Improving dynamical properties of stabilized discretizations of overdamped Langevin dynamics

M. Fathi[1] and G. Stoltz[2]

[1] LPMA, 4 place Jussieu, 75005 Paris, France

[2] Université Paris-Est, CERMICS (ENPC), INRIA, F-77455 Marne-la-Vallée, France

May 20, 2015

Abstract

The discretization of overdamped Langevin dynamics, through schemes such as the Euler-Maruyama method, may lead to numerical methods which are unstable when the forces are non-globally Lipschitz. One way to stabilize numerical schemes is to superimpose some acceptance/rejection rule, based on a Metropolis-Hastings criterion for instance. However, rejections perturb the dynamical consistency of the resulting numerical method with the reference dynamics. We present in this work some modifications of the standard stabilization of discretizations of overdamped Langevin dynamics by a Metropolis-Hastings procedure, which allow us to either improve the strong order of the numerical method, or to decrease the bias in the estimation of transport coefficients characterizing the effective dynamical behavior of the dynamics. For the latter approach, we rely on modified numerical schemes together with a Barker rule for the stabilization.

1 Introduction

Molecular simulation is nowadays a very common tool to quantitatively predict macroscopic properties of matter starting from a microscopic description. These macroscopic properties can be either static properties (such as the average pressure or energy in a system at fixed temperature and density), or transport properties (such as thermal conductivity or shear viscosity). Molecular simulation can be seen as the computational version of statistical physics, and is therefore often used by practitioners of the field as a black box to extract the desired macroscopic properties from some model of interparticle interactions. Most of the work in the physics and chemistry fields therefore focuses on improving the microscopic description, most notably developing force fields of increasing complexity. In comparison, less attention has been paid to the estimation of errors in the quantities actually computed by numerical simulation. Usually, due to the very high dimensionality of the systems under consideration, macroscopic properties are computed as ergodic averages over a very long trajectory of the system, evolved under some appropriate dynamics. There are two main types of errors in this approach: (i) statistical errors arising from incomplete sampling, and (ii) systematic errors (bias) arising from the fact that continuous dynamics are numerically integrated using a finite time-step $\Delta t > 0$.

The aim of this work is to reduce the bias arising from the use of finite time steps in the computation of dynamical quantities, for a certain type of dynamics called Brownian dynamics or overdamped Langevin dynamics in the chemistry literature. We denote by $\mathcal{M}$ the configurational space, which is in this work a compact domain with periodic boundary conditions such as $\mathbb{T}^d$ (with $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ the one-dimensional torus). Unbounded configuration spaces $\mathbb{R}^d$ can also be considered under some assumptions (see Remark 4). The overdamped Langevin dynamics is a stochastic differential equation on the configuration $q \in \mathcal{M}$ of the system:

$$dq_t = -\beta \nabla V(q_t)\, dt + \sqrt{2}\, dW_t, \tag{1}$$

where $\beta = 1/(k_{\mathrm{B}}T)$ is the inverse temperature (k_{B} being Boltzmann's constant and T being the temperature) and W_t is a standard d-dimensional Brownian motion. The function $V : \mathcal{M} \to \mathbb{R}$ is the potential

1

energy, assumed to be smooth for the mathematical analysis. The generator associated with (1) acts on smooth functions φ as

$$\mathcal{L}\varphi = -\beta \nabla V^T \nabla \varphi + \Delta \varphi. \tag{2}$$

The probability measure

$$\mu(dq) = Z^{-1} \, \mathrm{e}^{-\beta V(q)} \, dq, \qquad Z = \int_{\mathcal{M}} \mathrm{e}^{-\beta V} \tag{3}$$

is invariant by the dynamics (1) since a simple computation shows that, for appropriate functions φ (*e.g.* smooth and such that $\mathcal{L}\varphi$ is integrable with respect to μ),

$$\int_{\mathcal{M}} \mathcal{L}\varphi \, d\mu = 0.$$

Simple discretizations of (1) may fail to be ergodic when the dynamics is considered on unbounded spaces and the potential energy function is not globally Lipschitz [14]. In simulations of ionic solutions, potential energy functions with Coulomb-type singularities are used and it has been observed that the energy blows up (see for instance [8, Section 3.2]).

In order to stabilize numerical discretizations or simply to remove the bias in the invariant measure arising from the use of finite timesteps, it was suggested to consider the move obtained by the numerical scheme as a proposal to be accepted or rejected according to a Metropolis-Hastings ratio [15, 11]. The corresponding method is known as "Smart MC" in the chemistry literature [19], and was rediscovered later on in the computational statistics literature [18] where it is known as "Metropolis Adjusted Langevin Algorithm" (MALA). The interest of the acceptance/rejection step is that it ensures that the Markov chain samples the Gibbs measure μ, which prevents blow-ups. On the other hand, the use of an acceptance/rejection procedure limits the possible numerical schemes to be used as a proposal. Indeed, when generating a proposed move, the probability of coming back to the original configuration from the proposed move has to be computed, by evaluating the probability to observe the Gaussian increment necessary to perform the backward move. Therefore, it is unclear that proposals which are nonlinear in the Gaussian increment (such as the ones produced, say, by the Milstein's scheme, see [16] for instance) can be used, except in specific cases.

The previous works on the numerical analysis of dynamical properties of MALA established (i) strong error estimates over finite times [5], and, as a consequence, errors on finite time correlations [6]; (ii) exponential convergence rates towards the invariant measure, uniformly in the timestep [4] (which holds up to a small error term in Δt for systems in infinite volume); (iii) error estimates on the effective diffusion of the dynamics [8].

The aim of this work is to present new proposal functions, and also to advocate the use of acceptance/rejection rules different than the Metropolis one, in order to reduce the systematic bias in the estimation of dynamical quantities. More precisely, we propose in Section 2 a modified proposal which allows to obtain a strong error of order 1 rather than 3/4 for MALA (see Theorem 3). We also show in Section 3 that the error on the computed self-diffusion can be reduced from Δt for MALA to Δt^2 provided a modified proposal is considered together with a Barker rule [2] (see Theorems 6 and 8). As discussed in Section 3.5, the use of a Barker rule also reduces the bias on the self-diffusion for dynamics with multiplicative noise (*i.e.* a non-trivial diffusion matrix in front the Brownian motion in (1)). On the other hand, resorting to a Barker rule increases the statistical error in the simulation, roughly by a factor 2 since the rejection rate is 1/2 in the limit $\Delta t \to 0$ (as made precise in Remark 7). Such an increase in the variance was already shown in its full generality by Peskun in [17]. However, the reduction in the bias more than compensates the increase in the statistical error if one is interested in simulations with a given error tolerance. We provide numerical illustrations of our predictions at the end of Sections 2 and 3. The proofs of all our results are gathered in Section 4. As mentioned earlier, for simplicity we shall work under the following assumption:

Assumption 1. *The state spaces $\mathcal{M}$ is compact and has periodic boundary conditions, and the potential V is smooth.*

Compact spaces with periodic boundary conditions anyway is the relevant setting for self-diffusion since there is no effective diffusion for dynamics in unbounded spaces with confining potentials.

2 Improving the strong order

In this section, we show how a suitable modification of the proposal leads to a Metropolized scheme with a better strong accuracy at equilibrium that the standard MALA scheme. As was realized in [5], the local error over one timestep arises at dominant order from the rejections and not from the integration errors of the accepted move. The strategy to improve the strong error is therefore to add terms of lower order (in the timestep) to the proposal, which will not have a significant impact when the proposal is accepted, but which will significantly reduce the average rejection rate. As made precise in Lemma 1, the rejection rate scales as $\Delta t^{5/2}$ for the modified dynamics, instead of $\Delta t^{3/2}$ for MALA. Let us also mention that, instead of considering the modified dynamics we propose (see (7) below), we could in fact consider dynamics with an arbitrarily low rejection rate, see Remark 5.

We start by recalling the known error estimates for MALA in Section 2.1, before presenting the error bounds obtained for our modified dynamics in Section 2.2. Our theoretical predictions are illustrated by numerical simulations in Section 2.3.

2.1 Strong error estimates for the standard MALA algorithm

Let us first recall the definition of the MALA scheme. It is a Metropolis-Hastings algorithm whose proposal function is obtained by a Euler-Maruyama discretization of the dynamics (1). Given a timestep $\Delta t > 0$ and a previous configuration q^n, the proposed move is

$$\widetilde{q}^{n+1} = q^n - \beta \Delta t \, \nabla V(q^n) + \sqrt{2\Delta t} \, G^n,$$

where (here and elsewhere) $(G^n)_{n \geqslant 0}$ is a sequence of independent and identically distributed (i.i.d.) d-dimensional standard Gaussian random variables. For further purposes, it will be convenient to encode proposals using a function $\Phi_{\Delta t}(q, G)$ depending on the previous position q and the Gaussian increment G used to generate the proposed move. For the above Euler-Maruyama proposal,

$$\Phi_{\Delta t}^{\mathrm{EM}}(q, G) = q - \beta \Delta t \, \nabla V(q) + \sqrt{2\Delta t} \, G. \tag{4}$$

We next accept or reject the proposed move according to the Metropolis-Hastings ratio $A_{\Delta t}\left(q^n, \widetilde{q}^{n+1}\right)$:

$$A_{\Delta t}\left(q^n, \widetilde{q}^{n+1}\right) = \min\left(\frac{\mathrm{e}^{-\beta V(\widetilde{q}^{n+1})} T_{\Delta t}(\widetilde{q}^{n+1}, q^n)}{\mathrm{e}^{-\beta V(q^n)} T_{\Delta t}(q^n, \widetilde{q}^{n+1})}, 1\right), \tag{5}$$

where

$$T_{\Delta t}(q, q') = \left(\frac{1}{4\pi\Delta t}\right)^{d/2} \exp\left(-\frac{|q' - q + \beta\Delta t \, \nabla V(q)|^2}{4\Delta t}\right)$$

is the probability transition of the Markov chain encoded by (4). When the proposition is accepted, which is decided with probability $A_{\Delta t}\left(q^n, \widetilde{q}^{n+1}\right)$, we project $\widetilde{q}^{n+1}$ into the periodic simulation cell $\mathcal{M}$. If the proposal is rejected, the previous configuration is counted twice: $q^{n+1} = q^n$ (It is very important to count rejected configuration as many times as needed to ensure that the Boltzmann-Gibbs measure μ is invariant). In conclusion,

$$q^{n+1} = \Psi_{\Delta t}(q^n, G^n, U^n) = q^n + \mathbf{1}_{U^n \leqslant A_{\Delta t}\left(q^n, \Phi_{\Delta t}^{\mathrm{EM}}(q^n, G^n)\right)}\left(\Phi_{\Delta t}^{\mathrm{EM}}\left(q^n, G^n\right) - q^n\right), \tag{6}$$

where $(U^n)_{n \geqslant 0}$ is a sequence of i.i.d. random variables with i.i.d. components uniformly distributed in $[0, 1]$, and $\mathbf{1}_{U^n \leqslant \alpha}$ is an indicator function whose value is 1 when $U^n \leqslant \alpha$ and 0 otherwise.

It is easy to see that the Markov chain is irreducible with respect to the Lebesgue measure and that μ is an invariant probability measure. It is therefore ergodic, and in fact reversible with respect to μ (see for instance the references in [13, Section 2.2]). In particular, this guarantees the scheme does not blow up.

The following result on the strong convergence of MALA on finite time intervals has been obtained in [5, Theorem 2.1]. We state the results for dynamics in compact spaces although it has been obtained for dynamics in unbounded spaces, under some additional assumptions on the potential energy function. This allows us to write strong error estimates which are uniform with respect to the initial condition q_0, rather than average errors for initial conditions distributed according to the canonical measure μ (a restriction arising from the lack of geometric ergodicity for MALA in the case of unbounded spaces, see [18]).

Theorem 1 (Strong convergence of MALA [5]). *Consider the Markov chain*

$$q^{n+1} = \Psi_{\Delta t}(q^n, W_{(n+1)\Delta t} - W_{n\Delta t}, U^n)$$

started from $q^0 = q_0$, and denote by $q_t^{\Delta t}$ the piecewise constant function defined as $q_t^{\Delta t} = q^n$ for $t \in [n\Delta t, (n+1)\Delta t)$. Then, there exists $\Delta t^* > 0$ and, for any $T > 0$, a constant $C(T) > 0$ such that, for any $0 < \Delta t \leqslant \Delta t^*$ and all $t \in [0, T]$, and for all $q_0 \in \mathcal{M}$,

$$\left(\mathbb{E}_{q_0} \left[\left| q_t^{\Delta t} - q_t \right|^2 \right] \right)^{1/2} \leqslant C(T)\, \Delta t^{3/4},$$

where the expectation is over all realizations of the Brownian motion $(W_t)_{0 \leqslant t \leqslant T}$.

Numerical simulations confirm that the exponent 3/4 is sharp (see [5, Section 3.1] and Section 2.3 below). Let us emphasize that, in order to have correct strong error estimates, the Gaussian increments in the numerical scheme have to be consistent with the realization of the Brownian motion used to construct the solution of the continuous stochastic dynamics.

2.2 Strong error estimates for a modified dynamics

The un-Metropolized Euler scheme has strong order 1 for dynamics with additive noise such as (1). In order to improve the convergence rate $\Delta t^{3/4}$ given by Theorem 1, we modify (4) with terms of order $\Delta t^{3/2}$ and Δt^2:

$$\widetilde{q}^{n+1} = q^n + \Delta t \Big(- \beta \nabla V(q^n) + \Delta t\, F(q^n) \Big) + \sqrt{2\Delta t} \Big(\mathrm{Id} + \Delta t\, \sigma(q^n) \Big)^{-1/2} G^n. \tag{7}$$

This can be encoded by the proposal function

$$\Phi_{\Delta t}^{\mathrm{mod}}(q, G) = \Phi_{\Delta t}^{\mathrm{EM}}(q, G) + \Delta t^2 F(q) + \sqrt{2\Delta t} \left[(\mathrm{Id} + \Delta t\, \sigma(q))^{-1/2} - \mathrm{Id} \right] G.$$

The above definitions are formal since the matrix $\mathrm{Id} + \Delta t\, \sigma(q^n)$ should be symmetric definite positive in order for the inverse square root to make sense. As made clear below, this is indeed the case for sufficiently small timesteps since the matrix valued function σ is smooth hence bounded on $\mathcal{M}$ (see Remark 4 for unbounded spaces).

Remark 2 (Alternative expressions of the diffusion matrix). *It is possible to consider other diffusion matrices which agree with $(\mathrm{Id} + \Delta t\, \sigma(q))^{-1/2}$ at dominant order in Δt, such as $\mathrm{Id} - \Delta t\, \sigma(q)/2$ which need not be inverted but may be negative for large values of Δt. In [7], the authors suggest to consider $\exp(-\Delta t\, \sigma(q)/2)$, which is always non-negative but may be cumbersome to compute in practice.*

The Metropolis-Hastings ratio associated with the proposal (7) reads

$$A_{\Delta t}^{\mathrm{mod}}\left(q^n, \widetilde{q}^{n+1}\right) = \min \left(\frac{\mathrm{e}^{-\beta V(\widetilde{q}^{n+1})} T_{\Delta t}^{\mathrm{mod}}(\widetilde{q}^{n+1}, q^n)}{\mathrm{e}^{-\beta V(q^n)} T_{\Delta t}^{\mathrm{mod}}(q^n, \widetilde{q}^{n+1})}, 1 \right), \tag{8}$$

with a transition rate taking into account the spatial dependence of the diffusion:

$$T_{\Delta t}^{\mathrm{mod}}(q, q') := \left(\frac{1}{4\pi\Delta t} \right)^{d/2} \det \left(\mathrm{Id} + \Delta t\, \sigma(q) \right)^{1/2}$$

$$\times \exp \left(- \frac{\Big(q' - q + \Delta t\, [\beta \nabla V(q) - \Delta t F(q)] \Big)^T \Big(\mathrm{Id} + \Delta t\, \sigma(q) \Big) \Big(q' - q + \Delta t\, [\beta \nabla V(q) - \Delta t F(q)] \Big)}{4\Delta t} \right). \tag{9}$$

We denote by $\Psi_{\Delta t}^{\mathrm{mod}}(q, G, U)$ the random variable obtained by a single step of the modified MALA scheme (7) starting from q:

$$q^{n+1} = \Psi_{\Delta t}^{\mathrm{mod}}(q^n, G^n, U^n),$$

with

$$\Psi_{\Delta t}^{\mathrm{mod}}(q, G, U) = q + \mathbf{1}_{U \leqslant A_{\Delta t}^{\mathrm{mod}}\left(q, \Phi_{\Delta t}^{\mathrm{mod}}(q, G)\right)} \left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G) - q \right). \tag{10}$$

We can then state the following strong error estimate (see Section 4.2 for the proof).

4

Theorem 3 (Improved strong error estimates). *For a smooth potential V on the compact space $\mathcal{M}$, choose σ and F as follows:*

$$\sigma = \frac{\beta}{3}\nabla^2 V, \qquad F = \frac{\beta}{6}\Big(\nabla(\Delta V) - \beta\nabla^2 V \nabla V\Big). \tag{11}$$

Consider, for $\Delta t < 1/\|\sigma\|_{L^\infty}$, the Markov chain associated with the modified proposal (7) corrected by a Metropolis rule: $q^{n+1} = \Psi_{\Delta t}^{\mathrm{mod}}(q^n, W_{(n+1)\Delta t} - W_{n\Delta t}, U^n)$, and denote by $q_t^{\Delta t,\mathrm{mod}}$ the piecewise constant function defined as $q_t^{\Delta t,\mathrm{mod}} = q^n$ for $t \in [n\Delta t, (n+1)\Delta t)$. Then, there exists $\Delta t^* > 0$ and, for any $T > 0$, a constant $C(T) > 0$ such that, for any $0 < \Delta t \leqslant \Delta t^*$ and $t \in [0, T]$, and for all $q_0 \in \mathcal{M}$,

$$\left(\mathbb{E}_{q_0}\left[\left|q_t^{\Delta t,\mathrm{mod}} - q_t\right|^2\right]\right)^{1/2} \leqslant C(T)\Delta t.$$

The crucial estimate to prove Theorem 3 is the following bound on the rejection rate, which makes precise the fact that the rejection rate scales as $\Delta t^{5/2}$ rather than $\Delta t^{3/2}$ as for MALA (see Section 4.1 for the proof). As the proof shows, there is some algebraic miracle since the corrections terms, chosen to eliminate the dominant contribution of order $\Delta t^{3/2}$ in the rejection rate, in fact also eliminate the next order contribution of order Δt^2.

Lemma 1. *For any $\ell \geqslant 1$, there exists a constant $C_\ell > 0$ such that*

$$\forall q \in \mathbb{R}^d, \qquad \mathbb{E}_G\left(\left[1 - A_{\Delta t}^{\mathrm{mod}}\left(q, \Phi_{\Delta t}^{\mathrm{mod}}(q, G)\right)\right]^{2\ell}\right) \leqslant C_\ell \Delta t^{5\ell}.$$

Let us end this section with the following considerations on the extension of the improved strong error estimates to unbounded spaces.

