo', *Phys. Lett. B* **195**, 216–222.
A. B. Duncan, T. Lelièvre and G. A. Pavliotis (2016), 'Variance reduction using nonreversible Langevin samplers', *J. Statist. Phys.* **163**, 457–491.
P. Dupuis and H. Wang (2004), 'Importance sampling, large deviations, and differential games', *Stoch. and Stoch. Reports* **76**, 481–508.
P. Dupuis, K. Spiliopoulos and H. Wang (2012), 'Importance sampling for multiscale diffusions', *SIAM J. Multiscale Model. Simul.* **10**, 1–27.
P. Dupuis, K. Spiliopoulos and X. Zhou (2015), 'Escaping from an attractor: Importance sampling and rest points I', *Ann. Appl. Probab.* **25**, 2909–2958.
W. E and E. Vanden-Eijnden (2004), Metastability, conformation dynamics, and transition pathways in complex systems. In *Multiscale Modelling and Simulation*, Vol. 39 of *Lecture Notes in Computational Science and Engineering*, Springer, pp. 35–68.
W. E, W. Ren and E. Vanden-Eijnden (2002), 'String method for the study of rare events', *Phys. Rev. B* **66**, 052301.
W. E, W. Ren and E. Vanden-Eijnden (2005), 'Finite temperature string method for the study of rare events', *J. Phys. Chem. B* **109**, 6688–6693.
J.-P. Eckmann and M. Hairer (2003), 'Spectral properties of hypoelliptic operators', *Commun. Math. Phys.* **235**, 233–253.
P. Español and P. Warren (1995), 'Statistical-mechanics of dissipative particle dynamics', *Europhys. Lett.* **30**, 191–196.
S. N. Ethier and T. G. Kurtz (1986), *Markov Processes: Probability and Mathematical Statistics*, Wiley.
D. J. Evans and G. P. Morriss (2008), *Statistical Mechanics of Nonequilibrium Liquids*, Cambridge University Press.
L. Evans and R. Gariepy (1992), *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics, CRC.
E. Faou (2012), *Geometric Numerical Integration and Schrödinger Equations*, Zurich Lectures in Advanced Mathematics, European Mathematical Society.
A. Faradjian and R. Elber (2004), 'Computing time scales from reaction coordinates by milestoning', *J. Chem. Phys.* **120**, 10880–10889.
M. Fathi and G. Stoltz (2015), Improving dynamical properties of stabilized discretizations of overdamped Langevin dynamics. arXiv:1505.04905v1
M. Fathi, A.-A. Homman and G. Stoltz (2014), 'Error analysis of the transport properties of Metropolized schemes', *ESAIM Proc.* **48**, 341–363.
A. Fischer, S. Waldhausen, I. Horenko, E. Meerbach and C. Schütte (2007), 'Identification of biomolecular conformations from incomplete torsion angle observations by hidden Markov models', *J. Comp. Chem.* **28**, 2453–2464.

W. Fleming (1977), 'Exit probabilities and optimal stochastic control', *Appl. Math. Optim.* **4**, 329–346.
W. Fleming and H. Soner (2006), *Controlled Markov Processes and Viscosity Solutions*, Springer.
G. Fort, B. Jourdain, E. Kuhn, T. Lelièvre and G. Stoltz (2014), 'Efficiency of the Wang–Landau algorithm: A simple test case', *Appl. Math. Res. Express* **2**, 275–311.
G. Fort, B. Jourdain, E. Kuhn, T. Lelièvre and G. Stoltz (2015*a*), 'Convergence of the Wang–Landau algorithm', *Math. Comp.* **84**, 2297–2327.
G. Fort, B. Jourdain, T. Lelièvre and G. Stoltz (2015*b*), Self-healing umbrella sampling: Convergence and efficiency. Accepted for publication in *Statist. Comput.* arXiv:1410.2109
P. L. Freddolino, C. B. Harrison, Y. Liu and K. Schulten (2010), 'Challenges in protein folding simulations', *Nature Physics* **6**, 751–758.
M. Freidlin and A. Wentzell (1984), *Random Perturbations of Dynamical Systems*, Springer.
A. Friedman (1975), *Stochastic Differential Equations and Applications*, Vol. 1, Academic.
G. Gallavotti and E. G. D. Cohen (1995), 'Dynamical ensembles in stationary states', *J. Statist. Phys.* **80**, 931–970.
C. J. Geyer (1994), Estimating normalizing constants and reweighting mixtures in Markov chain Monte Carlo. Technical report 565, School of Statistics, University of Minnesota.
M. B. Giles (2008), 'Multilevel Monte Carlo path simulation', *Oper. Res.* **56**, 607–617.
M. B. Giles (2015), Multilevel Monte Carlo methods. In *Acta Numerica*, Vol. 24, Cambridge University Press, pp. 259–328.
D. Givon, R. Kupferman and A. Stuart (2004), 'Extracting macroscopic dynamics: Model problems and algorithms', *Nonlinearity* **17**, R55–R127.
J. B. Goodman and K. K. Lin (2009), 'Coupling control variates for Markov chain Monte Carlo', *J. Comput. Phys.* **228**, 7127–7136.
L. Gross (1975), 'Logarithmic Sobolev inequalities', *Amer. J. Math.* **97**, 1061–1083.
N. Grunewald, F. Otto, C. Villani and M. Westdickenberg (2009), 'A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit', *Ann. Inst. H. Poincaré Probab. Statist.* **45**, 302–351.
A. Guionnet and B. Zegarlinski (2003), Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités XXXVI*, Vol. 1801 of *Lecture Notes in Mathematics*, Springer, pp. 1–134.
A. Guyader (2015). Personal communication.
E. Hairer, C. Lubich and G. Wanner (2003), Geometric numerical integration illustrated by the Störmer–Verlet method. In *Acta Numerica*, Vol. 12, Cambridge University Press, pp. 399–450.
E. Hairer, C. Lubich and G. Wanner (2006), *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, Vol. 31 of *Springer Series in Computational Mathematics*, Springer.
M. Hairer and A. J. Majda (2010), 'A simple framework to justify linear response theory', *Nonlinearity* **23** , 909–922.

M. Hairer and J. C. Mattingly (2011), Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, Vol. 63 of *Progr. Probab.*, Birkhäuser/Springer, pp. 109–117.

M. Hairer and G. Pavliotis (2008), 'From ballistic to diffusive behavior in periodic potentials', *J. Statist. Phys.* **131**, 175–202.

C. Hartmann and C. Schütte (2012), 'Efficient rare event simulation by optimal nonequilibrium forcing', *J. Statist. Mech. Theory Exp.* **2012**, P11004.

C. Hartmann, C. Schütte, M. Weber and W. Zhang (2015), Importance sampling in path space for diffusion processes. arXiv:1502.07899

W. K. Hastings (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**, 97–109.

B. Helffer and F. Nier (2005), *Hypoelliptic Estimates and Spectral Theory for Fokker–Planck Operators and Witten Laplacians*, Vol. 1862 of *Lecture Notes in Mathematics*, Springer.

B. Helffer and F. Nier (2006), 'Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach: The case with boundary', *Mémoires de la SMF* **105**.

B. Helffer and J. Sjöstrand (1984), 'Multiple wells in the semiclassical limit I', *Comm. Partial Diff. Equations* **9**, 337–408.

B. Helffer and J. Sjöstrand (1985), 'Puits multiples en mécanique semi-classique IV: Etude du complexe de Witten', *Comm. Partial Diff. Equations* **10**, 245–340.

B. Helffer, M. Klein and F. Nier (2004), 'Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach', *Mat. Contemp.* **26**, 41–85.

S. G. Henderson (1997), Variance reduction via an approximating Markov process. PhD thesis, Department of Operations Research, Stanford University.

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

G. Henkelman and H. Jónsson (1999), 'A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives', *J. Chem. Phys.* **111**, 7010–7022.

F. Hérau and F. Nier (2004), 'Isotropic hypoellipticity and trend to equilibrium for the Fokker–Planck equation with a high-degree potential', *Arch. Rational Mech. Anal.* **171**, 151–218.

F. Hérau, M. Hitrik and J. Sjöstrand (2011), 'Tunnel effect and symmetries for Kramers–Fokker–Planck type operators', *J. Inst. Math. Jussieu* **10**, 567–634.