Remark 4 (Generalization to unbounded spaces). *It is possible to extend the results of Theorem 3 to unbounded spaces under appropriate assumptions on the potential, such as [5, Assumption 4.1] and [4, Assumptions 2.1]. These assumptions ensure that the potential energy function grows sufficiently fast at infinity, with derivatives bounded by V (up to order 5 in the present case), and a lower bound on the Hessian. In this setting, error estimates can be stated for initial conditions $q_0 \sim \nu$, where ν is such that $\mathbb{E}_\nu(|V(q^k)|^p) \leqslant R < +\infty$ for some integer p sufficiently large. A convenient choice is $\nu = \mu$, in which case $q^k \sim \mu$, and the finiteness of the moments of V is guaranteed under mild conditions on V. More generally, error estimates for generic initial conditions can be stated when the dynamics is geometrically ergodic. See also Remark 10 for technical precisions on the extension of Lemma 1.*

2.3   Numerical results

We illustrate the error bounds predicted by Theorems 1 and 3, and also check the scaling of the rejection rate obtained in Lemma 1. We consider a simple one-dimensional system with the potential energy function $V(q) = q^4$ already used as a test case in [5]. This example is particularly relevant since it can be shown that in absence of Metropolization the associated Euler-Maruyama scheme is transient [14].

We compute a reference trajectory over a given time interval $[0, T]$ with a very small timestep Δt_{ref}. We next compute trajectories for larger timesteps $\Delta t_k = k\Delta t_{\mathrm{ref}}$, using Brownian increments consistent with the ones used to generate the trajectory with the smallest timestep. More precisely, denoting by $(G_{\mathrm{ref}}^n)_{n=0,\dots,T/\Delta t_{\mathrm{ref}}-1}$ the Gaussian random variables used for the reference trajectory, the trajectories $(q_k^m)_{m\geqslant 0}$ with coarser timesteps $\Delta t_k = k\Delta t_{\mathrm{ref}}$ are computed with the iterative rule for $0 \leqslant m \leqslant T/\Delta t_k$ (assuming that $T/\Delta t_k$ is an integer):

$$q_k^{m+1} = \Psi_{\Delta t_k}(q_k^m, G_k^m, U_k^m),$$

with the Gaussian random variables

$$G_k^m = \frac{1}{k}\sum_{n=mk}^{(m+1)k-1} G_{\mathrm{ref}}^n,$$

but where the uniform random variables $(U_k^m)_{k,m}$ are i.i.d. and independent of the variables $(U_{\mathrm{ref}}^n)_{n\geqslant 0}$ used to generated the reference trajectory. We denote by d_k the maximal error between the reference
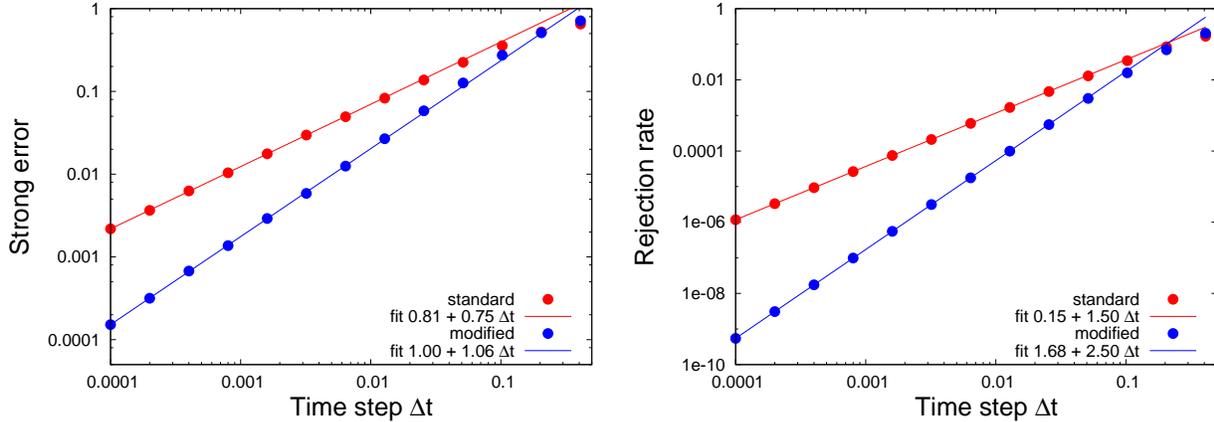
Figure 1: Left: Strong error over the trajectory as a function of the timestep. Right: Average rejection rate as a function of the timestep. The results obtained with MALA correspond to 'standard' while the ones obtained with the scheme (10) correspond to 'modified'.

trajectory and the trajectory generated with the timestep Δt_k, considered at the same times $m\Delta t_k = mk\Delta t_{\mathrm{ref}}$:

$$d_k = \max_{m=1,\ldots,T/\Delta t_k} \left| q_k^m - q_{\mathrm{ref}}^{mk} \right|.$$

We next perform I independent realizations of this procedure, henceforth generating trajectories $(q_k^{i,m})_{m \geqslant 0}$ and $(q_{\mathrm{ref}}^{i,m})_{m \geqslant 0}$. Denoting by d_k^i the so-obtained errors, the strong error for the timestep Δt_k is estimated by the empirical average

$$\widehat{E}_{k,I} = \sqrt{\frac{1}{I} \sum_{i=1}^{I} (d_k^i)^2}$$

In fact, confidence intervals on this error can be obtained thanks to a Central Limit theorem, which shows that

$$\widehat{E}_K \simeq \sqrt{\mathbb{E}(d^2)} \left(1 + \frac{\sigma_d}{2\mathbb{E}(d^2)\sqrt{K}} \mathcal{G} \right), \qquad \sigma_d^2 = \mathbb{E}(d^4) - \left(\mathbb{E}(d^2) \right)^2.$$

Estimates of the average rejection rates are obtained with the empirical average of $1 - A_{\Delta t}(q^n, \widetilde{q}^{n+1})$ computed along the generated trajectories.

Figure 1 presents the estimates $\widehat{E}_{k,I}$ of the strong error and the scaling of the rejection rate as a function of the time step Δt_k. The reference timestep is $\Delta t_{\mathrm{ref}} = 10^{-6}$, and we chose $k = 100 \times 2^l$ where $l = 0, \ldots, L$ with $L = 12$, as well as an integration time $T = 0.8192$ (which corresponds to $2\Delta t_L$). We fixed $\beta = 1$ and averaged over $I = 10^6$ realizations for the standard proposal (4) and $I = 5 \times 10^5$ realizations for the modified proposal (7). The first initial condition is $q_{\mathrm{ref}}^{0,0} = 0$, while the subsequent initial conditions are the end point of the previous trajectory: $q_{\mathrm{ref}}^{i+1,0} = q_{\mathrm{ref}}^{i,T/\Delta t_{\mathrm{ref}}}$. The maximal value of $\sigma_d/(2\mathbb{E}(d^2))$ was estimated to be less than 32 for all reported simulations, so that the maximal relative statistical error is lower than 0.05 in all cases, and often much smaller.

As predicted by Theorems 1 and 3, we find that the strong error decreases as $\Delta t^{3/4}$ for MALA (with a rejection rate scaling as $\Delta t^{3/2}$, as predicted by [5, Lemma 4.7]), while it scales as Δt for the modified dynamics (10). A key element to the improvement in the strong error is the fastest decrease of the rejection rate, indeed confirmed to be of order $\Delta t^{5/2}$.

3 Reducing the bias in the computation of transport coefficients

Our aim in this section is to modify the standard MALA algorithm in order to obtain better approximations of transport coefficients such as the self-diffusion. There are two complementary ways to do so: consider better proposal functions, and replace the Metropolis-Hastings rule with the Barker rule [2].

We start by reviewing in Section 3.1 the definitions of the self-diffusion for the continuous dynamics (1), and recall the known error estimates of the approximations of these quantities based on MALA in Section 3.2. We next propose in Section 3.3 modified schemes to compute more accurately transport coefficients, and state the related error estimates in Section 3.4. As we show in Section 3.5, the use of a Barker rule also allows to reduce errors for dynamics with multiplicative noise. Finally, we illustrate the theoretical results of this section by numerical experiments in Section 3.6.

3.1 Definition of the self-diffusion

We briefly recall the setting already presented in [8, Section 1.2]. Since the positions q_t are restricted to the periodic domain $\mathcal{M}$, they are uniformly bounded in time. To obtain a diffusive behavior from the evolution of q_t, we consider the following additive functional defined on the whole space $\mathbb{R}^d$: starting from $Q_0 = q_0$,

$$Q_t = Q_0 - \beta \int_0^t \nabla V(q_s)\, ds + \sqrt{2}\, W_t. \tag{12}$$

The difference with q_t is that Q_t is not reprojected in $\mathcal{M}$ by the periodization procedure (By this, we mean that we do not choose among all the images of Q_t by translations of the lattice $\mathbb{Z}^d$ the one for which all components are in the interval $[0, 1)$). The diffusion tensor is then given by the following limit (which exists in our context thanks to the existence of a spectral gap for the generator $\mathcal{L}$, see for instance [8, Theorem 1]):

$$\mathscr{D} = \lim_{t \to +\infty} \mathbb{E}\left(\frac{Q_t - Q_0}{\sqrt{t}} \otimes \frac{Q_t - Q_0}{\sqrt{t}} \right), \tag{13}$$

where the expectation is over all realizations of the continuous dynamics (1), starting from initial conditions distributed according to the Boltzmann-Gibbs measure μ defined in (3). In fact, as made precise in [8, Section 1.2], the self-diffusion constant $\mathcal{D}$, defined as the average mean-square displacement of the individual particles, has two equivalent expressions:

$$\mathcal{D} = \frac{1}{2d}\mathrm{Tr}(\mathscr{D}) = \lim_{t \to +\infty} \mathbb{E}\left(\frac{1}{2dt} \sum_{i=1}^d (Q_{i,t} - Q_{i,0})^2 \right) \tag{14}$$

$$= 1 - \frac{\beta^2}{d} \int_0^{+\infty} \mathbb{E}\left[\nabla V(q_t)^T \nabla V(q_0) \right] dt. \tag{15}$$

The expression (14) is called the *Einstein formula*. The second expression (15) involves an integrated autocorrelation function. In accordance with the standard physics and chemistry nomenclature, we call (15) the *Green-Kubo formula* for the self-diffusion in the sequel.

3.2 Error estimates for MALA

For the MALA scheme described in Section 2.1, the self-diffusion can be estimated by either discretizing (14) or (15). In both cases, the error is of order Δt. For the Green-Kubo approach, we proved in [8, Section 2.1] (by an adaption of the results obtained in [12]) that the integrated autocorrelation of smooth observables with average 0 with respect to μ can be estimated up to error terms of order Δt. More precisely, introducing

$$\widetilde{C}^\infty(\mathcal{M}) = \left\{ \psi \in C^\infty(\mathcal{M}) \,\middle|\, \int_\mathcal{M} \psi\, d\mu = 0 \right\},$$

it is proved in [8, Theorem 2] that, for any $\psi, \varphi \in \widetilde{C}^\infty(\mathcal{M})$, there exists $\Delta t^* > 0$ such that, for any $0 < \Delta t \leqslant \Delta t^*$,

$$\int_0^{+\infty} \mathbb{E}\left[\psi(q_t)\, \varphi(q_0) \right] dt = \Delta t \sum_{n=0}^{+\infty} \mathbb{E}_{\Delta t}\left[\psi(q^n)\, \varphi(q^0) \right] + \Delta t\, r_{\psi,\varphi,\Delta t}, \tag{16}$$

7

with $r_{\psi,\varphi,\Delta t}$ uniformly bounded (with respect to Δt), and where the expectation on the left hand side of the above equation is with respect to initial conditions $q_0 \sim \mu$ and over all realizations of the dynamics (1), while the expectation on the right hand side is with respect to initial conditions $q^0 \sim \mu$ and over all realizations of the MALA scheme. As a corollary,

$$\mathcal{D}_{\Delta t}^{\mathrm{GK}} = \mathcal{D} + \Delta t\, \widetilde{\mathcal{D}}_{\Delta t}, \tag{17}$$

where $\widetilde{\mathcal{D}}$ is uniformly bounded for Δt sufficiently small, and where the numerically computed self-diffusion reads

$$\mathcal{D}_{\Delta t}^{\mathrm{GK}} = 1 - \frac{\beta^2 \Delta t}{d} \sum_{n=0}^{+\infty} \mathbb{E}_{\Delta t}\left[\nabla V(q^n)^T \nabla V(q^0)\right]. \tag{18}$$

For the approach based on Einstein's formula (14), we introduce, in accordance with the definition (12), a discrete additive functional allowing to keep track of the diffuse behavior of the Markov chain: Starting from $Q^0 = q^0$,

$$Q^n = \sum_{k=0}^{n-1} \delta_{\Delta t}\left(q^k, G^k, U^k\right), \tag{19}$$

with

$$\delta_{\Delta t}\left(q^k, G^k, U^k\right) = \mathbf{1}_{U^k \leqslant A_{\Delta t}\left(q^k, \Phi_{\Delta t}(q^k, G^k)\right)}\left(\Phi_{\Delta t}\left(q^k, G^k\right) - q^n\right). \tag{20}$$

While the Markov chain $(q^n)_{n \geqslant 0}$ remains in $\mathcal{M}$, the additive functional $(Q^n)_{n \geqslant 0}$ has values in $\mathbb{R}^d$. The diffusion tensor actually computed by the numerical scheme is

$$\mathscr{D}_{\Delta t} = \lim_{n \to +\infty} \mathbb{E}_{\Delta t}\left[\frac{Q^n - Q^0}{\sqrt{n\Delta t}} \otimes \frac{Q^n - Q^0}{\sqrt{n\Delta t}}\right], \tag{21}$$

where the expectation on the right hand side is with respect to initial conditions $Q^0 = q^0 \sim \mu$ and for all realizations of the numerical scheme. It is shown in [8, Theorem 3] that there exists $\Delta t^* > 0$ such that, for any $0 < \Delta t \leqslant \Delta t^*$,

$$\mathscr{D}_{\Delta t} = \mathscr{D} + \Delta t\, \widetilde{\mathscr{D}}_{\Delta t},$$

where the coefficients of the symmetric matrix $\widetilde{\mathscr{D}}_{\Delta t} \in \mathbb{R}^{dN \times dN}$ are uniformly bounded, at least for Δt small enough. As an immediate corollary,

$$\mathcal{D}_{\Delta t}^{\mathrm{Einstein}} = \mathcal{D} + \Delta t\, \widetilde{\mathcal{D}}_{\Delta t}^{\mathrm{Einstein}}, \qquad \mathcal{D}_{\Delta t}^{\mathrm{Einstein}} = \frac{1}{2dN}\mathrm{Tr}\left(\mathscr{D}_{\Delta t}\right), \tag{22}$$

where $\widetilde{\mathcal{D}}_{\Delta t}^{\mathrm{Einstein}}$ is uniformly bounded for Δt sufficiently small.

3.3   Presentation of the modified schemes

The modified scheme presented in Section 2.2 has a very low rejection rate, but it can be checked that the choice (11) leads to a numerical scheme whose weak order is still 1, as for the standard Euler-Maruyama scheme and MALA. Therefore, the error in the computation of the transport coefficients for the scheme of Section 2.2 is still of order Δt. It is not sufficient to aim for lower rejection rates to improve transport properties.

We present in this section a way to decrease the errors in the transport coefficients, even if the rejection rate is still of order $\Delta t^{3/2}$ – and maybe even more surprisingly if the rejection rate is close to $1/2$! The first idea is to introduce two proposals $\widetilde{q}^{n+1} = \Phi_{\Delta t}(q^n, G^n)$ which are alternatives to the standard Euler-Maruyama proposal (4): an implicit midpoint

$$\widetilde{q}^{n+1} = q^n - \beta \Delta t\, \nabla V\left(\frac{\widetilde{q}^{n+1} + q^n}{2}\right) + \sqrt{2\Delta t}\, G^n, \tag{23}$$

and

$$\widetilde{q}^{n+1} = q^n - \beta \Delta t\, \nabla V\left(q^n + \frac{\sqrt{2\Delta t}}{2}\, G^n\right) + \sqrt{2\Delta t}\, G^n. \tag{24}$$

The latter proposal has already been used in [3]. Since the first proposal uses an implicit scheme, the timestep has to be sufficiently small for $\Phi_{\Delta t}$ to be well defined (see Lemma 2 below). For the second

8

proposal, it seems that computing the probability of reverse moves from $\widetilde{q}^{n+1}$ to q^n is going to be difficult in view of the explicit nonlinear dependence on the Gaussian increment G^n. It turns out however that (24) can be reformulated as one step of a reversible, symplectic scheme starting from a random momentum $p^n = \beta^{-1/2} G^n$ and with a timestep $h = \sqrt{2\beta\Delta t}$:

$$
\begin{cases}
q^{n+1/2} = q^n + \dfrac{h}{2}\, p^n, \\[2mm]
p^{n+1} = p^n - h\, \nabla V\left(q^{n+1/2}\right), \\[2mm]
\widetilde{q}^{n+1} = q^{n+1/2} + \dfrac{h}{2}\, p^{n+1}.
\end{cases}
\tag{25}
$$

As made precise below in (28), this allows to compute ratios of transition probabilities relying on Hybrid Monte Carlo algorithms, see for instance the presentation in [13, Section 2.1.3]. For practical purposes, (24) should be preferred since there is no extra computational cost contrarily to the implicit proposal (23), which has to be solved using fixed point iterations or Newton iterations.

Remark 5 (Higher order HMC schemes). *It is possible in principle to integrate the Hamiltonian dynamics using a reversible scheme of arbitrary accuracy, relying on appropriate composition methods (see [9, Sections II.4 and V.3]). This seemingly leads to proposals with an arbitrarily low rejection rate. In practice however, the efficiency of these schemes is limited by the numerical instabilities related to round off errors. We therefore focus on simple composition methods of order 2 such as* (25).