P. Hohenberg and W. Kohn (1964), 'Inhomogeneous electron gas', *Phys. Rev. B* **136**, 864–871.

R. Holley and D. Stroock (1987), 'Logarithmic Sobolev inequalities and stochastic Ising models', *J. Statist. Phys.* **46**, 1159–1194.

P. J. Hoogerbrugge and J. M. V. A. Koelman (1992), 'Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics', *Europhys. Lett.* **19**, 155–160.

W. G. Hoover (1985), 'Canonical dynamics: Equilibrium phase-space distributions', *Phys. Rev. A* **31**, 1695–1697.

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

L. Hörmander (1985), *The Analysis of Linear Partial Differential Operators I–IV*, Springer.

R. J. Hulse, R. Howley and W. Wilding (2005), 'Transient nonequilibrium molecular dynamic simulations of thermal conductivity I: Simple fluids', *Int. J. Thermophys.* **26**, 1–12.

G. Hummer (2004), 'From transition paths to transition states and rate coefficients', *J. Chem. Phys.* **120**, 516–523.

G. Hummer and A. Szabo (2001), 'Free energy reconstruction from nonequilibrium single-molecule pulling experiments', *Proc. Nat. Acad. Sci.* **98**, 3658–3661.

C.-R. Hwang, S.-Y. Hwang-Ma and S. J. Sheu (1993), 'Accelerating Gaussian diffusions', *Ann. Appl. Probab.* **3**, 897–913.

C.-R. Hwang, S.-Y. Hwang-Ma and S. J. Sheu (2005), 'Accelerating diffusions', *Ann. Appl. Probab.* **15**, 1433–1444.

J. A. Izaguirre and S. S. Hampton (2004), 'Shadow hybrid Monte Carlo: An efficient propagator in phase space of macromolecules', *J. Chem. Phys.* **200**, 581–604.

R. Jack and P. Sollich (2010), 'Large deviations and ensembles of trajectories in stochastic models', *Prog. Theor. Phys. Supp.* **184**, 304–317.

C. Jarzynski (1997), 'Equilibrium free energy differences from nonequilibrium measurements: A master equation approach', *Phys. Rev. E* **56**, 5018–5035.

M. Jerrum, J.-B. Son, P. Tetali and E. Vigoda (2004), 'Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains', *Ann. Appl. Probab.* **14**, 1741–1765.

H. Jónsson, G. Mills and K. Jacobsen (1998), Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations* (B. J. Berne, G. Ciccotti and D. F. Coker, eds), World Scientific, pp. 385–404.

W. L. Jorgensen and J. Tirado-Rives (2005), 'Potential energy functions for atomic-level simulations of water and organic and biomolecular systems', *Proc. Natl Acad. Sci. USA* **102**, 6665–6670.

R. Joubaud, G. Pavliotis and G. Stoltz (2015), 'Langevin dynamics with space-time periodic nonequilibrium forcing', *J. Statist. Phys.* **158**, 1–36.

B. Jourdain, T. Lelièvre and R. Roux (2010), 'Existence, uniqueness and convergence of a particle approximation for the adaptive biasing force process', *ESAIM-Math. Model. Num.* **44**, 831–865.

I. G. Kevrekidis and G. Samaey (2009), 'Equation-free multiscale computation: Algorithms and applications', *Annu. Rev. Phys. Chem.* **60**, 321–344.

R. Khasminskii (2012), *Stochastic Stability of Differential Equations*, second edition, Vol. 66 of *Stochastic Modelling and Applied Probability*, Springer.

C. Kipnis and C. Landim (1999), *Scaling Limits of Interacting Particle Systems*, Springer.

J. Kirkwood (1935), 'Statistical mechanics of fluid mixtures', *J. Chem. Phys.* **3**, 300–313.

W. Kliemann (1987), 'Recurrence and invariant measures for degenerate diffusions', *Ann. Probab.* **15**, 690–707.

P. E. Kloeden and E. Platen (1992), *Numerical Solution of Stochastic Differential Equations*, Vol. 23 of *Applications of Mathematics*, Springer.

W. Kohn and L. J. Sham (1965), 'Self-consistent equations including exchange and correlation effects', *Phys. Rev.* **140**, A1133–A1138.

T. Komorowski, C. Landim and S. Olla (2012), *Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation*, Vol. 345 of *Grundlehren der Mathematischen Wissenschaften*, Springer.

A. Kong, J. S. Liu and W. H. Wong (1994), 'Sequential imputation and Bayesian missing data problems', *J. Amer. Statist. Assoc.* **89**, 278–288.

A. Kong, P. McCullagh, X. L. Meng, D. Nicolae and Z. Tan (2003), 'A theory of statistical models for Monte-Carlo integration', *J. Roy. Statist. Soc. B* **65**, 585–618.

M. Kopec (2015), 'Weak backward error analysis for Langevin process', *BIT* **55**, 1057–1103.

M. Kopec (2015), 'Weak backward error analysis for overdamped Langevin processes', *IMA J. Numer. Anal.* **35**, 583–614.

O. Kum, B. Dickson, S. Stuart, B. Uberuaga and A. Voter (2004), 'Parallel replica dynamics with a heterogeneous distribution of barriers: Application to n-hexadecane pyrolysis', *J. Chem. Phys.* **121**, 9808–9819.

A. Kundu, S. Sabhapandit and A. Dhar (2011), 'Application of importance sampling to the computation of large deviations in nonequilibrium processes', *Phys. Rev. E* **83**, 031119.

J. Kurchan (1998), 'Fluctuation theorem for stochastic dynamics', *J. Phys.* A **31**, 3719–3729.

A. Laio and M. Parrinello (2002), 'Escaping free-energy minima', *Proc. Natl Acad. Sci. USA* **99**, 12562–12566.

D. Lamberton and G. Pagès (2002), 'Recursive computation of the invariant distribution of a diffusion', *Bernoulli* **8**, 367–405.

C. Le Bris (2005), Computational chemistry from the perspective of numerical analysis. In *Acta Numerica*, Vol. 14, Cambridge University Press, pp. 363–444.

C. Le Bris and P.-L. Lions (2008), 'Existence and uniqueness of solutions to Fokker–Planck type equations with irregular coefficients', *Comm. Partial Diff. Equations* **33**, 1272–1317.

C. Le Bris, T. Lelièvre, M. Luskin and D. Perez (2012), 'A mathematical formalization of the parallel replica dynamics', *Monte Carlo Methods Appl.* **18**, 119–146.

D. Le Peutrec, F. Nier and C. Viterbo (2013), 'Precise Arrhenius law for p-forms: The Witten Laplacian and Morse–Barannikov complex', *Ann. Henri Poincaré* **14**, 567–610.

J. L. Lebowitz and H. Spohn (1999), 'A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics', *J. Statist. Phys.* **95**, 333–365.

M. Ledoux (2001), Logarithmic Sobolev inequalities for unbounded spin systems revisited. In *Séminaire de Probabilités XXXV*, Vol. 1755 of *Lecture Notes in Mathematics*, Springer, pp. 167–194.

E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas and K. Schulten (2009), 'Discovery through the computational microscope', *Structure* **17**, 1295–1306.