The second idea to improve the computation of transport coefficients is to consider acceptance/rejection rules different from the Metropolis procedure. In fact, we consider the Barker rule [2]. More precisely, the acceptance rates for the Metropolis and Barker rules respectively read

$$
A_{\Delta t}^{\mathrm{MH}}(q^n, \widetilde{q}^{n+1}) = \min\left(1, \mathrm{e}^{-\alpha_{\Delta t}(q^n, \widetilde{q}^{n+1})}\right), \qquad
A_{\Delta t}^{\mathrm{Barker}}(q^n, \widetilde{q}^{n+1}) = \frac{\mathrm{e}^{-\alpha_{\Delta t}(q^n, \widetilde{q}^{n+1})}}{1 + \mathrm{e}^{-\alpha_{\Delta t}(q^n, \widetilde{q}^{n+1})}},
\tag{26}
$$

with, for the midpoint proposal (23),

$$
\begin{aligned}
\alpha_{\Delta t}(q, q') &= \beta\left(V(q') - V(q)\right) + \frac{1}{4\Delta t}\left(\left[q - q' + \beta\Delta t\, \nabla V\left(\frac{q+q'}{2}\right)\right]^2 - \left[q' - q + \beta\Delta t\, \nabla V\left(\frac{q+q'}{2}\right)\right]^2\right) \\
&= \beta\left(V(q') - V(q) - \nabla V\left(\frac{q+q'}{2}\right)\cdot(q' - q)\right),
\end{aligned}
\tag{27}
$$

while, for (24), we define the ratio only in the case $q' = \Phi_{\Delta t}(q, G)$:

$$
\alpha_{\Delta t}\left(q, \Phi_{\Delta t}(q, G)\right) = \beta\left(V(\Phi_{\Delta t}(q, G)) - V(q)\right) + \frac{1}{2}\left[\left(G - \sqrt{2\Delta t}\,\beta\nabla V\left(q + \frac{\sqrt{2\Delta t}}{2}\, G\right)\right)^2 - G^2\right].
\tag{28}
$$

With the notation used in (25), the right-hand side of the above equation can be rewritten as $\beta\left[H(\widetilde{q}^{n+1}, p^{n+1}) - H(q^n, p^n)\right]$, where $H(q, p) = V(q) + p^2/2$ is the Hamiltonian of the system, and $p^n = \beta^{-1/2} G^n$. The new configuration is in all cases

$$
q^{n+1} = \Psi_{\Delta t}(q^n, G^n, U^n) = q^n + \mathbf{1}_{U^n \leqslant A_{\Delta t}(q^n, \Phi_{\Delta t}(q^n, G^n))}\left(\Phi_{\Delta t}\left(q^n, G^n\right) - q^n\right),
$$

where $A_{\Delta t}$ is one of the rates defined in (26). Note that, for the Barker rule, the average rejection rate is close to $1/2$ in the limit $\Delta t \to 0$ (as made precise in (55)). Let us also emphasize that the use of a Barker rule leads to a Markov chain which is reversible with respect to μ (see for instance [13, Section 2.1.2.2]).

Let us now discuss the ergodic properties of the schemes based on the proposals (23) and (24). We introduce the evolution operator

$$
P_{\Delta t}\psi(q) = \mathbb{E}\left(\psi\left(q^{n+1}\right)\,\middle|\, q^n = q\right),
$$

as well as the discrete generator

$$
\begin{aligned}
\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\psi(q) &= \frac{1}{\Delta t}\left[\mathbb{E}\left(\psi\left(q^{n+1}\right)\,\middle|\, q^n = q\right) - \psi(q^n)\right] \\
&= \int_{\mathbb{R}^d} A_{\Delta t}(q, \Phi_{\Delta t}(q, g))\frac{\psi(\Phi_{\Delta t}(q, g)) - \psi(q)}{\Delta t}\,\frac{\mathrm{e}^{-g^2/2}}{(2\pi)^{d/2}}\, dg,
\end{aligned}
\tag{29}
$$

9

where the expectation in the first line is over G^n, U^n. We also define the space of bounded functions with average 0 with respect to μ:

$$\widetilde{L}^\infty(\mathcal{M}) = \left\{ \varphi \in L^\infty(\mathcal{M}) \,\middle|\, \int_\mathcal{M} f \, d\mu = 0 \right\}.$$

We can then state the following lemma (see Section 4.4 for the proof).

Lemma 2. *There exists $\Delta t^* > 0$ such that, for any $0 < \Delta t \leqslant \Delta t^*$, the proposal* (23) *is well defined whatever $q^n \in \mathcal{M}$ and $G^n \in \mathbb{R}^d$. In addition, there exist $C, \lambda > 0$ such that, for all $n \in \mathbb{N}$ and $0 < \Delta t \leqslant \Delta t^*$, and any $f \in \widetilde{L}^\infty(\mathcal{M})$, the evolutions generated by either the proposal* (23) *or* (24)*, together with a Metropolis-Hastings rule or a Barker rule, satisfy*

$$\|P_{\Delta t}^n f\|_{L^\infty} \leqslant C \, e^{-\lambda n \Delta t} \|f\|_{L^\infty}. \tag{30}$$

As a consequence, there exists $K > 0$ such that, uniformly in $\Delta t \in (0, \Delta t^*]$,

$$\left\| \left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t} \right)^{-1} \right\|_{\mathcal{B}(\widetilde{L}^\infty)} \leqslant K. \tag{31}$$

Such ergodicity results are already provided in [4, 8] for proposals based on a simple Euler-Maruyama discretization and a Metropolis-Hastings rule. The resolvent bound (31) is a crucial ingredient to appropriately estimate errors in transport coefficients.

3.4 Error estimates on the self-diffusion for the modified schemes

We are now able to state error estimates concerning the numerical approximation of Green-Kubo formulas.

Theorem 6 (Improved Green-Kubo formula). *Set $a = 1/2$ and $\alpha = 2$ when the Barker rule is used with either the midpoint proposal* (23) *or the HMC proposal* (24)*, and $a = 1$ and $\alpha = 3/2$ when the Metropolis-Hastings rule is used for these proposals. Consider two observables $\psi, \varphi \in \widetilde{C}^\infty(\mathcal{M})$. Then, there exists $\Delta t^* > 0$ such that, for any $0 < \Delta t \leqslant \Delta t^*$,*

$$\int_0^{+\infty} \mathbb{E}\Big[\psi(q_t)\,\varphi(q_0)\Big] dt = \Delta t \left(a \sum_{n=0}^{+\infty} \mathbb{E}_{\Delta t} \left[\psi(q^n)\,\varphi(q^0) \right] - \frac{1}{2} \int_\mathcal{M} \psi\varphi \, d\mu \right) + \Delta t^\alpha r_{\psi,\varphi,\Delta t},$$

with $r_{\psi,\varphi,\Delta t}$ uniformly bounded (with respect to Δt), and where the expectation on the left hand side of the above equation is with respect to initial conditions $q_0 \sim \mu$ and over all realizations of the dynamics (1)*, while the expectation on the right hand side is with respect to initial conditions $q^0 \sim \mu$ and over all realizations of the numerical scheme.*

Note that, for the Barker rule, the quadrature formula corresponds to the standard way of computing Green-Kubo formulas (as on the right-hand side of (16)) except that the first term $n = 0$ is discarded and a global factor $1/2$ is required in order to correct for the fact that the average rejection rate of the Barker algorithm is close to $1/2$.

Remark 7 (Increase in the asymptotic variance for the Barker rule). *The asymptotic variance of the time average of a function of interest f for a Markov chain $(q^n)_{n\geqslant 0}$ with invariant measure μ is (assuming without loss of generality that the average of f with respect to μ vanishes)*

$$\sigma^2(f) = \int_\mathcal{M} f^2 \, d\mu + 2 \sum_{n=1}^{+\infty} \mathbb{E}\Big(f(q^n) f(q^0) \Big),$$

where the expectation is with respect to initial conditions $q^0 \sim \mu$ and over all realizations of the Markov chain; see for instance [13, Section 2.3.1.2] and references therein. In view of the estimates provided by Theorem 6, the asymptotic variance of time averages computed along trajectories of schemes based on the Metropolis rule are, for Δt sufficiently small,

$$\Delta t \, \sigma^2_{\mathrm{MH},\Delta t}(f) = \int_0^{+\infty} \mathbb{E}\Big[f(q_t)\,f(q_0) \Big] dt + \mathrm{O}(\Delta t),$$

10

while, for the Barker rule,

$$\Delta t \, \sigma_{\mathrm{Barker}, \Delta t}^2(f) = 2 \int_0^{+\infty} \mathbb{E}\Big[f(q_t) \, f(q_0) \Big] dt + \mathrm{O}(\Delta t).$$

This shows that $\Delta t \, \sigma_{\mathrm{Barker}, \Delta t}^2(f) = 2 \Delta t \, \sigma_{\mathrm{MH}, \Delta t}^2(f) + \mathrm{O}(\Delta t)$, which allows to quantify the increase in the variance as $\Delta t \to 0$ due to the increased rejection rate of the Barker rule.

As a corollary of Theorem 6, we obtain error bounds on the computation of the self-diffusion as

$$\mathcal{D}_{\Delta t}^{\mathrm{GK}} = \mathcal{D} + \Delta t^\alpha \, \widetilde{\mathcal{D}}_{\Delta t}^{\mathrm{GK}}, \tag{32}$$

where $\widetilde{\mathcal{D}}_{\Delta t}^{\mathrm{GK}}$ is uniformly bounded for Δt sufficiently small, and where the numerically computed self-diffusion reads, for the Metropolis rule,

$$\mathcal{D}_{\Delta t}^{\mathrm{GK}} = 1 - \frac{\beta^2 \Delta t}{d} \left(\frac{1}{2} \int_{\mathcal{M}} |\nabla V|^2 \, d\mu + \sum_{n=1}^{+\infty} \mathbb{E}_{\Delta t} \left[\nabla V(q^n)^T \nabla V(q^0) \right] \right).$$

while, for the Barker rule,

$$\mathcal{D}_{\Delta t}^{\mathrm{GK}} = 1 - \frac{\beta^2 \Delta t}{2d} \sum_{n=1}^{+\infty} \mathbb{E}_{\Delta t} \left[\nabla V(q^n)^T \nabla V(q^0) \right]. \tag{33}$$

Compared to the results of [8] (recalled in (17)), (32) allows to compute the diffusion coefficient up to error terms of order Δt^2 rather than Δt provided the Barker rule is used.

We next present error estimates on the numerical approximation of Einstein's formula. We still define Q^n by (19), but replace the proposal function $\Phi_{\Delta t}$ by the ones associated with the schemes presented in Section 3.3, and consider the appropriate acceptance rates (26).

Theorem 8 (Improved Einstein fluctuation formula). *Set $a = 1/2$ and $\alpha = 2$ when the Barker rule is used with either the midpoint or the HMC proposal, and $a = 1$ and $\alpha = 3/2$ when the Metropolis-Hastings rule is used for these proposals. Consider the unperiodized displacement Q^n defined (19), with $Q^0 = q^0 \sim \mu$. Then,*

$$\mathscr{D}_{\Delta t} = \lim_{n \to +\infty} \mathbb{E}\left[\frac{Q^n - Q^0}{\sqrt{n \Delta t}} \otimes \frac{Q^n - Q^0}{\sqrt{n \Delta t}} \right] = a \mathscr{D} + \Delta t^\alpha \, \widetilde{\mathscr{D}}_{\Delta t},$$

where the expectation on the left hand side of the above equation is with respect to initial conditions $q^0 \sim \mu$ and over all realizations of the numerical scheme, and where $\widetilde{\mathscr{D}}_{\Delta t}^{\mathrm{Einstein}}$ has uniformly bounded coefficients.

It is possible to generalize such fluctuation formulas to more general increments $f(q, \delta_{\Delta t}(q, G, U))$ instead of $\delta_{\Delta t}(q, G, U)$, but the corresponding formulas are quite complicated and we therefore refrain from stating them. An immediate corollary of Theorem 8 is the following error estimate on the self-diffusion:

$$\mathcal{D}_{\Delta t}^{\mathrm{Einstein}} = \mathcal{D} + \Delta t^\alpha \, \widetilde{\mathcal{D}}_{\Delta t}^{\mathrm{Einstein}}, \qquad \mathcal{D}_{\Delta t}^{\mathrm{Einstein}} = \frac{1}{2ad} \mathrm{Tr}\left(\mathscr{D}_{\Delta t} \right), \tag{34}$$

where we recall $a = 1$ and $\alpha = 3/2$ for the Metropolis-Hastings rule, while $a = 1/2$ and and $\alpha = 2$ for the Barker rule. The error estimates provided by this formula are consistent with the ones provided by the Green-Kubo approach (see (32)).

The proofs of Theorems 6 and 8, which can be read in Sections 4.5 and 4.6 respectively, crucially rely on the following weak-type expansion of the discrete generator associated with the proposals (23) or (24).

Lemma 3. *For any $\psi \in C^\infty(\mathcal{M})$,*

$$\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t} \psi = a \left(\mathcal{L}\psi + \frac{\Delta t}{2} \mathcal{L}^2 \psi \right) + \Delta t^\alpha \, r_{\psi, \Delta t}, \tag{35}$$

where $r_{\psi, \Delta t}$ is uniformly bounded in $L^\infty(\mathcal{M})$ for Δt sufficiently small; and $\alpha = 2$, $a = 1/2$ for schemes using the Barker rule, while $\alpha = 3/2$, $a = 1$ for Metropolis-based methods.

The proof of this lemma is provided in Section 4.3. Let us emphasize that the proposals corrected by the Barker rule do not lead to evolution operators $P_{\Delta t}$ which have weak second order, and cannot even be written as $e^{\Delta t \mathcal{L}/2}$ up to corrections of order Δt^3. This is the reason why we call (35) a "second order type weak expansion", by which we mean that the Δt^2 term in the expansion of $P_{\Delta t}$ is proportional to $\mathcal{L}^2$. This is in fact all we need to perform the proofs of Theorems 6 and 8.

Remark 9. *As made precise in Section 4.3, it is possible to replace the drift $-\nabla V$ in the proposals with more general drifts, namely $-\nabla V + \Delta t\, \widetilde{F}$ for the midpoint proposal (23), or $-\nabla V + \Delta t\, \nabla \widetilde{V}$ for the HMC proposal (24). The choice $\widetilde{F} = -\nabla(\Delta V)/4$ ensures that the underlying scheme is of weak order 2 (see [1, Theorem 3.2]). However, even when $\widetilde{F} = 0$, in which case the underlying scheme is only of weak order 1, the average drift introduced by the acceptance/rejection procedures automatically corrects the drift and ensures that the expansion (35) holds whatever the expression of $\widetilde{F}$.*

3.5 Extension to diffusions with multiplicative noise

3.5.1 Description of the dynamics

In this section, we extend the results obtained for the dynamics (1) with additive noise to dynamics with multiplicative noise. More precisely, we consider a position-dependent diffusion matrix $M \in \mathbb{R}^{d \times d}$, assumed to be a smooth function of the positions $q \in \mathcal{M}$ with values in the space of symmetric, definite, positive matrices. The corresponding overdamped Langevin dynamics reads

$$dq_t = \Big(-\beta M(q_t)\nabla V(q_t) + \operatorname{div}(M)(q_t)\Big)dt + \sqrt{2}M^{1/2}(q_t)\, dW_t, \tag{36}$$

where $\operatorname{div} M$ is the vector whose ith component is the divergence of the ith column of the matrix $M = [M_1, \dots, M_d]$:

$$\operatorname{div} M = (\operatorname{div} M_1, \dots, \operatorname{div} M_d)^T .$$

The generator of (36) acts on test functions φ as

$$\mathcal{L}\varphi = \Big(-\beta M\nabla V + \operatorname{div}(M)\Big)^T \nabla\varphi + M : \nabla^2\varphi.$$

A simple computation gives, for any smooth functions φ, ϕ,

$$\langle \varphi, \mathcal{L}\phi \rangle_{L^2(\mu)} = -\int_{\mathcal{M}} (\nabla\varphi)^T M (\nabla\phi)\, d\mu.$$

This shows that the canonical measure μ is invariant by the dynamics (36) (choose $\varphi = 1$).

3.5.2 Definition of the self-diffusion

To simplify the notation, we introduce the total drift

$$F(q) = -\beta M(q)\nabla V(q) + \operatorname{div}(M)(q).$$

A straightforward adaption of the proof of [8, Theorem 1] shows that the effective diffusion tensor associated with the dynamics (36) is well defined and that

$$\mathscr{D} = \lim_{t\to+\infty} \mathbb{E}\left(\frac{Q_t - Q_0}{\sqrt{t}} \otimes \frac{Q_t - Q_0}{\sqrt{t}} \right) = 2\left(\int_{\mathcal{M}} M(q)\, \mu(dq) - \int_0^{+\infty} \mathbb{E}\left[F(q_t) \otimes F(q_0)\right] dt \right).$$

where

$$Q_t = q_0 + \int_0^t F(q_s)\, ds + \sqrt{2}\int_0^t M^{1/2}(q_s)\, dW_s,$$

has values in $\mathbb{R}^d$ and where the expectations are over all initial conditions q_0 distributed according to μ, and over all realizations of the overdamped Langevin dynamics (36). The associated effective self-diffusion

$$\mathcal{D} = \frac{1}{2d}\operatorname{Tr}(\mathscr{D}) = \frac{1}{d}\left(\int_{\mathcal{M}} \operatorname{Tr}(M)\, d\mu - \int_0^{+\infty} \mathbb{E}\left[F(q_t)^T F(q_0)\right] dt \right)$$

can be estimated either via Green-Kubo or Einstein formulas. Note that the above equalities reduce to (14)-(15) when $M = \operatorname{Id}$.

3.5.3 Proposal functions

Proposals that can be used in conjunction with a Metropolis-Hastings procedure, and which avoid the computation of the divergence of M, are presented in [3]. The resulting numerical method is a discretization of the dynamics (36) of the same weak order as the Metropolis-Hastings method based on proposals obtained by the standard Euler-Maruyama scheme (compare the results of [3, Section 5] and Lemma 4 below). In order to simplify the presentation, we therefore restrict ourselves to the simple scheme

$$\widetilde{q}^{n+1} = \Phi_{\Delta t}(q^n, G^n) = q^n + F(q^n)\Delta t + \sqrt{2\Delta t}\, M^{1/2}(q^n)\, G^n, \tag{37}$$

although its practical implementation may be cumbersome. Our results can however straightforwardly be extended to the proposal function described in [3].

As in the previous sections, we consider two types of acceptance/rejection procedures to correct the proposal (37): a Metropolis-Hastings rule and a Barker rule. The spatial dependence of the diffusion matrix has to be taken into account in the acceptance/rejection criteria. In fact, we still consider the acceptance rates (26) but change the definition of $\alpha_{\Delta t}$ as follows:

$$
\alpha_{\Delta t}(q, q') = \beta\Big(V(q') - V(q)\Big) + \frac{1}{4\Delta t}\Big(|q - q' - \Delta t F(q')|^2_{M(q')} - |q' - q - \Delta t F(q)|^2_{M(q)}\Big) \\
+ \frac{1}{2}\Big[\ln\big(\det M(q')\big) - \ln\big(\det M(q)\big)\Big], \tag{38}
$$

where, for a symmetric, definite, positive matrix M, we introduced the norm $|q|_M = \sqrt{q^T M^{-1} q}$. The crucial estimate to obtain error bounds on Green-Kubo formulas and the effective diffusion is the following result, which is the equivalent of Lemma 3 for diffusions with multiplicative noise (see Section 4.7 for the proof).

Lemma 4. *For any $\psi \in C^\infty(\mathcal{M})$,*

$$\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\psi = a\mathcal{L}\psi + \Delta t^\alpha\, r_{\psi, \Delta t}, \tag{39}$$

where $a = 1$ and $\alpha = 1/2$ when a Metropolis-Hastings rule is used, while $a = 1/2$ and $\alpha = 1$ when a Barker rule is considered.

Note that the remainder terms are much larger than for diffusions with additive noise (where $\alpha = 3/2$ for the Metropolis-Hastings rule and $\alpha = 2$ for the Barker rule).