F. Legoll and T. Lelièvre (2010), 'Effective dynamics using conditional expectations', *Nonlinearity* **23**, 2131–2163.
B. Leimkuhler and C. Matthews (2013*a*), 'Rational construction of stochastic numerical methods for molecular sampling', *Appl. Math. Res. Express* pp. 34–56.
B. Leimkuhler and C. Matthews (2013*b*), 'Robust and efficient configurational molecular sampling via Langevin dynamics', *J. Chem. Phys.* **138**, 174102.
B. Leimkuhler and C. Matthews (2015), *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods*, Springer.
B. Leimkuhler and S. Reich (2004), *Simulating Hamiltonian Dynamics*, Cambridge University Press.
B. Leimkuhler, C. Matthews and G. Stoltz (2015), 'The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics', *IMA J. Numer. Anal.*, in press.
B. Leimkuhler, E. Noorizadeh and F. Theil (2009), 'A gentle stochastic thermostat for molecular dynamics', *J. Statist. Phys.* **135**, 261–277.
T. Lelièvre (2009), 'A general two-scale criteria for logarithmic Sobolev inequalities', *J. Funct. Anal.* **256**, 2211–2221.
T. Lelièvre (2013), Two mathematical tools to analyze metastable stochastic processes. In *Numerical Mathematics and Advanced Applications 2011* (A. Cangiani, R. L. Davidchack, E. Georgoulis, A. N. Gorban, J. Levesley and M. V. Tretyakov, eds), Springer, pp. 791–810.
T. Lelièvre and K. Minoukadeh (2011), 'Long-time convergence of an adaptive biasing force method: The bi-channel case', *Arch. Rational Mech. Anal.* **202**, 1–34.
T. Lelièvre and F. Nier (2015), 'Low temperature asymptotics for quasi-stationary distributions in a bounded domain', *Analysis & PDE*, **8**, 561–628.
T. Lelièvre, F. Nier and G. A. Pavliotis (2013), 'Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion', *J. Statist. Phys.* **152**, 237–274.
T. Lelièvre, M. Rousset and G. Stoltz (2007), 'Computation of free energy profiles with adaptive parallel dynamics', *J. Chem. Phys.* **126**, 134111.
T. Lelièvre, M. Rousset and G. Stoltz (2008), 'Long-time convergence of an adaptive biasing force method', *Nonlinearity* **21**, 1155–1181.
T. Lelièvre, M. Rousset and G. Stoltz (2010), *Free Energy Computations: A Mathematical Perspective*, Imperial College Press.
T. Lelièvre, M. Rousset and G. Stoltz (2012), 'Langevin dynamics with constraints and computation of free energy differences', *Math. Comp.* **81**, 2071–2125.
S. Lepri, R. Livi and A. Politi (2003), 'Thermal conduction in classical low-dimensional lattices', *Phys. Rep.* **377**, 1–80.
E. H. Lieb (1983), 'Density functional theory for Coulomb systems', *Int. J. Quant. Chem.* **24**, 243–277.
F. Lipparini, L. Lagardère, C. Raynaud, B. Stamm, E. Cancès, B. Mennucci, M. Schnieders, P. Ren, Y. Maday and J.-P. Piquemal (2015), 'Polarizable molecular dynamics in a polarizable continuum solvent', *J. Chem. Theory Comput.* **11**, 623–634.
J. Lu and J. Nolen (2015), 'Reactive trajectories and the transition path process', *Probab. Theory Rel. Fields* **161**, 195–244.

M. Luskin and C. Ortner (2013), Atomistic-to-continuum coupling. In *Acta Numerica*, Vol. 22, Cambridge University Press, pp. 397–508.

N. Madras and D. Randall (2002), 'Markov chain decomposition for convergence rate analysis', *Ann. Appl. Probab.* **12**, 581–606.

R. S. Maier and D. L. Stein (1993), 'Escape problem for irreversible systems', *Phys. Rev. E* **48**, 931–938.

R. S. Maier and D. L. Stein (1997), 'Limiting exit location distributions in the stochastic exit problem', *SIAM J. Appl. Math.* **57**, 752–790.

X. Mao (2008), *Stochastic Differential Equations and Applications*, second edition, Horwood, Chichester.

L. Maragliano, E. Vanden-Eijnden and B. Roux (2009), 'Free energy and kinetics of conformational transitions from Voronoi tessellated milestoning with restraining potentials', *J. Chem. Theory Comput.* **5**, 2589–2594.

P. A. Markowich and C. Villani (2000), 'On the trend to equilibrium for the Fokker–Planck equation: An interplay between physics and functional analysis', *Mat. Contemp.* **19**, 1–29.

J. E. Marsden and M. West (2001), Discrete mechanics and variational integrators. In *Acta Numerica*, Vol. 10, Cambridge University Press, pp. 357–514.

S. Marsili, A. Barducci, R. Chelli, P. Procacci and V. Schettino (2006), 'Self-healing Umbrella Sampling: A non-equilibrium approach for quantitative free energy calculations', *J. Phys. Chem. B* **110**, 14011–14013.

B. J. Matkowsky and Z. Schuss (1979), 'The exit problem: A new approach to diffusion across potential barriers', *SIAM J. Appl. Math.* **36**, 604–623.

J. C. Mattingly, A. M. Stuart and D. J. Higham (2002), 'Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise', *Stoch. Proc. Appl.* **101**, 185–232.

S. Melchionna (2007), 'Design of quasisymplectic propagators for Langevin dynamics', *J. Chem. Phys.* **127**, 044108.

X.-L. Meng and W. H. Wong (1996), 'Simulating ratios of normalizing constants via a simple identity: A theoretical exploration', *Statist. Sinica* **6**, 831–860.

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

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller (1953), 'Equations of state calculations by fast computing machines', *J. Chem. Phys.* **21**, 1087–1091.

S. P. Meyn and R. L. Tweedie (2009), *Markov Chains and Stochastic Stability*, second edition, Cambridge University Press.

G. N. Milstein and M. V. Tretyakov (2004), *Stochastic Numerics for Mathematical Physics*, Scientific Computation, Springer.

K. Minoukadeh, C. Chipot and T. Lelièvre (2010), 'Potential of mean force calculations: a multiple-walker adaptive biasing force approach', *J. Chem. Theory Comput.* **6**, 1008–1017.

R. Miron and K. Fichthorn (2003), 'Accelerated molecular dynamics with the bond–boost method', *J. Chem. Phys.* **119**, 6210–6216.

H. Mori (1965), 'Transport, collective motion, and Brownian motion', *Prog. Theor. Phys.* **33**, 423–450.

N. Mousseau and G. Barkema (1998), 'Traveling through potential energy landscapes of disordered materials: The activation–relaxation technique', *Phys. Rev. E* **57**, 2419–2424.

T. Naeh, M. M. Klosek, B. J. Matkowsky and Z. Schuss (1990), 'A direct approach to the exit problem', *SIAM J. Appl. Math.* **50**, 595–627.

F. Nier (2014), Boundary conditions and subelliptic estimates for geometric Kramers–Fokker–Planck operators on manifolds with boundaries. arXiv:1309.5070

S. Nosé (1984*a*), 'A molecular-dynamics method for simulations in the canonical ensemble', *Mol. Phys.* **52**, 255–268.

S. Nosé (1984*b*), 'A unified formulation of the constant temperature molecular-dynamics methods', *J. Chem. Phys.* **81**, 511–519.

E. Nummelin (1984), *General Irreducible Markov Chains and Nonnegative Operators*, Cambridge University Press.

E. Olivieri and M. Vares (2005), *Large Deviations and Metastability*, Cambridge University Press.

S. Olla and V. Letizia (2015), Non-equilibrium isothermal transformations in a temperature gradient from a microscopic dynamics. arXiv:1505.05002v1

F. Otto and M. Reznikoff (2007), 'A new criterion for the logarithmic Sobolev inequality and two applications', *J. Funct. Anal.* **243**, 121–157.

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

E. Pardoux and A. Y. Veretennikov (2005), 'On the Poisson equation and diffusion approximation 3', *Ann. Probab.* **33**, 1111–1133.

S. Piana and A. Laio (2007), 'A bias-exchange approach to protein folding', *J. Phys. Chem. B* **111**, 4553–4559.

E. Platen (1999), An introduction to numerical methods for stochastic differential equations. In *Acta Numerica*, Vol. 8, Cambridge University Press, pp. 197–246.

S. Redon, G. Stoltz and Z. Trstanova (2016), Error analysis of modified Langevin dynamics. arXiv:1601.07411

M. Reed and B. Simon (1975*a*), *Methods of Modern Mathematical Physics*, Vol. I: *Functional Analysis*, Academic.

M. Reed and B. Simon (1975*b*), *Methods of Modern Mathematical Physics*, Vol. IV: *Analysis of Operators*, Academic.

P. Reimann (2002), 'Brownian motors: Noisy transport far from equilibrium', *Phys. Rep.* **61**, 57–265.

D. Revuz and M. Yor (1999), *Continuous Martingales and Brownian Motion*, Grundlehren der Mathematischen Wissenschaften, Springer.

L. Rey-Bellet (2006), Ergodic properties of Markov processes. In *Open Quantum Systems II* (S. Attal, A. Joye and C.-A. Pillet, eds), Vol. 1881 of *Lecture Notes in Mathematics*, Springer, pp. 1–39.

L. Rey-Bellet and K. Spiliopoulos (2015), 'Irreversible Langevin samplers and variance reduction: A large deviations approach', *Nonlinearity* **28**, 2081–2103.