3.5.4 Error estimates on the self-diffusion

A straightforward adpation of the proofs of Lemma 2 (see Remark 13), Theorems 6 and 8 shows that, keeping the same notation as in Section 3.4,

$$\int_0^{+\infty} \mathbb{E}\Big[\psi(q_t)\,\varphi(q_0)\Big]dt = a\sum_{n=0}^{+\infty} \mathbb{E}_{\Delta t}\big[\psi(q^n)\,\varphi(q^0)\big] + \mathrm{O}\left(\Delta t^\alpha\right),$$

so that

$$\frac{1}{d}\left(\int_{\mathcal{M}} \mathrm{Tr}(M)\,d\mu - a\sum_{n=0}^{+\infty} \mathbb{E}_{\Delta t}\big[\psi(q^n)\,\varphi(q^0)\big]\right) = \mathcal{D} + \mathrm{O}\left(\Delta t^\alpha\right), \tag{40}$$

while

$$\lim_{n \to +\infty} \frac{1}{a}\mathbb{E}\left[\left(\frac{Q^n - Q^0}{\sqrt{n\Delta t}}\right)^2\right] = \mathcal{D} + \mathrm{O}\left(\Delta t^\alpha\right), \tag{41}$$

with $a = 1$ and $\alpha = 1/2$ when a Metropolis-Hastings rule is used, while $a = 1/2$ and $\alpha = 1$ when a Barker rule is considered. Note that, once again, the Barker rule allows to reduce the bias in the computation of the self-diffusion, here from $\sqrt{\Delta t}$ to Δt.
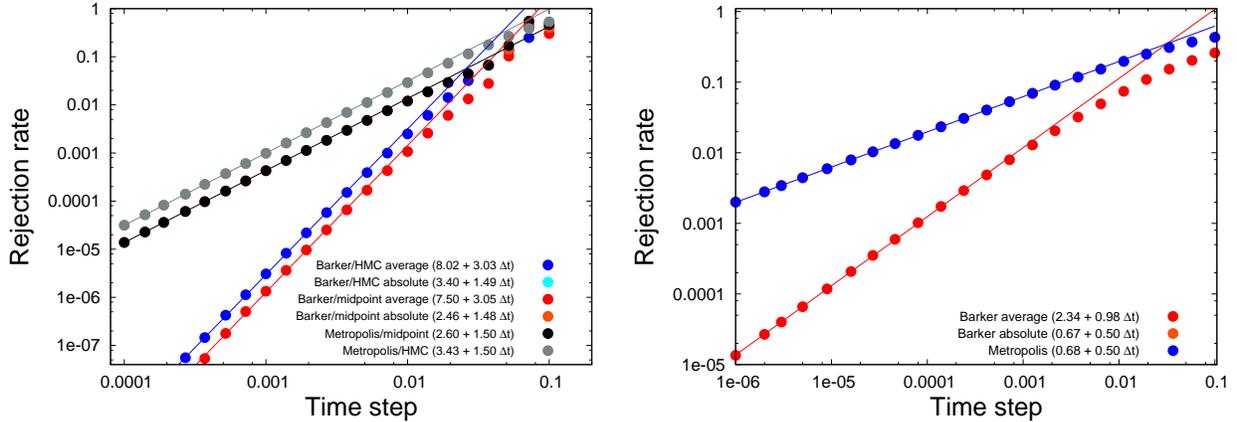
Figure 2: Scaling of the rejection rates. 'Metropolis' corresponds to (42), while 'Barker average' corresponds to the first inequality in (43) and 'Barker absolute' to the second one. Left: dynamics with additive noise. Right: dynamics with multiplicative noise. In both cases, 'Barker absolute' and 'Metropolis' are almost on top of each other.

3.6 Numerical results

We illustrate on a simple example the results obtained in the previous sections. We consider a one-dimensional system with $\mathcal{M} = \mathbb{T}$, at the inverse temperature $\beta = 1$, and for the simple potential $V(q) = \cos(2\pi q)$ already used in [8]. For the dynamics with multiplicative noise, we consider the positive diffusion coefficient

$$M(q) = \left(\frac{1.5 + \cos(2\pi q)}{2} \right)^2.$$

We first study the average rejection rates, based on the computations performed in Section 4.3 and 4.7. For the Metropolis rule, the predictions are

$$0 \leqslant \mathbb{E}\left(1 - A_{\Delta t}^{\mathrm{MH}}\Big(q, \Phi_{\Delta t}(q, G)\Big) \right) \leqslant K_{\mathrm{MH}} \Delta t^{\alpha}, \tag{42}$$

with $\alpha = 3/2$ for dynamics with additive noise (see (59)), and $\alpha = 1/2$ for dynamics with multiplicative noise (see (77)). For the Barker rule, the rejection rate satisfies,

$$\left| 2\mathbb{E}\left(A_{\Delta t}^{\mathrm{Barker}}\Big(q, \Phi_{\Delta t}(q, G)\Big) \right) - 1 \right| \leqslant K_{\mathrm{Barker}} \Delta t^{\alpha}, \qquad \mathbb{E}\left| 2A_{\Delta t}^{\mathrm{Barker}}\Big(q, \Phi_{\Delta t}(q, G)\Big) - 1 \right| \leqslant \widetilde{K}_{\mathrm{Barker}} \Delta t^{\gamma}, \tag{43}$$

with $\alpha = 3$ and $\gamma = 3/2$ for dynamics with additive noise (see (55) and (57)), and $\alpha = 1$ and $\gamma = 1/2$ for dynamics with multiplicative noise (see (78)). In fact,

$$\mathbb{E}\left| 2A_{\Delta t}^{\mathrm{Barker}}\Big(q, \Phi_{\Delta t}(q, G)\Big) - 1 \right| \simeq \mathbb{E}\left(1 - A_{\Delta t}^{\mathrm{MH}}\Big(q, \Phi_{\Delta t}(q, G)\Big) \right).$$

The rejection rates were averaged over 10^9 iterations at $\beta = 1$, starting from $q^0 = 0$. For the midpoint scheme, the proposed configuration is computed using a fixed point algorithm whose tolerance is set to 10^{-8} (this tolerance should be decreased in order to check the scaling of the rejection for smaller timesteps). About 10 fixed point iterations were needed for convergence for $\Delta t = 0.01$, and less iterations for smaller timesteps. The results presented in Figure 2 confirm the predicted rates.

Figure 3 presents the estimated self-diffusions as a function of the timestep for the dynamics with additive noise (1). We refer to [8, Section 3.1] for the precise definitions of the numerical estimators. Expectations are replaced by empirical averages over K realizations. For Green-Kubo formulas, the sum over the iterations is truncated to some maximal iteration index $\lfloor \tau/\Delta t \rfloor$, where τ is a given time. For the Einstein formula, we monitor the estimated mean-square displacement $\mathbb{E}[(Q^n - Q^0)^2]$ as a function of time up to a maximal iteration index, the slope of this curve (obtained via a least square fit) being proportional
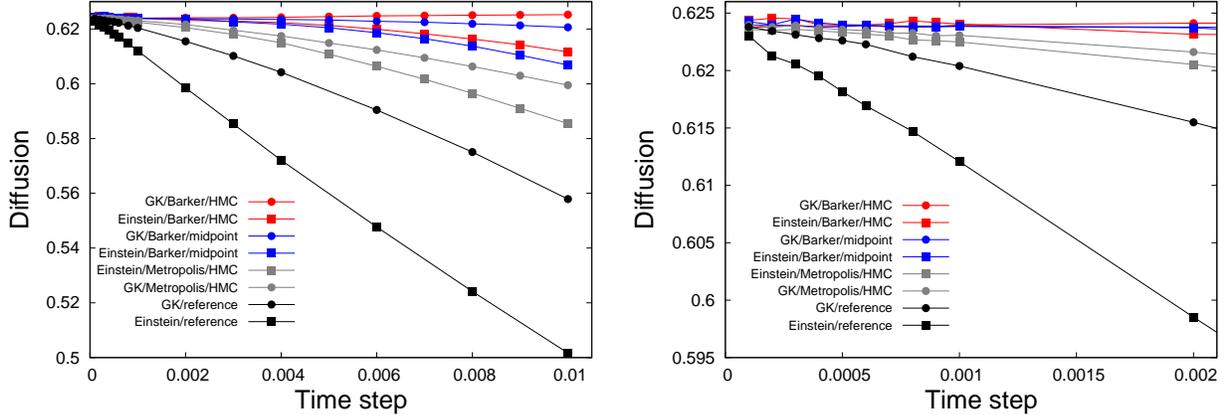
Figure 3: Left: estimated diffusion as a function of the time-step Δt for the dynamics with additive noise. Right: zoom on the smallest values of Δt. The observed scalings of the error correspond to the ones predicted by the theory (namely, Δt^2 when the Barker rule is used with the mipoint or the HMC scheme, $\Delta t^{3/2}$ when a Metropolis rule is used with the latter two proposals, and Δt in the reference case corresponding to the standard MALA scheme).

to the self-diffusion constant for the given timestep. The Green-Kubo estimation (33) was computed using $K = 10^9$ realizations and an integration time $\tau = 0.6$, with initial conditions subsampled every 20 steps from a preliminary trajectory computed with $\Delta t_{\text{thm}} = 0.005$. The Einstein estimation (34) was computed using $K = 10^7$ trajectories integrated over $N_{\text{Einstein}} = 3 \times 10^5$ iterations, with initial conditions subsampled every 1000 steps from a preliminary trajectory computed with $\Delta t_{\text{thm}} = 0.005$. As can be seen from the various curves in Figure 3, the error is dramatically reduced by using the Barker rule, the midpoint and HMC schemes performing quite comparably. We also checked that the expected scalings of the errors as a function of Δt are indeed the ones predicted by our theoretical results. The errors are always larger with a Metropolis rule, though somewhat smaller with the HMC scheme compared to the standard Euler-Maruyama discretization. Also, the Green-Kubo formula leads to more reliable results in this simple case, as already noted in [8]. In fact, with the most precise method (HMC and Barker rule), there is almost no bias up to timesteps of the order of $\Delta t = 0.01$.

Figure 4 presents the estimated self-diffusion constant as a function of the timestep for the dynamics with multiplicative noise (36). The Green-Kubo estimates (40) are obtained with $K = 5 \times 10^6$ realizations for $\Delta t = 10^{-5}$, with a number of realizations increasing proportionally to the timestep up to $K = 10^9$. The integration time is $\tau = 2$. The Einstein estimates (41) were computed with $K = 10^6$ trajectories integrated over $N_{\text{Einstein}} = 10^7$ iterations. The use of a Barker rule instead of a Metropolis rule allows to drastically reduce the bias due to the timestep in the estimation of the self-diffusion constant. Again, the Green-Kubo formula seems more reliable. Finally, note that the error indeed seems to behave as $\sqrt{\Delta t}$ for very small Δt when the Metropolis rule is used. In any case, the variations of the estimated self-diffusion are quite sharp around $\Delta t = 0$ with the Metropolis procedure. In contrast, the estimates obtained with the Barker rule are better behaved, which makes it possible to resort to Romberg extrapolation for moderately small values of the timestep.

4 Proofs

The nth order differential of a function ψ applied to n vectors $v_1, \ldots, v_n \in \mathbb{R}^d$ is written

$$D^n \psi(q) \cdot (v_1 \otimes \cdots \otimes v_n) = \sum_{i_1, \ldots, i_n = 1}^{d} \partial^n_{i_1 \ldots i_n} \psi(q) v_{1,i_1} \ldots v_{n,i_n},$$

15
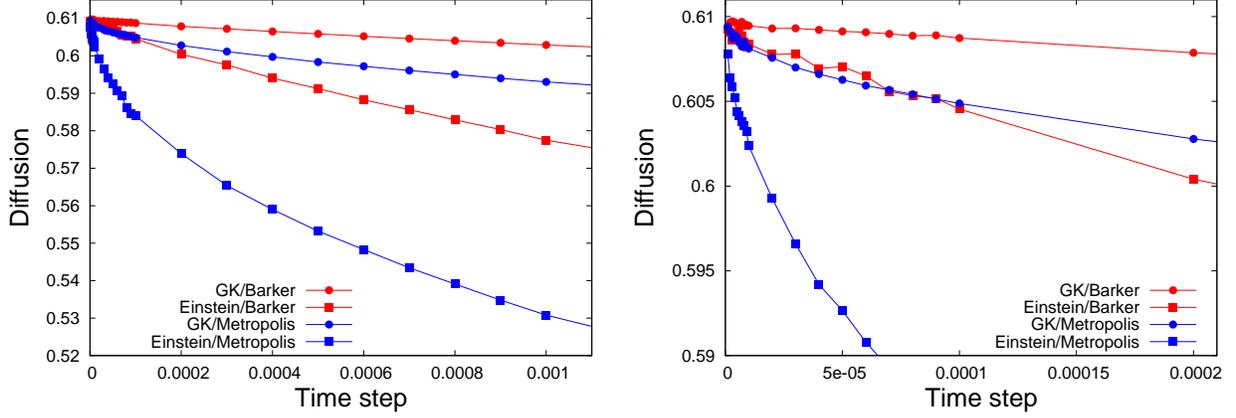
Figure 4: Left: estimated diffusion as a function of the time-step Δt for the dynamics with multiplicative noise. Right: zoom on the smallest values of Δt. The observed scalings of the error correspond to the ones predicted by the theory (namely, Δt for the Barker rule and $\sqrt{\Delta t}$ for the Metropolis rule).

where $v_{k,i}$ is the ith component of the vector v_k. Note that $D^n\psi(q)\cdot(v_1\otimes\cdots\otimes v_n) = D^n\psi(q)\cdot(v_{\sigma(1)}\otimes\cdots\otimes v_{\sigma(n)})$ for any permutation σ. When $v_1 = \cdots = v_n$, the element $v_1\otimes\cdots\otimes v_n$ is simply denoted by $v_1^{\otimes n}$. These definitions can be extended to matrix-valued functions: for a given element $(v_1,\ldots,v_n) \in (\mathbb{R}^d)^n$, the matrix $D^n M(q)\cdot v$ has components $D^n M_{ij}(q)\cdot(v_1\otimes\cdots\otimes v_n)$.

In order to simplify the notation, we write all proofs for $\beta = 1$. The constants $C > 0$ in the inequalities may also vary from one line to another.

4.1   Proof of Lemma 1

In view of (8), we can rewrite $A_{\Delta t}^{\mathrm{mod}}(q,q')$ as $\min(\mathrm{e}^{-\alpha_{\Delta t}(q,')},1)$ with

$$
\begin{aligned}
\alpha_{\Delta t}(q,q') = {} & V(q') - V(q) - \frac{1}{2}\Big[\ln\det\big(\mathrm{Id} + \Delta t\,\sigma(q')\big) - \ln\det\big(\mathrm{Id} + \Delta t\,\sigma(q)\big)\Big] \\
& + \frac{1}{4\Delta t}\big(q - q' + \Delta t\nabla V(q') - \Delta t^2 F(q')\big)^T(\mathrm{Id} + \Delta t\,\sigma(q'))\big(q - q' + \Delta t\nabla V(q') - \Delta t^2 F(q')\big) \\
& - \frac{1}{4\Delta t}\big(q' - q + \Delta t\nabla V(q) - \Delta t^2 F(q)\big)^T(\mathrm{Id} + \Delta t\,\sigma(q))\big(q' - q + \Delta t\nabla V(q) - \Delta t^2 F(q)\big).
\end{aligned}
$$

We first perform an expansion of $\alpha_{\Delta t}\big(q,\Phi_{\Delta t}^{\mathrm{mod}}(q,G)\big)$ up to terms of order $\Delta t^{5/2}$ in order to determine the correction terms $F(q)$ and $\sigma(q)$. The desired conclusion then follows from the inequality

$$
0 \leqslant 1 - \min(1,\mathrm{e}^{-x}) \leqslant \max(0,x). \tag{44}
$$

We rewrite $\alpha_{\Delta t}(q,q')$ as

$$
\begin{aligned}
\alpha_{\Delta t}(q,q') = {} & V(q') - V(q) - \frac{1}{2}(q' - q)^T\big(\nabla V(q) + \nabla V(q')\big) + \frac{\Delta t}{4}\big(|\nabla V(q')|^2 - |\nabla V(q)|^2\big) \\
& - \frac{1}{2}\Big[\ln\det\big(\mathrm{Id} + \Delta t\,\sigma(q')\big) - \ln\det\big(\mathrm{Id} + \Delta t\,\sigma(q)\big)\Big] + \frac{\Delta t}{2}(q' - q)^T\big(F(q) + F(q')\big) \\
& + \frac{1}{4}(q' - q)^T\big(\sigma(q') - \sigma(q)\big)(q' - q) - \frac{\Delta t}{2}(q' - q)^T\big(\sigma(q')\nabla V(q') + \sigma(q)\nabla V(q)\big) \\
& + \Delta t^2\widetilde{\alpha}_{\Delta t}(q,q'),
\end{aligned}
$$

16

where

$$\widetilde{\alpha}_{\Delta t}(q, q') = -\frac{1}{2}\left(F(q')^T \nabla V(q') - F(q)^T \nabla V(q)\right) + \frac{\Delta t}{4}\left(|F(q')|^2 - |F(q)|^2\right)$$

$$+ \frac{1}{2}(q' - q)^T\left(\sigma(q')F(q') + \sigma(q)F(q)\right)$$

$$+ \frac{1}{4}\left[\left(\nabla V(q') - \Delta t F(q')\right)^T \sigma(q')\left(\nabla V(q') - \Delta t F(q')\right) - \left(\nabla V(q) - \Delta t F(q)\right)^T \sigma(q)\left(\nabla V(q) - \Delta t F(q)\right)\right].$$

As made precise below in (46), the remainder $\widetilde{\alpha}_{\Delta t}(q, q')$ can be thought of as being of order $\sqrt{\Delta t}$ when $q' = \Phi_{\Delta t}^{\mathrm{mod}}(q, G)$. We therefore focus on the first terms in the expression of $\alpha_{\Delta t}(q, q')$. A simple computation shows that

$$V(q') - V(q) - \frac{1}{2}(q' - q)^T\left(\nabla V(q) + \nabla V(q')\right) = -\frac{1}{12}D^3V(q) \cdot (q' - q)^{\otimes 3} - \frac{1}{24}D^4V(q) \cdot (q' - q)^{\otimes 4}$$

$$- \frac{1}{24}\int_0^1 (1 - t)^4 D^5V\left((1 - t)q + tq'\right) \cdot (q' - q)^{\otimes 5}\, dt,$$

$$(45)$$

so that

$$V\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G)\right) - V(q) - \frac{1}{2}\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G) - q\right)^T\left(\nabla V(q) + \nabla V\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G)\right)\right)$$

$$= -\frac{\sqrt{2}}{6}\Delta t^{3/2}\, D^3V(q) \cdot G^{\otimes 3} + \Delta t^2\left(\frac{1}{2}D^3V(q) \cdot (\nabla V(q) \otimes G^{\otimes 2}) - \frac{1}{6}D^4V(q) \cdot G^{\otimes 4}\right) + \Delta t^{5/2}\mathcal{R}_1(q, G),$$

where the remainder $\mathcal{R}_1(q, G)$ can be thought of as being uniformly bounded (see (47) below and Remark 10). In the sequel, we no longer explicitly give the expressions of the integral remainder terms such as the last term on the right-hand side of (45) and simply write $\mathrm{O}(|q' - q|^r)$ for some integer power r. With this notation,

$$|\nabla V(q')|^2 - |\nabla V(q)|^2 = 2\nabla V(q)^T \nabla^2 V(q)(q' - q) + (q' - q)^T\left(\nabla^2 V(q)\right)^2(q' - q)$$

$$+ D^3V(q) \cdot \left((q' - q)^{\otimes 2} \otimes \nabla V(q)\right) + \mathrm{O}\left(|q' - q|^3\right),$$

so that

$$\frac{\Delta t}{4}\left(\left|\nabla V\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G)\right)\right|^2 - |\nabla V(q)|^2\right) = \frac{\sqrt{2}\Delta t^{3/2}}{2}\nabla V(q)^T \nabla^2 V(q)G$$

$$+ \frac{\Delta t^2}{2}\left(-\nabla V(q)^T \nabla^2 V(q)\nabla V(q) + G^T\left(\nabla^2 V(q)\right)^2 G + D^3V(q) \cdot \left(\nabla V(q) \otimes G^{\otimes 2}\right)\right) + \Delta t^{5/2}\mathcal{R}_2(q, G).$$