L. Rey-Bellet and K. Spiliopoulos (2016), Improving the convergence of reversible samplers. arXiv:1601.08118

G. O. Roberts and J. S. Rosenthal (1998), 'Optimal scaling of discrete approximations to Langevin diffusions', *J. Roy. Statist. Soc. B* **60**, 255–268.
G. O. Roberts, A. Gelman and W. R. Gilks (1997), 'Weak convergence and optimal scaling of random walk Metropolis algorithms', *Ann. Appl. Probab.* **7**, 110–120.
M. A. Rohrdanz, W. Zheng, M. Maggioni and C. Clementi (2011), 'Determination of reaction coordinates via locally scaled diffusion map', *J. Chem. Phys.* **134**, 124116.
P. J. Rossky, J. D. Doll and H. L. Friedman (1978), 'Brownian dynamics as smart Monte Carlo simulation', *J. Chem. Phys.* **69**, 4628–4633.
G. Royer (2007), *An Initiation to Logarithmic Sobolev Inequalities*, AMS.
W. Rudin (1987), *Real and Complex Analysis*, McGraw-Hill.
M. Rupp (2015), Preface to special issue on machine learning and quantum mechanics, *Int. J. Quantum Chem.* **115**, 1003–1004.
A. Samanta and W. E (2012), 'Atomistic simulations of rare events using gentlest ascent dynamics', *J. Chem. Phys.* **136**, 124104.
M. Sarich and C. Schütte (2013), *Metastability and Markov state Models in Molecular Dynamics*, Vol. 24 of *Courant Lecture Notes*, AMS.
M. Sarich, F. Noé and C. Schütte (2010), 'On the approximation quality of Markov state models', *Multiscale Model. Simul.* **8**, 1154–1177.
T. Schlick (2002), *Molecular Modeling and Simulation*, Springer.
C. Schütte (1998), Conformational dynamics: Modelling, theory, algorithm and application to biomolecules. Habilitation dissertation, Freie Universität Berlin.
C. Schütte, F. Noé, J. Lu, M. Sarich and E. Vanden-Eijnden (2011), 'Markov state models based on milestoning', *J. Chem. Phys.* **134**, 204105.
C. Schütte, S. Winkelmann and C. Hartmann (2012), 'Optimal control of molecular dynamics using Markov state models', *Math. Program.* **134**, 259–282.
T. Shardlow and Y. B. Yan (2006), 'Geometric ergodicity for dissipative particle dynamics', *Stoch. Dynam.* **6**, 123–154.
M. Shirts and J. Chodera (2008), 'Statistically optimal analysis of samples from multiple equilibrium states', *J. Chem. Phys.* **124**, 124105.
D. Siegmund (1976), 'Importance sampling in the Monte Carlo study of sequential tests', *Ann. Statist.* **4**, 673–684.
B. Simon (1984), 'Semiclassical analysis of low lying eigenvalues II: Tunneling', *Ann. of Math.* **120**, 89–118.
R. D. Skeel and J. A. Izaguirre (2002), 'An impulse integrator for Langevin dynamics', *Mol. Phys.* **100**, 3885–3891.
M. Sorensen and A. Voter (2000), 'Temperature-accelerated dynamics for simulation of infrequent events', *J. Chem. Phys.* **112**, 9599–9606.
M. Sprik and G. Ciccoti (1998), 'Free energy from constrained molecular dynamics', *J. Chem. Phys.* **109**, 7737–7744.
D. Talay (2002), 'Stochastic Hamiltonian dissipative systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme', *Markov Proc. Rel. Fields* **8**, 163–198.
D. Talay and L. Tubaro (1990), 'Expansion of the global error for numerical schemes solving stochastic differential equations', *Stoch. Proc. Appl.* **8**, 94–120.

Z. Tan (2004), 'On a likelihood approach for Monte-Carlo integration', *J. Amer. Statist. Assoc.* **99**, 1027–1036.

J. Tersoff (1989), 'Modeling solid-state chemistry: Interatomic potentials for multicomponent systems', *Phys. Rev. B* **39**, 5566–5568.

F. Thalmann and J. Farago (2007), 'Trotter derivation of algorithms for Brownian and dissipative particle dynamics', *J. Chem. Phys.* **127**, 124109.

L. Tierney (1994), 'Markov chains for exploring posterior distributions', *Ann. Statist.* **22**, 1701–1762.

M. Tuckerman (2010), *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press.

S. Vaikuntanathan and C. Jarzynski (2008), 'Escorted free energy simulations: Improving convergence by reducing dissipation', *Phys. Rev. Lett.* **100**, 190601.

A. C. T. van Duin, S. Dasgupta, F. Lorant and W. A. Goddard III (2001), 'ReaxFF: A reactive force field for hydrocarbons', *J. Phys. Chem. A* **105**, 9396–9409.

T. van Erp and P. Bolhuis (2005), 'Elaborating transition interface sampling methods', *J. Comp. Phys.* **205**, 157–181.

T. van Erp, D. Moroni and P. Bolhuis (2003), 'A novel path sampling method for the calculation of rate constants', *J. Chem. Phys.* **118**, 7762–7774.

E. Vanden-Eijnden and J. Weare (2012), 'Rare event simulation of small noise diffusions', *Comm. Pure Appl. Math.* **65**, 1770–1803.

E. Vanden-Eijnden, M. Venturoli, G. Ciccotti and R. Elber (2008), 'On the assumptions underlying milestoning', *J. Chem. Phys.* **129**, 174102.

L. Verlet (1967), 'Computer "experiments" on classical fluids I: Thermodynamical properties of Lennard–Jones molecules', *Phys. Rev.* **159**, 98–103.

C. Villani (2002), A review of mathematical topics in collisional kinetic theory. In *Handbook of Mathematical Fluid Dynamics*, Vol. I, North-Holland, pp. 71–305.

C. Villani (2003), *Topics in Optimal Transportation*, Vol. 58 of *Graduate Studies in Mathematics*, AMS.

C. Villani (2009), *Hypocoercivity*, Vol. 950 of *Memoirs of the American Mathematical Society*, AMS.

A. F. Voter (1997), 'A method for accelerating the molecular dynamics simulation of infrequent events', *J. Chem. Phys.* **106**, 4665–4677.

A. F. Voter (1998), 'Parallel replica method for dynamics of infrequent events', *Phys. Rev. B* **57**, R13 985.

A. F. Voter (2007), Introduction to the Kinetic Monte Carlo method. In *Radiation Effects in Solids* (K. E. Sickafus, E. A. Kotomin and B. P. Uberuaga, eds), Vol. 235 of *NATO Science Series*, Springer, pp. 1–23.

A. F. Voter, F. Montalenti and T. C. Germann (2002), 'Extending the time scale in atomistic simulation of materials', *Ann. Rev. Mater. Res.* **32**, 321–346.

D. Wales (2003), *Energy Landscapes*, Cambridge University Press.

A. Wallqvist and R. D. Mountain (1999), 'Molecular models of water: Derivation and description', *Rev. Comput. Chem.* **13**, 183–247.

F. Wang and D. Landau (2001*a*), 'Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram', *Phys. Rev. E* **64**, 056101.

F. Wang and D. Landau (2001*b*), 'Efficient, multiple-range random walk algorithm to calculate the density of states', *Phys. Rev. Lett.* **86**, 2050–2053.

B. Widom (1963), 'Some topics in the theory of fluids', *J. Chem. Phys* **39**, 2808–2812.

E. Witten (1982), 'Supersymmetry and Morse theory', *J. Differential Geom.* **17**, 661–692.

J. Zhang and Q. Du (2012), 'Shrinking dimer dynamics and its applications to saddle point search', *SIAM J. Numer. Anal.* **50**, 1899–1921.

W. Zhang (2001), *Lectures on Chern–Weil Theory and Witten Deformations*, Vol. 4 of *Nankai Tracts in Mathematics*, World Scientific.

W. Zhang, H. Wang, C. Hartmann, M. Weber and C. Schütte (2014), 'Applications of the cross-entropy method to importance sampling and optimal control of diffusions', *SIAM J. Sci. Comput.* **36**, A2654–A2672.

R. Zhao, J. Shen and R. Skeel (2010), 'Maximum flux transition paths of conformational change', *J. Chem. Theory Comput.* **6**, 2411–2423.

R. Zwanzig (1954), 'High-temperature equation of state by a perturbation method: I. nonpolar gases', *J. Chem. Phys.* **22**, 1420–1426.

R. Zwanzig (1973), 'Nonlinear generalized Langevin equations', *J. Statist. Phys.* **9**, 215–220.