Here and in the sequel, remainders $\mathcal{R}_k(q, G)$ satisfy the bounds made precise in (47). For the next term to consider in the expression of $\alpha_{\Delta t}(q, q')$, we use $\det\left(\mathrm{Id} + \Delta t\,\sigma(q)\right) = 1 + \Delta t\,\mathrm{Tr}(\sigma(q)) + \Delta t^2 r_{\mathrm{det},\Delta t}(q)$ where the remainder $r_{\mathrm{det},\Delta t}$ is a smooth function of q. This leads to

$$\ln \det\left(\mathrm{Id} + \Delta t\,\sigma(q')\right) - \ln \det\left(\mathrm{Id} + \Delta t\,\sigma(q)\right) = \Delta t\,\mathrm{Tr}\left(\sigma(q') - \sigma(q)\right) + \Delta t^2\left(r_{\mathrm{det},\Delta t}(q') - r_{\mathrm{det},\Delta t}(q)\right),$$

so that

$$-\frac{1}{2}\left(\ln \det\left[\mathrm{Id} + \Delta t\,\sigma\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G)\right)\right] - \ln \det\left[\mathrm{Id} + \Delta t\,\sigma(q)\right]\right) = -\frac{\sqrt{2}\Delta t^{3/2}}{2}G^T \nabla\left(\mathrm{Tr}\,\sigma\right)(q)$$

$$+ \frac{\Delta t^2}{2}\left(\nabla V(q)^T \nabla\left(\mathrm{Tr}\,\sigma\right)(q) - \mathrm{Tr}\left[D^2\sigma(q) \cdot G^{\otimes 2}\right]\right) + \Delta t^{5/2}\mathcal{R}_3(q, G).$$

Moreover,

$$\frac{1}{2}(q' - q)^T\left(F(q) + F(q')\right) = F(q)^T(q' - q) + \frac{1}{2}(q' - q)^T\left(DF(q) \cdot (q' - q)\right) + \mathrm{O}\left(|q' - q|^3\right),$$

so that

$$\frac{\Delta t}{2}\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G) - q\right)^T\left(F(q) + F\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G)\right)\right)$$

$$= \sqrt{2}\Delta t^{3/2}F(q)^T G + \Delta t^2\left(G^T\left(DF(q) \cdot G\right) - F(q)^T \nabla V(q)\right) + \Delta t^{5/2}\mathcal{R}_4(q, G).$$

17

In addition,
$$(q' - q)^T \left(\sigma(q') - \sigma(q) \right)(q' - q) = (q' - q)^T \left(D\sigma(q) \cdot (q' - q) \right)(q' - q)$$
$$+ \frac{1}{2}(q' - q)^T \left(D^2\sigma(q) \cdot (q' - q)^{\otimes 2} \right)(q' - q) + \mathrm{O}\left(|q' - q|^5 \right),$$

so that

$$\frac{1}{4}\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G) - q \right)^T \left(\sigma\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G) \right) - \sigma(q) \right)\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G) - q \right) = \frac{\sqrt{2}\Delta t^{3/2}}{2} G^T \left(D\sigma(q) \cdot G \right) G$$
$$+ \Delta t^2 \left(-\nabla V(q)^T \left(D\sigma(q) \cdot G \right) G - \frac{1}{2} G^T \left(D\sigma(q) \cdot \nabla V(q) \right) G + \frac{1}{2} G^T \left(D^2\sigma(q) \cdot G^{\otimes 2} \right) G \right)$$
$$+ \Delta t^{5/2}\mathcal{R}_5(q, G).$$

The last term to consider is
$$(q' - q)^T \left(\sigma(q')\nabla V(q') + \sigma(q)\nabla V(q) \right) = 2(q' - q)^T \sigma(q)\nabla V(q) + (q' - q)^T \left(D\sigma(q) \cdot (q' - q) \right)\nabla V(q)$$
$$+ (q' - q)^T \sigma(q)\nabla^2 V(q)(q' - q) + \mathrm{O}\left(|q' - q|^3 \right),$$

so that

$$-\frac{\Delta t}{2}\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G) - q \right)^T \left(\sigma\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G) \right)\nabla V\left(\Phi_{\Delta t}^{\mathrm{mod}}(q, G) \right) + \sigma(q)\nabla V(q) \right) = -\sqrt{2}\Delta t^{3/2} G^T \sigma(q)\nabla V(q)$$
$$+ \Delta t^2 \left(\nabla V(q)^T \sigma(q)\nabla V(q) - G^T \left(D\sigma(q) \cdot G \right)\nabla V(q) - G^T \sigma(q)\nabla^2 V(q)G \right) + \Delta t^{5/2}\mathcal{R}_6(q, G).$$

In conclusion,
$$\alpha_{\Delta t}\left(q, \Phi_{\Delta t}^{\mathrm{mod}}(q, G) \right) = \Delta t^{3/2}\overline{\alpha}_{3/2}(q, G) + \Delta t^2\overline{\alpha}_2(q, G) + \Delta t^{5/2}\mathcal{R}(q, G),$$

with

$$\overline{\alpha}_{3/2}(q, G) = \frac{\sqrt{2}}{2}\left(-\frac{1}{3}D^3 V(q) \cdot G^{\otimes 3} + G^T \left(D\sigma(q) \cdot G \right) G \right)$$
$$+ \sqrt{2}G^T \left(\frac{1}{2}\nabla^2 V(q)\nabla V(q)^T - \frac{1}{2}\nabla\left(\mathrm{Tr}\,\sigma \right)(q) + F(q) - \sigma(q)\nabla V(q) \right).$$

The choice $\sigma(q) = \nabla^2 V(q)/3$ allows to cancel the cubic term in G, while

$$F(q) = \left(\sigma(q) - \frac{1}{2}\nabla^2 V(q) \right)\nabla V(q) + \frac{1}{2}\nabla\left(\mathrm{Tr}\,\sigma \right)(q) = \frac{1}{6}\left(-\nabla^2 V(q)\nabla V(q) + \frac{1}{6}\nabla(\Delta V)(q) \right)$$

ensures that the linear term in G disappear. For these choices of σ and F, it is easy to see that there exist $K > 0$ and an integer p such that

$$\left| \widetilde{\alpha}_{\Delta t}\left(q, \Phi_{\Delta t}^{\mathrm{mod}}(q, G) \right) \right| \leqslant K(1 + |G|^p)\sqrt{\Delta t}, \tag{46}$$

and

$$|\mathcal{R}_k(q, G)| \leqslant K(1 + |G|^p). \tag{47}$$

In fact, it turns out, quite unexpectedly, that the choice of F and σ also allows to cancel $\overline{\alpha}_2(q, G)$. Indeed,

$$\overline{\alpha}_2(q, G) = -\frac{1}{6}D^4 V(q) \cdot G^{\otimes 4} + \frac{1}{2}G^T \left(D^2\sigma(q) \cdot G^{\otimes 2} \right) G$$
$$+ D^3 V(q) \cdot \left(\nabla V(q) \otimes G^{\otimes 2} \right) + \frac{1}{2}G^T \left(\nabla^2 V(q) \right)^2 G - \frac{1}{2}\mathrm{Tr}\left[D^2\sigma(q) \cdot G^{\otimes 2} \right] + G^T \left(DF(q) \cdot G \right)$$
$$- \frac{1}{2}G^T \left(D\sigma(q) \cdot \nabla V(q) \right) G - 2G^T \left(D\sigma(q) \cdot G \right)\nabla V(q) - G^T \sigma(q)\nabla^2 V(q)G$$
$$+ \nabla V(q)^T \left(\left[\sigma(q) - \frac{1}{2}\nabla^2 V(q) \right]\nabla V(q) + \frac{1}{2}\nabla\left(\mathrm{Tr}\,\sigma \right)(q) - F(q) \right).$$

A simple computation then shows that $\overline{\alpha}_2(q, G) = 0$ (it is clear that the first and last lines in the above expression vanish; the fact that the sum of the second and third lines also vanishes requires a few additional manipulations).

This finally allows to conclude that $\alpha_{\Delta t}\left(q, \Phi_{\Delta t}^{\mathrm{mod}}(q, G) \right) = \Delta t^{5/2}\mathcal{R}(q, G)$ with a remainder $\mathcal{R}(q, G)$ satisfying estimates similar to (47), and gives the desired conclusion in view of (44).

18

Remark 10. *The above proof can be extended to unbounded spaces by following the strategy presented in the proof of [4, Lemma 5.5], assuming somme additional bounds on the derivatives of V, up to order 5.*

4.2   Proof of Theorem 3

The proof follows the same lines as the one performed in [5, Section 4.2] for the usual MALA scheme (trivially adapted to the case of bounded configuration spaces), with the crucial improvement on the rejection rate made precise in Lemma 1. The main ingredient is an improved estimate on the single-step accuracy of the MALA scheme, which is itself obtained from error estimates for the modified (but un-metropolized) Euler scheme. The estimate on the strong accuracy of the scheme finally follows by a discrete Gronwall argument where local errors are summed up over finite time-intervals.

Single-step accuracy of the un-metropolized scheme

We first prove that the modified Euler scheme has the same accuracy as the standard Euler scheme, compare with [5, Lemma 4.4].

Lemma 5 (Single step accuracy of the modified Euler scheme). *There exists $C > 0$ and $\Delta t^* > 0$ such that, for any $0 < \Delta t \leqslant \Delta t^*$ and any $q_0 \in \mathcal{M}$,*

$$\mathbb{E}_{q_0} \left[\left| \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}) - q_{\Delta t} \right|^2 \right] \leqslant C \Delta t^3,$$

$$\left| \mathbb{E}_{q_0} \left(\Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}) - q_{\Delta t} \right) \right| \leqslant C \Delta t^2,$$

where $q_{\Delta t}$ is the solution at time Δt of (1) with initial condition q_0 at $t = 0$, and where the expectation is over all realizations of the Brownian motion $(W_t)_{0 \leqslant t \leqslant \Delta t}$.

Proof. In order to rely on the results already obtained in [5], we introduce the solution of the standard Euler-Maruyama scheme

$$\Phi_{\Delta t}^{\mathrm{EM}}(q, G) = q - \Delta t \nabla V(q) + \sqrt{2\Delta t}\, G.$$

Then, in view of the equality $\Phi_{\Delta t}^{\mathrm{mod}}(q, G) = \Phi_{\Delta t}^{\mathrm{EM}}(q, G) + \Delta t^2 F(q) + \sqrt{2\Delta t}\left[(\mathrm{Id} + \Delta t\, \sigma(q))^{-1/2} - \mathrm{Id} \right] G$ and by a Cauchy-Schwarz inequality,

$$\mathbb{E}_{q_0}\left[\left| \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}) - q_{\Delta t} \right|^2 \right] \leqslant 3\mathbb{E}_{q_0}\left[\left| \Phi_{\Delta t}^{\mathrm{EM}}(q_0, W_{\Delta t}) - q_{\Delta t} \right|^2 \right] + 6\Delta t\, \mathbb{E}_{q_0}\left[\left| \left[(\mathrm{Id} + \Delta t\, \sigma(q_0))^{-1/2} - \mathrm{Id} \right] W_{\Delta t} \right|^2 \right]$$
$$+ 3\Delta t^4 |F(q_0)|^2.$$

It is proved in [5, Lemma 4.4] that $\mathbb{E}_{q_0}\left[\left| \Phi_{\Delta t}^{\mathrm{EM}}(q, W_{\Delta t}) - q_{\Delta t} \right|^2 \right] \leqslant C_{\mathrm{EM}} \Delta t^3$. Moreover,

$$\mathbb{E}_{q_0}\left[\left| \left[(\mathrm{Id} + \Delta t\, \sigma(q_0))^{-1/2} - \mathrm{Id} \right] W_{\Delta t} \right|^2 \right] = \Delta t\, \mathrm{Tr}\left(\left[(\mathrm{Id} + \Delta t\, \sigma(q_0))^{-1/2} - \mathrm{Id} \right]^2 \right) := \Delta t\, g(\Delta t).$$

The function $t \mapsto g(t)$ is well defined for $|t| < 3/\kappa$, with $\kappa := -\min_{\mathcal{M}} \sigma$, and $g(0) = 0$. Therefore, $g(\Delta t) = \Delta t g'(\theta_{\Delta t} \Delta t)$ for some $\theta_{\Delta t} \in [0, 1]$. Now, a simple computation shows that

$$g'(t) = \mathrm{Tr}\left(\sigma(q_0)\big(\mathrm{Id} + t\, \sigma(q_0)\big)^{-3/2} \left[\big(\mathrm{Id} + t\, \sigma(q_0)\big)^{-1/2} - \mathrm{Id} \right] \right),$$

Therefore,

$$\left| g'(t) \right| \leqslant \|\sigma(q_0)\|_{\mathrm{F}} \left(\left\| \big(\mathrm{Id} + t\, \sigma(q_0)\big)^{-3/2} \right\|_{\mathrm{F}} + \left\| \big(\mathrm{Id} + t\, \sigma(q_0)\big)^{-2} \right\|_{\mathrm{F}} \right)$$

$$\leqslant K_d \frac{\|\sigma(q_0)\|_{\mathrm{F}}}{\|\mathrm{Id} + t\, \sigma(q_0)\|_{\mathrm{F}}^{3/2}} \left(1 + \frac{1}{\|\mathrm{Id} + t\, \sigma(q_0)\|_{\mathrm{F}}^{1/2}} \right),$$

for some constant $K_d > 0$ depending on the dimension d. In the above inequalities, we introduced the Frobenius norm $\|M\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}(M^T M)}$ and used the fact that $\mathrm{Id} + t\, \sigma$ is real, positive, symmetric,

19

and that all matrix norms are equivalent, to obtain the second inequality by spectral calculus. Since $\mathrm{Id} + t\,\sigma \geqslant (1 - \kappa t)\mathrm{Id}$ for $t \geqslant 0$, we conclude that

$$\left|g'(t)\right| \leqslant \widetilde{K}_d \frac{\|\sigma(q_0)\|_{\mathrm{F}}}{(1 - \kappa t)^2},$$

for some modified constant $\widetilde{K}_d > 0$. We finally deduce that

$$\mathbb{E}_{q_0}\left[\left|\left[(\mathrm{Id} + \Delta t\,\sigma(q_0))^{-1/2} - \mathrm{Id}\right]W_{\Delta t}\right|^2\right] \leqslant C_1\Delta t^2. \tag{48}$$

In addition, $|F(q)|^2$ is bounded in view of the expression of F, which gives the first estimate.

For the second estimate, we note that

$$\mathbb{E}_{q_0}\left(\Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}) - q_{\Delta t}\right) = \mathbb{E}_{q_0}\left(\Phi_{\Delta t}^{\mathrm{EM}}(q_0, W_{\Delta t}) - q_{\Delta t}\right) + \Delta t^2\,F(q_0).$$

The conclusion then follows from the estimates provided in [5, Lemma 4.4]. $\qquad\square$

Single-step accuracy of the Metropolized scheme

Recall that $\Psi_{\Delta t}^{\mathrm{mod}}(q, G, U)$ is defined in (10).

Lemma 6 (Local accuracy of MALA). *There exists $C > 0$ and $\Delta t^* > 0$ such that, for any $0 < \Delta t \leqslant \Delta t^*$ and any $q_0 \in \mathcal{M}$,*

$$\mathbb{E}_{q_0}\left[\left|\Psi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}, U) - q_{\Delta t}\right|^2\right] \leqslant C\Delta t^3,$$

$$\left|\mathbb{E}_{q_0}\left[\Psi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}, U) - q_{\Delta t}\right]\right| \leqslant C\Delta t^2,$$

where the expectation is over all realizations of the Brownian motion $(W_t)_{0 \leqslant t \leqslant \Delta t}$ and of the random variable $U \sim \mathcal{U}[0, 1]$.

Proof. By definition of the scheme, we have

$$\mathbb{E}_{q_0}\left[\left|\Psi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}, U) - q_{\Delta t}\right|^2\right] = \mathbb{E}_{q_0}\left[A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right)\left|\Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}) - q_{\Delta t}\right|^2\right]$$
$$+ \mathbb{E}_{q_0}\left[\left(1 - A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right)\right)|q_0 - q_{\Delta t}|^2\right].$$

The first term on the right-hand side is bounded using the estimates provided in Lemma 5 and the inequality $0 \leqslant A_{\Delta t}^{\mathrm{mod}}(q, q') \leqslant 1$ for any $q, q' \in \mathbb{R}^d$. For the second term, a Cauchy-Schwarz inequality gives

$$\mathbb{E}_{q_0}\left[\left(1 - A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right)\right)|q_0 - q_{\Delta t}|^2\right]$$
$$\leqslant \mathbb{E}_{q_0}\left[|q_0 - q_{\Delta t}|^4\right]^{1/2}\mathbb{E}_{q_0}\left[\left(1 - A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right)\right)^2\right]^{1/2}.$$

From [5, Lemma 4.2], we know that, for any integer $\ell \geqslant 1$, there exists a constant $K_\ell > 0$ such that

$$\mathbb{E}_{q_0}\left[|q_0 - q_{\Delta t}|^{2\ell}\right] \leqslant K_\ell\Delta t^\ell. \tag{49}$$

The conclusion then follows by the above inequality in the case $\ell = 2$ and by using the estimates stated in Lemma 1.

The proof of the second estimate is based on similar arguments. First,

$$\left|\mathbb{E}_{q_0}\left[\Psi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}, U) - q_{\Delta t}\right]\right| \leqslant \left|\mathbb{E}_{q_0}\left[A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right)\left(\Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}) - q_{\Delta t}\right)\right]\right|$$
$$+ \left|\mathbb{E}_{q_0}\left[\left(1 - A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right)\right)(q_0 - q_{\Delta t})\right]\right|$$

The second term can be bounded by a Cauchy-Schwarz inequality as

$$\left| \mathbb{E}_{q_0} \left[\left(1 - A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right)\right)(q_0 - q_{\Delta t}) \right] \right|$$

$$\leqslant \mathbb{E}_{q_0} \left[\left(1 - A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right)\right)^2 \right]^{1/2} \mathbb{E}_{q_0} \left[|q_0 - q_{\Delta t}|^2 \right]^{1/2},$$

together with (49) in the case $\ell = 1$ and Lemma 1. To estimate the first term, we write

$$\mathbb{E}_{q_0} \left[A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right) \left(\Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}) - q_{\Delta t}\right) \right] = \mathbb{E}_{q_0} \left[\Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}) - q_{\Delta t} \right]$$

$$- \mathbb{E}_{q_0} \left[\left(1 - A_{\Delta t}^{\mathrm{mod}}\left(q_0, \Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t})\right)\right) \left(\Phi_{\Delta t}^{\mathrm{mod}}(q_0, W_{\Delta t}) - q_{\Delta t}\right) \right],$$

and use Lemma 5 to bound the first term on the right-hand side and a Cauchy-Schwarz inequality together with Lemmas 1 and 5 for the second. □

Global accuracy of the Metropolized scheme

We now have all the tools we need to prove Theorem 3. For $k = 0, .., \lfloor T/\Delta t \rfloor$, we introduce $t_k := k\Delta t$ and

$$\varepsilon_k := \mathbb{E}_{q_0} \left[\left| q^k - q_{t_k} \right|^2 \right],$$

where we recall that the expectation is over all realizations of the Brownian motion $(W_t)_{0 \leqslant t \leqslant T}$ for a given initial position $q_0 = q^0$. The Gaussian increments used in the Metropolis scheme are consistent with the realization of the Brownian motion used to integrate the continuous dynamics. More precisely, starting from $q^0 = q_0$, we consider, for $k \geqslant 1$,

$$q^k = \Psi_{\Delta t}^{\mathrm{mod}}\left(q^{k-1}, W_{t_k} - W_{t_{k-1}}, U^k\right).$$

We claim that there exist $K_1, K_2 > 0$ and $\Delta t^* > 0$ such that, for any $0 < \Delta t \leqslant \Delta t^*$,

$$\varepsilon_{k+1} \leqslant (1 + K_1 \Delta t)\varepsilon_k + K_2 \Delta t^3. \tag{50}$$

Theorem 3 then follows by a discrete Gronwall inequality.

Let us now prove (50). We denote by $Q_{t,s}(q)$ be the value at time t of the solution of the SDE (1) starting at time s from q, which depends on the realization of the underlying Brownian motion. Let $\mathcal{F}_k$ be the sigma-algebra of events up to the time t_k. It holds

$$\left| q^{k+1} - q_{t_{k+1}} \right|^2 = \left| q^{k+1} - Q_{t_{k+1}, t_k}(q^k) + Q_{t_{k+1}, t_k}(q^k) - Q_{t_{k+1}, t_k}(q_{t_k}) \right|^2$$

$$= \left| q^{k+1} - Q_{t_{k+1}, t_k}(q^k) \right|^2 + \left| Q_{t_{k+1}, t_k}(q^k) - Q_{t_{k+1}, t_k}(q_{t_k}) \right|^2$$

$$+ 2 \left(q^{k+1} - Q_{t_{k+1}, t_k}(q^k) \right)^T \left(Q_{t_{k+1}, t_k}(q^k) - Q_{t_{k+1}, t_k}(q_{t_k}) \right).$$

Lemma 6 implies that

$$\mathbb{E}\left[\left| q^{k+1} - Q_{t_{k+1}, t_k}(q^k) \right|^2 \Big| \mathcal{F}_k \right] \leqslant C\Delta t^3,$$

so that

$$\mathbb{E}_{q_0}\left[\left| q^{k+1} - Q_{t_{k+1}, t_k}(q^k) \right|^2 \right] \leqslant C\Delta t^3.$$

Similarly, using [5, Lemma 4.3], there exists $\widetilde{K}_1 > 0$ such that

$$\mathbb{E}_{q_0}\left[\left| Q_{t_{k+1}, t_k}(q^k) - Q_{t_{k+1}, t_k}(q_{t_k}) \right|^2 \right] \leqslant \left(1 + \widetilde{K}_1 \Delta t\right) \varepsilon_k. \tag{51}$$

It therefore remains to bound the third term on the right-hand side of (51). Setting

$$\Delta_k := Q_{t_{k+1}, t_k}(q^k) - Q_{t_{k+1}, t_k}(q_{t_k}) - \left(q^k - q_{t_k}\right),$$

we can rewrite the term under consideration as

$$\left(q^{k+1} - Q_{t_{k+1},t_k}(q^k)\right)^T \left(q^k - q_{t_k}\right) + \left(q^{k+1} - Q_{t_{k+1},t_k}(q^k)\right)^T \Delta_k.$$

Using a Cauchy-Schwarz inequality and Lemma 6,

$$\mathbb{E}_{q_0}\left[\left(q^{k+1} - Q_{t_{k+1},t_k}(q^k)\right)^T \left(q^k - q_{t_k}\right)\right] = \mathbb{E}_{q_0}\left[\left(\mathbb{E}\left[q^{k+1} - Q_{t_{k+1},t_k}(q^k)\middle|\mathcal{F}_k\right]\right)^T \left(q^k - q_{t_k}\right)\right]$$

$$\leqslant \mathbb{E}_{q_0}\left[\left|\mathbb{E}\left[q^{k+1} - Q_{t_{k+1},t_k}(q^k)\middle|\mathcal{F}_k\right]\right|^2\right]^{1/2} \varepsilon_k^{1/2}$$

$$\leqslant C\Delta t^2 \varepsilon_k^{1/2} \leqslant \frac{C\Delta t}{2}(\varepsilon_k + \Delta t^2).$$

Similarly,

$$\mathbb{E}_{q_0}\left[\left(q^{k+1} - Q_{t_{k+1},t_k}(q^k)\right)^T \Delta_k\right] \leqslant \mathbb{E}_{q_0}\left(\mathbb{E}\left[\left|q^{k+1} - Q_{t_{k+1},t_k}(q^k)\right|^2\middle|\mathcal{F}_k\right]\right)^{1/2} \mathbb{E}_{q_0}\left(\mathbb{E}\left[|\Delta_k|^2\middle|\mathcal{F}_k\right]\right)^{1/2}$$

$$\leqslant C\Delta t^{3/2}\,\mathbb{E}_{q_0}\left(\mathbb{E}\left[|\Delta_k|^2\middle|\mathcal{F}_k\right]\right)^{1/2}.$$

According to [5, Lemma 4.3],

$$\mathbb{E}\left[|\Delta_k|^2\middle|\mathcal{F}_k\right] \leqslant K\Delta t^2\left|q^k - q_{t_k}\right|,$$

so that, by a Cauchy-Schwarz inequality,

$$\mathbb{E}_{q_0}\left(\mathbb{E}\left[|\Delta_k|^2\middle|\mathcal{F}_k\right]\right) \leqslant K\Delta t^2\,\mathbb{E}_{q_0}\left|q^k - q_{t_k}\right| \leqslant K\Delta t^2\,\varepsilon_k^{1/2}.$$

Therefore, there exists $\widetilde{C} > 0$ such that

$$\mathbb{E}_{q_0}\left[\left(q^{k+1} - Q_{t_{k+1},t_k}(q^k)\right)^T \Delta_k\right] \leqslant \widetilde{C}\Delta t^{5/2}\varepsilon_k^{1/4} \leqslant \frac{3\widetilde{C}}{4}(\Delta t\,\varepsilon_k + \Delta t^3),$$

which concludes the proof of (50).

4.3 Proof of Lemma 3

Midpoint proposal with Barker rule. We first consider the midpoint proposal (23) together with a Barker rule. We also write the proof for a general drift $F_{\Delta t} = -\nabla V + \Delta t\,\widetilde{F}$ in order to prove the statements of Remark 9. We start by rewriting (27) as

$$\alpha_{\Delta t}(q,q') = V\left(\frac{q+q'}{2} + \frac{1}{2}(q'-q)\right) - V\left(\frac{q+q'}{2} - \frac{1}{2}(q'-q)\right) + F_{\Delta t}\left(\frac{q+q'}{2}\right)\cdot(q'-q)$$

$$= \left[\nabla V\left(\frac{q+q'}{2}\right) + F_{\Delta t}\left(\frac{q+q'}{2}\right)\right]\cdot(q'-q) + \frac{1}{24}\left[D^3 V\left(\frac{q+q'}{2}\right)\right](q'-q)^{\otimes 3} + \mathrm{O}\left(|q'-q|^5\right)$$

$$= \Delta t\,\widetilde{F}\left(\frac{q+q'}{2}\right)\cdot(q'-q) + \frac{1}{24}\left[D^3 V\left(\frac{q+q'}{2}\right)\right](q'-q)^{\otimes 3} + \mathrm{O}\left(|q'-q|^5\right).$$

Note that the remainder is of order $\mathrm{O}\left(|q'-q|^5\right)$ and not $\mathrm{O}\left(|q'-q|^4\right)$. Next,

$$\Phi_{\Delta t}(q,G) = q + \sqrt{2\Delta t}\,G - \Delta t\,\nabla V(q) - \frac{\sqrt{2}}{2}\Delta t^{3/2}\nabla^2 V(q)\cdot G$$

$$+ \Delta t^2\left(-\frac{1}{4}D^3 V(q)\cdot G^{\otimes 2} + \frac{1}{2}\nabla^2 V(q)\cdot\nabla V(q) + \widetilde{F}(q)\right) + \Delta t^{5/2}\,Q_{\Delta t}(q^n,G^n). \tag{52}$$

This leads to

$$\alpha_{\Delta t}\left(q,\Phi_{\Delta t}(q,G)\right) = \Delta t\,\widetilde{F}\left(\frac{q+\Phi_{\Delta t}(q,G)}{2}\right)\cdot\left(\sqrt{2\Delta t}G - \Delta t\,\nabla V(q)\right)$$

$$+ \frac{1}{24}D^3 V\left(\frac{q+\Phi_{\Delta t}(q,G)}{2}\right)\cdot\left((2\Delta t)^{3/2}G^{\otimes 3} - 6\Delta t^2 G^{\otimes 2}\otimes\nabla V(q)\right) + \Delta t^{5/2}\widehat{\alpha}_{\Delta t}(q,G)$$

$$= \Delta t^{3/2}\xi_{3/2}(q,G) + \Delta t^2\xi_2(q,G) + \Delta t^{5/2}\xi_{5/2}(q,G) + \Delta t^3\widetilde{\alpha}_{\Delta t}(q,G),$$

22

with

$$\xi_{3/2}(q,G) = \sqrt{2}\left(\widetilde{F}(q)\cdot G + \frac{1}{12}D^3V(q)\cdot(G)^{\otimes 3}\right),$$

$$\xi_2(q,G) = -\widetilde{F}(q)\cdot\nabla V(q) + D\widetilde{F}(q)\cdot(G)^{\otimes 2} - \frac{1}{4}D^3V(q)\cdot\left((G)^{\otimes 2}\otimes\nabla V(q)\right) + \frac{1}{12}D^4V(q)\cdot(G)^{\otimes 4},$$

and where $\xi_{5/2}(q,G)$ involves only odd powers of G and satisfies an estimate of the form

$$|\xi_{5/2}(q,G)| \leqslant L(1+|G|^p) \tag{53}$$

for some integer p and some constant $L > 0$. A similar bound is satisfied by the remainder $\widetilde{\alpha}_{\Delta t}(q,G)$. Using

$$\frac{e^{-a}}{1+e^{-a}} = \frac{1}{2} - \frac{a}{4} + O(a^3), \tag{54}$$

it follows that

$$A_{\Delta t}\left(q,\Phi_{\Delta t}(q,G)\right) = \frac{1}{2} - \frac{\Delta t^{3/2}}{4}\xi_{3/2}(q,G) - \frac{\Delta t^2}{4}\xi_2(q,G) - \frac{\Delta t^{5/2}}{4}\xi_{5/2}(q,G) + \Delta t^3\widetilde{A}_{\Delta t}(q,G), \tag{55}$$

where the remainder $\widetilde{A}_{\Delta t}(q,G)$ satisfies an estimate of the form (53) uniformly in Δt.

Remark 11 (Average rejection rate)**.** *Note that* $\mathbb{E}_G(\xi_{3/2}(q,G)) = \mathbb{E}_G(\xi_{5/2}(q,G)) = 0$*, while*

$$\overline{\xi}_2(q) = \mathbb{E}_G\left[\xi_2(q,G)\right] = -\widetilde{F}(q)\cdot\nabla V(q) + \mathrm{div}\left(\widetilde{F}\right)(q) - \frac{1}{4}\nabla(\Delta V)\cdot\nabla V(q) + \frac{1}{4}\Delta^2 V(q)$$

is such that

$$\int_{\mathcal{M}}\overline{\xi}_2\,d\mu = 0,$$

since

$$\int_{\mathcal{M}}\Delta^2 V\,d\mu = \int_{\mathcal{M}}\nabla(\Delta V)\cdot\nabla V\,d\mu. \tag{56}$$

Therefore, the average acceptance rate at equilibrium is

$$\mathbb{E}_\mu\mathbb{E}_G\left[A_{\Delta t}\left(q,\Phi_{\Delta t}(q,G)\right)\right] = \frac{1}{2} + O\left(\Delta t^3\right), \tag{57}$$

where the expectation is over $q\sim\mu$ *and all realizations of the Gaussian random variable* G*.*

On the other hand, the Taylor expansion

$$\psi\left(\Phi_{\Delta t}(q,G)\right) - \psi(q) = \nabla\psi(q)\cdot(\Phi_{\Delta t}(q,G) - q) + \frac{1}{2}D^2\psi(q)\cdot(\Phi_{\Delta t}(q,G) - q)^{\otimes 2} + \frac{1}{6}D^3\psi(q)\cdot(\Phi_{\Delta t}(q,G) - q)^{\otimes 3} + \ldots$$

leads to

$$\psi\left(\Phi_{\Delta t}(q,G)\right) - \psi(q) = \nabla\psi(q)\cdot\left(\sqrt{2\Delta t}\,G - \Delta t\,\nabla V\left(\frac{q+\Phi_{\Delta t}(q,G)}{2}\right) + \Delta t^2\widetilde{F}(q)\right)$$

$$+ \frac{1}{2}\,D^2\psi(q)\cdot\left(2\Delta t\,G^{\otimes 2} - 2^{3/2}\Delta t^{3/2}G\otimes\nabla V\left(\frac{q+\Phi_{\Delta t}(q,G)}{2}\right) + \Delta t^2\nabla V(q)\otimes\nabla V(q)\right)$$

$$+ \frac{1}{6}\,D^3\psi(q)\cdot\left(2^{3/2}\Delta t^{3/2}G^{\otimes 3} - 6\Delta t^2\nabla V(q)\otimes G^{\otimes 2}\right)$$

$$+ \frac{\Delta t^2}{6}\,D^4\psi(q)\cdot G^{\otimes 4} + \Delta t^{5/2}\widetilde{\Psi}_{5/2}(q,G) + \Delta t^3\widehat{\Psi}_{\Delta t}(q,G),$$

with $\widetilde{\Psi}_{5/2}(q,G)$ involving only odd powers of G. Upon further expanding

$$\nabla V\left(\frac{q+\Phi_{\Delta t}(q,G)}{2}\right) = \nabla V(q) + \frac{1}{2}D^2V(q)\cdot(\Phi_{\Delta t}(q,G) - q) + \frac{1}{8}D^3V(q)\cdot(\Phi_{\Delta t}(q,G) - q)^{\otimes 2} + \ldots$$

$$= \nabla V(q) + \frac{1}{2}D^2V(q)\cdot\left(\sqrt{2\Delta t}\,G - \Delta t\,\nabla V(q)\right) + \frac{\Delta t}{4}D^3V(q)\cdot G^{\otimes 2} + \Delta t^{3/2}\widetilde{V}_{\Delta t}(q,G),$$

23

it follows

$$
\begin{aligned}
\psi\left(\Phi_{\Delta t}(q, G)\right) - \psi(q) &= \sqrt{2\Delta t}\,\nabla\psi(q) \cdot G \\
&\quad + \Delta t\left(-\nabla V(q) \cdot \nabla\psi(q) + D^2\psi(q) \cdot G^{\otimes 2}\right) \\
&\quad + \Delta t^{3/2}\left(-\frac{\sqrt{2}}{2}D^2 V(q) \cdot (G \otimes \nabla\psi(q)) - \sqrt{2}D^2\psi(q) \cdot (G \otimes \nabla V(q)) + \frac{\sqrt{2}}{3}D^3\psi(q) \cdot G^{\otimes 3}\right) \\
&\quad + \Delta t^2 T_{4,\psi}(q, G) + \Delta t^{5/2}\Psi_{5/2}(q, G) + \Delta t^3 \widetilde{\Psi}_{\Delta t}(q, G),
\end{aligned}
\tag{58}
$$

with $\Psi_{5/2}(q, G)$ involving only odd powers of G and

$$
\begin{aligned}
T_{4,\psi}(q, G) &= \widetilde{F}(q) \cdot \nabla\psi(q) + \frac{1}{2}D^2 V(q) \cdot (\nabla V(q) \otimes \nabla\psi(q)) - \frac{1}{4}D^3 V(q) \cdot \left(\nabla\psi(q) \otimes G^{\otimes 2}\right) \\
&\quad - D^2\psi(q) \cdot \left(G \otimes \left(D^2 V(q) \cdot G\right)\right) + \frac{1}{2}D^2\psi(q) \cdot (\nabla V(q) \otimes \nabla V(q)) \\
&\quad - D^3\psi(q) \cdot \left(\nabla V(q) \otimes G^{\otimes 2}\right) + \frac{1}{6}D^4\psi(q) \cdot G^{\otimes 4}.
\end{aligned}
$$

A simple computation shows that $\overline{T_{4,\psi}}(q) = \mathbb{E}_G[T_{4,\psi}(q, G)]$ is equal to

$$
\overline{T_{4,\psi}} = \widetilde{F}\cdot\nabla\psi + \frac{1}{2}(\nabla V)^T \nabla^2 V(\nabla\psi) - \frac{1}{4}\nabla(\Delta V)\cdot\nabla\psi - \nabla^2 V : \nabla^2\psi + \frac{1}{2}(\nabla V)^T\nabla^2\psi(\nabla V) - \nabla V\cdot\nabla(\Delta\psi) + \frac{1}{2}\Delta^2\psi.
$$

Recalling $\mathcal{L} = -\nabla V \cdot \nabla + \Delta$ and using the expression of $\mathcal{L}^2$ provided in [12, Section 4.9],

$$
\overline{T_{4,\psi}} = \frac{1}{2}\mathcal{L}^2\psi + \left(\widetilde{F} + \frac{1}{4}\nabla(\Delta V)\right) \cdot \nabla\psi.
$$

Therefore, irrespectively of the choice of $\widetilde{F}$, and in view of (55) and the definition of the discrete generator (29),

$$
\begin{aligned}
\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\psi(q) &= \frac{1}{2}\mathbb{E}_G\left[\frac{\psi(\Phi_{\Delta t}(q, G)) - \psi(q)}{\Delta t}\right] - \frac{\sqrt{2}\Delta t}{4}\mathbb{E}_G\left[\xi_{3/2}(q, G)\left(\nabla\psi(q) \cdot G\right)\right] + \Delta t^2\widehat{r}_{\psi,\Delta t}(q) \\
&= \frac{1}{2}\left[\mathcal{L}\psi(q) + \Delta t\left(\frac{1}{2}\mathcal{L}^2\psi(q) + \left(\widetilde{F} + \frac{1}{4}\nabla(\Delta V)\right) \cdot \nabla\psi(q)\right)\right] \\
&\quad - \frac{\Delta t}{2}\left[\left(\widetilde{F} + \frac{1}{4}\nabla(\Delta V)\right) \cdot \nabla\psi\right](q) + \Delta t^2 r_{\psi,\Delta t}(q) \\
&= \frac{1}{2}(\mathcal{L}\psi)(q) + \frac{\Delta t}{4}(\mathcal{L}^2\psi)(q) + \Delta t^2 r_{\psi,\Delta t}(q),
\end{aligned}
$$

which gives the claimed result (35) for the midpoint proposal (23) when a Barker rule is used.

Midpoint proposal with Metropolis rule. By distinguishing the cases $x \leqslant 0$ and $x \geqslant 0$,

$$
x_+ - \frac{x_+^2}{2} \leqslant 1 - \min(1, \mathrm{e}^{-x}) \leqslant x_+, \qquad x_+ = \max(0, x).
$$

Therefore, when the proposal (23) is considered with a modified drift $-\nabla V + \Delta t\,\widetilde{F}$ in conjunction with a Metropolis-Hastings algorithm, the rejection rate can be expanded as

$$
A_{\Delta t}^{\mathrm{MH}}(q^n, \widetilde{q}^{n+1}) = 1 - \Delta t^{3/2}\min\left(0, \xi_{3/2}(q^n, G^n)\right) + \Delta t^2\,\widetilde{A}_{\Delta t}^{\mathrm{MH}}(q^n, G^n),
\tag{59}
$$

with a remainder satisfying an inequality similar to (53). However, the remainder has a non-trivial average with respect to G as $\Delta t \to 0$, in contrast to the case where a Barker rule is used. Therefore, with computations similar to the ones of [8, Section 5.2],

$$
\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\psi = \mathcal{L}\psi + \Delta t\left(\frac{1}{2}\mathcal{L}^2 + \left(\widetilde{F} + \frac{1}{4}\nabla(\Delta V) + f_{3/2}\right) \cdot \nabla\psi\right) + \Delta t^{3/2}r_{\psi,\Delta t},
$$

24

where

$$f_{3/2}(q) = -\sqrt{2} \int_{\mathbb{R}^d} \min\left(0, \xi_{3/2}(q,g)\right) g \, \frac{\mathrm{e}^{-g^2/2}}{(2\pi)^{d/2}} \, dg.$$

Let us insist on the fact that the remainder now is of order $\Delta t^{3/2}$ rather than Δt^2 as in the Barker case. As in [3, Section 5], it is possible to obtain a simpler expression of $f_{3/2}$ in view of the symmetry property

$$\xi_{3/2}(q,g)g = -\xi_{3/2}(q,-g)g. \tag{60}$$

Introducing $\Omega(q) = \{g \in \mathbb{R}^d \,|\, \xi_{3/2}(q,g) \geqslant 0\}$,

$$f_{3/2}(q) = -\sqrt{2} \int_{\Omega(q)} \xi_{3/2}(q,g) \, g^T \nabla\psi \, \frac{\mathrm{e}^{-g^2/2}}{(2\pi)^{d/2}} \, dg$$

$$= -\frac{\sqrt{2}}{2} \int_{\mathbb{R}^d} \xi_{3/2}(q,g) \, g^T \nabla\psi \, \frac{\mathrm{e}^{-g^2/2}}{(2\pi)^{d/2}} \, dg = -\left(\widetilde{F}(q) + \frac{1}{4}\nabla(\Delta V)(q) \right).$$

Therefore, the Metropolis algorithm based on the midpoint proposal is of weak order 2, but with a fractional remainder of order $\Delta t^{5/2}$.

HMC proposal with Barker rule. We now set $\widetilde{F} = 0$ since the previous computations show that $\widetilde{F}$ does not change the weak type properties of the algorithm. For the HMC proposal (24), the expansion (52) is changed as

$$\Phi_{\Delta t}(q,G) = q + \sqrt{2\Delta t}\,G - \Delta t\,\nabla V(q) - \frac{\sqrt{2}}{2}\Delta t^{3/2}\nabla^2 V(q)\cdot G$$
$$- \frac{\Delta t^2}{4}D^3 V(q)\cdot G^{\otimes 2} + \Delta t^{5/2}\,Q_{\Delta t}(q^n,G^n). \tag{61}$$

Note that only the term in Δt^2 changes. Therefore, (58) holds upon changing $T_{4,\psi}(q,G)$ to

$$T_{4,\psi}(q,G) = -\frac{1}{4}D^3 V(q)\cdot \left(\nabla\psi(q) \otimes G^{\otimes 2}\right)$$
$$- D^2\psi(q)\cdot \left(G \otimes \left(D^2 V(q)\cdot G\right)\right) + \frac{1}{2}\,D^2\psi(q)\cdot \left(\nabla V(q) \otimes \nabla V(q)\right)$$
$$- D^3\psi(q)\cdot \left(\nabla V(q) \otimes G^{\otimes 2}\right) + \frac{1}{6}\,D^4\psi(q)\cdot G^{\otimes 4}.$$

The rate $\alpha_{\Delta t}(q, \Phi_{\Delta t}(q,G))$ defined in (28) is computed by replacing ψ by V in (58) (with the new definition of $T_{4,V}$):

$$V(\Phi_{\Delta t}(q,G)) - V(q) = \sqrt{2\Delta t}\,\nabla V(q)\cdot G + \Delta t\left(-|\nabla V(q)|^2 + D^2 V(q)\cdot G^{\otimes 2}\right)$$
$$+ \Delta t^{3/2}\left(-\frac{3\sqrt{2}}{2}D^2 V(q)\cdot (G \otimes \nabla V(q)) + \frac{\sqrt{2}}{3}\,D^3 V(q)\cdot G^{\otimes 3}\right)$$
$$+ \Delta t^2 T_{4,V}(q,G) + \Delta t^{5/2}\mathcal{V}_{5/2}(q,G) + \Delta t^3\widetilde{\mathcal{V}}_{\Delta t}(q,G),$$

and expanding

$$\nabla V\left(q + \frac{\sqrt{2\Delta t}}{2}\,G\right) = \nabla V(q) + \frac{\sqrt{2\Delta t}}{2}\,D^2 V(q)\cdot G + \frac{\Delta t}{4}D^3 V(q)\cdot G^{\otimes 2} + \frac{\sqrt{2}\Delta t^{3/2}}{24}D^4 V(q)\cdot G^{\otimes 3} + \dots$$

so that

$$\frac{1}{2}\left[\left(G - \sqrt{2\Delta t}\,\nabla V\left(q + \frac{\sqrt{2\Delta t}}{2}\,G\right)\right)^2 - G^2\right]$$

$$= -\sqrt{2\Delta t}\,G \cdot \nabla V\left(q + \frac{\sqrt{2\Delta t}}{2}\,G\right) + \Delta t\left|\nabla V\left(q + \frac{\sqrt{2\Delta t}}{2}\,G\right)\right|^2$$

$$= -\sqrt{2\Delta t}\,G \cdot \nabla V(q) + \Delta t\left(|\nabla V(q)|^2 - D^2 V(q) \cdot G^{\otimes 2}\right)$$

$$+ \Delta t^{3/2}\left(-\frac{\sqrt{2}}{4}\,D^3 V(q) \cdot G^{\otimes 3} + \sqrt{2}D^2 V(q) \cdot (G \otimes \nabla V(q))\right)$$

$$+ \Delta t^2\left(-\frac{1}{12}D^4 V(q) \cdot G^{\otimes 4} + \frac{1}{2}D^3 V(q) \cdot (G^{\otimes 2} \otimes \nabla V(q)) + \frac{1}{2}\left|\nabla^2 V(q) \cdot G\right|^2\right)$$

$$+ \Delta t^{5/2}\mathcal{K}_{5/2}(q, G) + \Delta t^3 \mathcal{K}_{\Delta t}(q).$$

Finally,

$$\alpha_{\Delta t}(q, \Phi_{\Delta t}(q, G)) = \Delta t^{3/2}\xi_{3/2}(q, G) + \Delta t^2 \xi_2(q, G) + \Delta t^{5/2}\xi_{5/2}(q, G) + \Delta t^3 \widetilde{\alpha}_{\Delta t}(q, G),$$

with

$$\xi_{3/2}(q, G) = \sqrt{2}\left(\frac{1}{12}D^3 V(q) \cdot (G)^{\otimes 3} - \frac{1}{2}D^2 V(q) \cdot (G \otimes \nabla V(q))\right),$$

$$\xi_2(q, G) = -\frac{3}{4}D^3 V(q) \cdot \left(\nabla V(q) \otimes G^{\otimes 2}\right) + \frac{1}{2}D^2 V(q) \cdot (\nabla V(q) \otimes \nabla V(q)) + \frac{1}{12}D^4 V(q) \cdot G^{\otimes 4}$$

$$- \frac{1}{2}\left|\nabla^2 V(q) \cdot G\right|^2,$$

and where $\xi_{5/2}(q, G)$ involves only odd powers of G. In conclusion, the term $D^2 V(q) \cdot (\nabla V(q) \otimes \nabla \psi(q))$, which is absent in the expression of $T_{4,\psi}$ compared to the corresponding expression for the midpoint proposal, is compensated by an extra term in the expression of $\xi_{3/2}$. It is then easy to prove that (35) holds.

Remark 12 (Average rejection rate). *As for the midpoint proposal used with the Barker rule, it is possible to prove that the average rejection rate at equilibrium is $1/2 + \mathrm{O}(\Delta t^3)$ (see (57) above). This computation relies on the fact that*

$$\overline{\xi}_2(q) = \mathbb{E}_G\left[\xi_2(q, G)\right] = -\frac{3}{4}\nabla(\Delta V) \cdot \nabla V(q) + \frac{1}{4}\Delta^2 V(q) + \frac{1}{2}\nabla V^T \nabla^2 V \nabla V(q) - \frac{1}{2}\operatorname{Tr}\left[(\nabla^2 V(q))^2\right]$$

has a vanishing average with respect to μ. To prove the latter statement, we compute

$$\int_{\mathcal{M}} \operatorname{Tr}\left[(\nabla^2 V(q))^2\right]\, d\mu = \sum_{i,j=1}^d \int_{\mathcal{M}}\left(\partial^2_{q_i, q_j} V\right)^2 d\mu = -\sum_{i,j=1}^d \int_{\mathcal{M}} \partial_{q_i} V\left(\partial^3_{q_i, q_j, q_j} V - \partial^2_{q_i, q_j} V \partial_{q_j} V\right)\, d\mu$$

$$= -\int_{\mathcal{M}} \nabla(\Delta V) \cdot \nabla V\, d\mu + \int_{\mathcal{M}} \nabla V^T \nabla^2 V \nabla V\, d\mu,$$

and use (56).

HMC proposal with Metropolis rule. The result is obtained by a straightforward modification of the argument for the midpoint scheme. We therefore omit the proof.

4.4   Proof of Lemma 2

Well posedness of the midpoint scheme. To prove that the implicit method is well defined, we use a fixed-point argument. For a given $q \in \mathcal{M}$ and $G \in \mathbb{R}^d$, we define $Q^0 = q$ and $Q^{k+1} = \mathscr{F}_{\Delta t}(Q^k)$ with

$$\mathscr{F}_{\Delta t}(Q) = q - \Delta t\,\nabla V\left(\frac{q + Q}{2}\right) + \sqrt{2\Delta t}\,G.$$

Note that, upon introducing the global Lipschitz constant L_V of $-\nabla V$,

$$|\mathscr{F}_{\Delta t}(Q) - \mathscr{F}_{\Delta t}(Q')| = \Delta t \left| \nabla V \left(\frac{q+Q}{2} \right) - \nabla V \left(\frac{q+Q'}{2} \right) \right| \leqslant \frac{L_V \Delta t}{2} |Q - Q'|.$$

When $\Delta t < 2/L_V$, the mapping $\mathscr{F}_{\Delta t}$ is a contraction, so that the existence and uniqueness of $\widetilde{q}^{n+1}$ is ensured by the Banach fixed point theorem.

Geometric ergodicity. We prove the geometric ergodicity of schemes based on the midpoint proposal (23), the computations for the HMC proposal being similar. Our aim is to prove that, for a given physical time $T > 0$, there exists $\alpha > 0$ and a probability measure ν on $\mathcal{M}$ such that the following uniform minorization condition holds (see for instance [4, 12] for related estimates):

$$P_{\Delta t}^{\lceil T/\Delta t \rceil}(q, \cdot) \geqslant \alpha \, \nu, \tag{62}$$

where $\lceil x \rceil$ denotes the smallest integer larger than x. The term "uniform" refers to estimates which are independent of the timestep Δt. To obtain such estimates, we have to consider evolutions over fixed physical times, which amounts to iterating the elementary evolution $P_{\Delta t}$ over $\lceil T/\Delta t \rceil$ timesteps. By the results of [10] for instance, (62) implies that there exists $\lambda, \widetilde{C} > 0$ such that, for any $f \in \widetilde{L}^\infty(\mathcal{M})$,

$$\left\| P_{\Delta t}^{\lceil 1/\Delta t \rceil} f \right\|_{L^\infty} \leqslant \widetilde{C} \, \mathrm{e}^{-\lambda} \|f\|_{L^\infty},$$

from which (30) follows.

The strategy of the proof of (62) is the following. We denote by $\widetilde{P}_{\Delta t}$ the transition kernel associated with the Markov chain where we perform a move according to the proposal function, and always accept it. We first show in Lemma 7 that $\widetilde{P}_{\Delta t}^n$ satisfies a uniform minorization condition when iterated for a number of steps larger than a fraction of $T/\Delta t$. We next show that this property is transferred to the scheme $P_{\Delta t}$ where acceptance/rejection is performed according to the Barker or Metropolis rules.

Lemma 7 (Uniform minorization condition for schemes without rejection)**.** *There exist* $\Delta t^*, \widetilde{\alpha} > 0$ *and a probability measure* ν *such that, for any bounded measurable non-negative function* f, *any* $0 < \Delta t \leqslant \Delta t^*$ *and* $q \in \mathcal{M}$,

$$\forall n \in \left\{ \left\lfloor \frac{T}{4\Delta t} \right\rfloor, \dots, \left\lceil \frac{T}{\Delta t} \right\rceil \right\}, \qquad \left(P_{\Delta t}^n f \right)(q) \geqslant \widetilde{\alpha} \int_{\mathcal{M}} f \, d\nu.$$

Proof. It is sufficient to prove the result for indicator functions of Borel sets $E \subset \mathcal{M}$ (see [20]). Denoting by $\widetilde{q}^{k+1} = \Phi_{\Delta t}(\widetilde{q}^k, G^k)$ the iterates of the Markov chain, we therefore aim at proving

$$\forall n \in \left\{ \left\lfloor \frac{T}{4\Delta t} \right\rfloor, \dots, \left\lceil \frac{T}{\Delta t} \right\rceil \right\}, \qquad \mathbb{P}\left(\widetilde{q}^n \in E \mid \widetilde{q}^0 = q \right) \geqslant \widetilde{\alpha} \, \nu(E),$$

for a well chosen probability measure ν and a constant $\widetilde{\alpha} > 0$. The idea of the proof is to explicitly rewrite $\widetilde{q}^n$ as a perturbation of the reference evolution corresponding to $\nabla V = 0$. Since we consider smooth potentials and the position space is compact, the perturbation can be uniformly controlled. More precisely,

$$\widetilde{q}^n = \widetilde{q}^0 + \mathscr{G}^n + \mathscr{F}^n, \tag{63}$$

with

$$\mathscr{G}^n = \sqrt{2\Delta t} \sum_{k=0}^{n-1} G^k, \qquad \mathscr{F}^n = -\Delta t \sum_{k=0}^{n-1} \nabla V \left(\frac{\widetilde{q}^k + \widetilde{q}^{k+1}}{2} \right).$$

Note that $|\mathscr{F}^n| \leqslant \|\nabla V\|_{L^\infty} n \Delta t \leqslant \|\nabla V\|_{L^\infty}(T + \Delta t)$, while $\mathscr{G}^n$ is a Gaussian random variable with covariance matrix $2n\Delta t \, \mathrm{Id}_d$. Therefore,

$$\mathbb{P}\left(\widetilde{q}^n \in E \mid \widetilde{q}^0 = q \right) = \mathbb{P}\left(\mathscr{G}^n \in E - q - \mathscr{F}^n \right) \geqslant \left(\frac{1}{4\pi n \Delta t} \right)^{d/2} \int_{E-q-\mathscr{F}^n} \exp\left(-\frac{|g|^2}{4n\Delta t} \right) dg. \tag{64}$$

In the latter expression and in the sequel, we consider that the random variable $\mathscr{G}^n$ has values in $\mathbb{R}^d$ rather than $\mathcal{M}$ and understand $E - q - \mathscr{F}^n$ as a subset of $\mathbb{R}^d$ rather than $\mathcal{M}$. This amounts to neglecting

the possible periodic images, and henceforth leads to the second inequality. Now, for Δt sufficiently small, it holds $T/8 \leqslant n\Delta t \leqslant 2T$, so that

$$\left(\frac{1}{2\pi n\Delta t}\right)^{d/2} \int_{E-q-\mathscr{F}^n} \exp\left(-\frac{|g|^2}{2n\Delta t}\right) dg \geqslant \left(\frac{1}{4\pi T}\right)^{d/2} \int_{E-q-\mathscr{F}^n} \exp\left(-\frac{4|g|^2}{T}\right) dg.$$

Since the state space is compact, there exists $R > 0$ such that $|q + \mathscr{F}^n| \leqslant R$ for any $q \in \mathcal{M}$. We can then consider the probability measure

$$\nu(E) = Z_R^{-1} \inf_{|Q| \leqslant R} \int_{E+Q} \exp\left(-\frac{4|g|^2}{T}\right) dg,$$

and the constant

$$\widetilde{\alpha} = Z_R \left(\frac{1}{4\pi T}\right)^{d/2},$$

which gives the claimed result. $\qquad\qquad\square$

Let us now show how to adapt the proof of Lemma 7 to the case when the proposals are accepted or rejected according to some rule (Metropolis or Barker). We set $a = 1/2$ for the Barker rule and $a = 1$ for the Metropolis one. Note first that (63) is modified as

$$q^n = q^0 + \mathscr{G}^n + \mathscr{F}^n,$$

with

$$\mathscr{G}^n = \sqrt{2\Delta t} \sum_{k=0}^{n-1} \mathbf{1}_{U^k \leqslant A_{\Delta t}\left(q^k, \Phi_{\Delta t}(q^k, G^k)\right)} G^k,$$

$$\mathscr{F}^n = -\Delta t \sum_{k=0}^{n-1} \mathbf{1}_{U^k \leqslant A_{\Delta t}\left(q^k, \Phi_{\Delta t}(q^k, G^k)\right)} \nabla V\left(\frac{\widetilde{q}^k + \Phi_{\Delta t}(q^k, G^k)}{2}\right).$$

It still holds $|\mathscr{F}^n| \leqslant \|\nabla V\|_{L^\infty}(T + \Delta t)$. To characterize more precisely $\mathscr{G}^n$, we decompose it as $\mathscr{G}^n = \widetilde{\mathscr{G}}^n + \widehat{\mathscr{G}}^n$, where

$$\widetilde{\mathscr{G}}^n = \sqrt{2\Delta t} \sum_{k=0}^{n-1} \mathbf{1}_{U^k \leqslant a} G^k$$

and

$$\widehat{\mathscr{G}}^n = \sqrt{2\Delta t} \sum_{k=0}^{n-1} \left(\mathbf{1}_{U^k \leqslant A_{\Delta t}\left(q^k, \Phi_{\Delta t}(q^k, G^k)\right)} - \mathbf{1}_{U^k \leqslant a}\right) G^k.$$

The latter random variable can be thought of as being small. To quantify this statement, we rewrite each term in the sum defining $\widehat{\mathscr{G}}^n$ as some drift plus a martingale increment, independent of the previous increments. More precisely,

$$\left(\mathbf{1}_{U^k \leqslant A_{\Delta t}\left(q^k, \Phi_{\Delta t}(q^k, G^k)\right)} - \mathbf{1}_{U^k \leqslant a}\right) G^k = D(q^k) + M^k,$$

where $\mathbb{E}(M^k | \mathcal{F}_k) = 0$ ($\mathcal{F}_k$ denoting the filtration of events up to iteration k), and

$$D(q) = \mathbb{E}_{U,G}\left[\left(\mathbf{1}_{U \leqslant A_{\Delta t}(q, \Phi_{\Delta t}(q, G))} - \mathbf{1}_{U \leqslant a}\right) G\right] = \mathbb{E}_G\left[\left(A_{\Delta t}\left(q, \Phi_{\Delta t}(q, G)\right) - a\right) G\right]. \tag{65}$$

In view of (55) and (59), the drift term is of order $\Delta t^{3/2}$: there exists $C > 0$ such that

$$|D(q)| \leqslant C\Delta t^{3/2}. \tag{66}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}\left[\left(M^k\right)^2 \Big| \mathcal{F}_k\right] &= \mathbb{E}_{U,G}\left[\left(\mathbf{1}_{U \leqslant A_{\Delta t}\left(q^k, \Phi_{\Delta t}(q^k, G)\right)} - \mathbf{1}_{U^k \leqslant a}\right)^2 G^2\right] - D(q^k)^2 \\
&= \mathbb{E}_G\left(\left[a + A_{\Delta t}\left(q^k, \Phi_{\Delta t}(q^k, G)\right) - 2\min\left(A_{\Delta t}\left(q^k, \Phi_{\Delta t}(q^k, G)\right), a\right)\right] G^2\right) - D(q^k)^2 \\
&\leqslant K\Delta t^{3/2},
\end{aligned}$$

so that

$$\mathbb{E}_{q^0}\left[\left(\sum_{k=0}^{n-1} M^k\right)^2\right] \leqslant nK\Delta t^{3/2}.$$

Therefore, by the Chebyshev inequality,

$$\mathbb{P}\left(\left|\mathscr{G}^n - \sqrt{2\Delta t}\sum_{k=0}^{n-1} D(q^k)\right| \geqslant \eta\,\sqrt{2Kn\Delta t^{5/2}}\right) = \mathbb{P}\left(\left|\sum_{k=0}^{n-1} M^k\right| \geqslant \eta\,\sqrt{Kn\Delta t^{3/2}}\right) \leqslant \frac{1}{\eta^2}.$$

By considering $\eta = \widetilde{\eta}\Delta t^{-3/4}$, it follows that there exists $C > 0$ such that

$$\forall n \leqslant \left\lceil\frac{T}{\Delta t}\right\rceil, \qquad \mathbb{P}\left(\left|\mathscr{G}^n - \sqrt{2\Delta t}\sum_{k=0}^{n-1} D(q^k)\right| \geqslant \frac{1}{2}\right) \leqslant C\Delta t^{3/2}. \tag{67}$$

Since, by (66),

$$\left|\sqrt{2\Delta t}\sum_{k=0}^{n-1} D(q^k)\right| \leqslant \widetilde{C}\Delta t,$$

it finally holds, for Δt sufficiently small,

$$\forall n \leqslant \left\lceil\frac{T}{\Delta t}\right\rceil, \qquad \mathbb{P}\left(\left|\widehat{\mathscr{G}}^n\right| \geqslant 1\right) \leqslant C\Delta t^{3/2} \leqslant \frac{1}{2}.$$

We next write, as in (64),

$$\mathbb{P}\left(q^n \in E \,\middle|\, q^0 = q\right) = \mathbb{P}\left(\widetilde{\mathscr{G}}^n \in E - q - \mathscr{F}^n - \widehat{\mathscr{G}}^n\right)$$

$$\geqslant \mathbb{P}\left(\widetilde{\mathscr{G}}^n \in E - q - \mathscr{F}^n - \widehat{\mathscr{G}}^n \,\middle|\, \left|\widehat{\mathscr{G}}^n\right| \leqslant 1\right)\mathbb{P}\left(\left|\widehat{\mathscr{G}}^n\right| \leqslant 1\right)$$

$$\geqslant \frac{1}{2}\mathbb{P}\left(\widetilde{\mathscr{G}}^n \in E - q - \mathscr{F}^n - \widehat{\mathscr{G}}^n \,\middle|\, \left|\widehat{\mathscr{G}}^n\right| \leqslant 1\right).$$

In view of the bounds on $\mathscr{F}^n$, there exists $R \in (0, +\infty)$ such that $\left|q + \mathscr{F}^n + \widehat{\mathscr{G}}^n\right| \leqslant R$ when $\left|\widehat{\mathscr{G}}^n\right| \leqslant 1$. Therefore,

$$\mathbb{P}\left(q^n \in E \,\middle|\, q^0 = q\right) \geqslant \frac{1}{2}\inf_{|Q| \leqslant R}\mathbb{P}\left(\widetilde{\mathscr{G}}^n \in E - Q\right). \tag{68}$$

In order to conclude the proof, we need to determine the law of $\widetilde{\mathscr{G}}^n$. When the Metropolis rule is used, $\widetilde{\mathscr{G}}^n$ simply is a Gaussian random variable of mean 0 and covariance matrix $2n\Delta t\,\mathrm{Id}_d$, and the desired conclusion follows by the same manipulations as the one performed below (64). The case of the Barker rule requires some additional work. Let us first introduce the random variable $\mathcal{N}_n$ which counts the number of times $U^k \leqslant a$ for $0 \leqslant k \leqslant n - 1$. Of course, $\mathcal{N}_n$ is a binomial law of parameters $1/2$ and n, hence its expectation is $\mathbb{E}(\mathcal{N}_n) = n/2$ while its variance is $\mathrm{Var}(\mathcal{N}_n) = n/4$. Therefore, by the Chebyshev inequality,

$$\mathbb{P}\left(\left|\mathcal{N}_n - \frac{n}{2}\right| \geqslant \eta\,\frac{\sqrt{n}}{2}\right) \leqslant \frac{1}{\eta^2},$$

so that

$$\mathbb{P}\left(\mathcal{N}_n \geqslant \frac{n}{4}\right) \geqslant 1 - \frac{4}{n}.$$

On the other hand, conditionally to $\mathcal{N}_n = m$, the random variable $\widetilde{\mathscr{G}}^n$ is a Gaussian random variable of mean 0 and covariance matrix $2m\Delta t\,\mathrm{Id}_d$. Therefore, for a given set $\widetilde{E} \subset \mathcal{M}$,

$$\mathbb{P}\left(\widetilde{\mathscr{G}}^n \in \widetilde{E}\right) \geqslant \mathbb{P}\left(\widetilde{\mathscr{G}}^n \in \widetilde{E} \,\middle|\, \mathcal{N}_n \geqslant \frac{n}{4}\right)\mathbb{P}\left(\mathcal{N}_n \geqslant \frac{n}{4}\right)$$

$$\geqslant \left(1 - \frac{4}{n}\right)\left(\frac{1}{4\pi n\Delta t}\right)^{d/2}\int_{\widetilde{E}} \exp\left(-\frac{|g|^2}{n\Delta t}\right)\,dg.$$

Together with (68), this allows to conclude, as at the end of the proof of Lemma 7, that (62) holds.

Remark 13 (Extension to the case of dynamics with multiplicative noise). *To extend the above proof to discretization of dynamics such as* (36), *the key point is to appropriately bound* $\widehat{\mathscr{G}}^n$ *since the rejections are encoded in this random variable. To this end, we note that the average drift* (65), *which seems to be of order* $\sqrt{\Delta t}$, *in fact is of order* Δt *in view of* (77)-(78) *and Lemma 10; while a bound similar to* (67) *holds with* $C\sqrt{\Delta t}$ *on the right-hand side since the variance of the martingale increments is of order* $\sqrt{\Delta t}$ *rather than* $\Delta t^{3/2}$.

4.5 Proof of Theorem 6

We follow the strategy of [12, Section 3.8] (as already done in [8, Section 5.4]) and write an approximation of $\mathcal{L}^{-1}$ using the discrete evolution operator $P_{\Delta t}$. We write the proof in the general case when

$$\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\psi = a_1\mathcal{L}\psi + a_2\Delta t\mathcal{L}^2\psi + \Delta t^\alpha\, r_{\psi,\Delta t}, \tag{69}$$

for $a_1 > 0$, $a_2 \geqslant 0$ and $1 < \alpha \leqslant 2$, and $r_{\psi,\Delta t}$ is uniformly bounded for Δt sufficiently small. The cases of interest are given by (35). In particular, $a_2 = a_1/2$ in all cases. Note that (with equalities in $\widetilde{L}^\infty(\mathcal{M})$),

$$(-\mathcal{L})^{-1}\,\psi = \left(\Delta t\sum_{n=0}^{+\infty} P_{\Delta t}^n\right)\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)\left(-\mathcal{L}^{-1}\right)\psi$$

$$= \left(\Delta t\sum_{n=0}^{+\infty} P_{\Delta t}^n\right)\left((a_1 + a_2\Delta t\mathcal{L})\,\psi + \Delta t^\alpha r_{\mathcal{L}^{-1}\psi,\Delta t}\right).$$

Since $\mathcal{L}^{-1}\psi$ still is a smooth function (by elliptic regularity), the remainder $r_{\mathcal{L}^{-1}\psi,\Delta t}$ is uniformly bounded in $L^\infty(\mathcal{M})$ by Lemma 3. Note also that since $(\mathrm{Id} - P_{\Delta t})\mathcal{L}^{-1}\psi$ and $\mathcal{L}\psi$ have vanishing averages with respect to μ, the remainder $r_{\mathcal{L}^{-1}\psi,\Delta t}$ has a vanishing average with respect to μ. Moreover, in view of (69),

$$\mathcal{L}\psi = -\frac{1}{a_1}\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\psi + \Delta t\,\widehat{r}_{\psi,\Delta t},$$

with a remainder uniformly bounded in Δt and with vanishing average with respect to μ, so that

$$\left(\Delta t\sum_{n=0}^{+\infty} P_{\Delta t}^n\right)\mathcal{L}\psi = \left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)^{-1}\mathcal{L}\psi = -\frac{1}{a_1}\psi + \Delta t\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)^{-1}\widehat{r}_{\psi,\Delta t}.$$

The above equalities show that

$$\int_{\mathcal{M}}\left(-\mathcal{L}^{-1}\psi\right)\varphi\,d\mu = a_1\Delta t\sum_{n=0}^{+\infty}\int_{\mathcal{M}}\left(P_{\Delta t}^n\psi\right)\varphi\,d\mu - \frac{a_2}{a_1}\Delta t\int_{\mathcal{M}}\psi\,\varphi\,d\mu$$

$$+ \Delta t^\alpha\int_{\mathcal{M}}\left[\left(\frac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)^{-1}\left(r_{\mathcal{L}^{-1}\psi,\Delta t} + \Delta t^{2-\alpha}\widehat{r}_{\psi,\Delta t}\right)\right]\varphi\,d\mu,$$

where the sum is convergent in view of (30). This gives the result, in view of the boundedness of the operator $\left(\dfrac{\mathrm{Id} - P_{\Delta t}}{\Delta t}\right)^{-1}$ (given by (31)).

4.6 Proof of Theorem 8

The proof follows the lines of the proof of [8, Theorem 3]. We write it in the more general case when $P_{\Delta t}$ satisfies (69). We need preliminary results on the average behavior of the increments. The first result is obtained by considering, for a given function f, the conditional increment $\overline{f}(q) = \mathbb{E}\left(f(q^1 - q^0)\,\middle|\,q^0 = q\right)$ and using the expansion of $P_{\Delta t}$ in powers of Δt. In order to state the result, we introduce the operator

$$(\mathcal{L}_Q^i f)(q) = (\mathcal{L}^i\tau_Q f)(q), \qquad \tau_a f(\delta) = f(\delta - a).$$

To be more explicit, $(\mathcal{L}_Q f)(q) = -\nabla V(q)\cdot\nabla f(q - Q) + \Delta f(q - Q)$, so that, in particular, $(\mathcal{L}_q f)(q) = -\nabla V(q)\cdot\nabla f(0) + \Delta f(0)$. The expression of $\left(\mathcal{L}_q^2 f\right)(q)$ is similarly obtained by writing the expression of $\mathcal{L}^2 f(q)$ and replacing the arguments of f by 0 everywhere.

Lemma 8. *Consider two smooth functions $f, g : \mathbb{R}^d \to \mathbb{R}$. Then,*

$$\overline{f}(q) = \mathbb{E}_{G,U}\Big(f\big(\delta_{\Delta t}(q, G, U)\big)\Big) = f(0) + a_1 \Delta t (\mathcal{L}_q f)(q) + a_2 \Delta t^2 \left(\mathcal{L}_q^2 f\right)(q) + \Delta t^{\alpha+1} r_{f,\Delta t}, \qquad (70)$$

and

$$\frac{1}{\Delta t}\Big(\mathbb{E}_{G,U}\Big(f\big(\delta_{\Delta t}(q, G, U)\big)g\big(\delta_{\Delta t}(q, G, U)\big)\Big) - \overline{f}(q)\overline{g}(q)\Big) = 2a_1 \nabla f(0) \cdot \nabla g(0) + \Delta t C_{f,g}(q) + \Delta t^\alpha r_{f,g,\Delta t}, \tag{71}$$

where

$$C_{f,g}(q) = a_2 \Big(\left[\mathcal{L}_q^2(fg)\right](q) - f(0)\left(\mathcal{L}_q^2 g\right)(q) - \left(\mathcal{L}_q^2 f\right)(q)g(0)\Big) - a_1^2(\mathcal{L}_q f)(q)(\mathcal{L}_q g)(q). \qquad (72)$$

The equality (71) is obtained by an application of (70) to the function fg. The second result on the average behavior of the increments is the following.

Lemma 9. *Set $\alpha = 3/2$, $\gamma = 2$ and $a = 1$ for schemes based on the Metropolis rule, while $\alpha = 2$, $\gamma = 3$ and $a = 1/2$ for schemes based on the Barker rule. For Δt sufficiently small,*

$$\overline{\delta}_{\Delta t}(q) = \mathbb{E}_{G,U}[\delta_{\Delta t}(q, G, U)] = -\Delta t\,(a_1 + a_2 \Delta t\,\mathcal{L})\,\nabla V(q) + \Delta t^{\alpha+1} r_{\Delta t,\delta}(q), \qquad (73)$$

where $r_{\Delta t,\delta}$ is uniformly bounded in $L^\infty(\mathcal{M})$ for Δt sufficiently small. The function $\overline{\delta}_{\Delta t}$ has average 0 with respect to μ, and the unique solution in $L^\infty(\mathcal{M})$ of the Poisson equation

$$(P_{\Delta t} - \mathrm{Id})\,N_{\Delta t} = \overline{\delta}_{\Delta t}, \qquad \int_{\mathcal{M}} N_{\Delta t}\,d\mu = 0,$$

can be expanded as

$$N_{\Delta t} = -\mathcal{L}^{-1}\nabla V + \Delta t^\alpha \widetilde{N}_\alpha + \Delta t^\gamma \widetilde{N}_{\gamma,\Delta t}, \qquad (74)$$

where $\widetilde{N}_\alpha$ is smooth and $\widetilde{N}_{\gamma,\Delta t}$ is uniformly bounded in $L^\infty(\mathcal{M})$ for Δt sufficiently small.

Proof. The expansion (73) is a direct consequence of Lemma 8 with the choice $f(\delta) = \delta$. The fact that $\overline{\delta}_{\Delta t}(q)$ has average 0 can be proved as in the proof of [8, Lemma 4]. From (69),

$$\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\left(\mathcal{L}^{-1}\nabla V\right) = a_1 \nabla V + a_2 \Delta t\,\mathcal{L}\nabla V + \Delta t^{\alpha+1} r_{\nabla V,\Delta t},$$

so that, in view of the equation satisfied by $N_{\Delta t}$ and the expansion of $\overline{\delta}_{\Delta t}$,

$$(P_{\Delta t} - \mathrm{Id})\left(N_{\Delta t} + \mathcal{L}^{-1}\nabla V\right) = \Delta t^{\alpha+1} r_{\nabla V,\Delta t}.$$

In view of (31), this shows that $N_{\Delta t} = -\mathcal{L}^{-1}\nabla V + \Delta t^\alpha \widetilde{N}_{\alpha,\Delta t}$. We need at this stage to obtain weak type expansions such as (35) at higher order. More precisely,

$$\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\psi = a\left(\mathcal{L}\psi + \frac{\Delta t}{2}\mathcal{L}^2\psi\right) + \Delta t^\alpha S\psi + \Delta t^\gamma\,r_{\psi,\Delta t},$$

where S is some differential operator of finite order, and the remainder is uniformly bounded for Δt sufficiently small. The proof is a slight extension of the proof of Lemma 3 performed in Section 4.3 and is therefore omitted. The important point to note is that $\gamma = 3$ when the Barker rule is used since terms with fractional powers of Δt always come with odd powers of G; while in contrast $\gamma = 2$ in the Metropolis case since all terms $\Delta t^{k/2}$ contribute for $k \geqslant 1$. By the same computations as above,

$$(P_{\Delta t} - \mathrm{Id})\left(N_{\Delta t} + \mathcal{L}^{-1}\nabla V\right) = \Delta t^{\alpha+1} S\mathcal{L}^{-1}\nabla V + \Delta t^\gamma r_{\nabla V,\Delta t}.$$

This allows us to identify $\widetilde{N}_\alpha = \mathcal{L}^{-1}S\mathcal{L}^{-1}\nabla V$ since

$$\left(\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\right)^{-1} S\mathcal{L}^{-1}\nabla V = \mathcal{L}^{-1}S\mathcal{L}^{-1}\nabla V + \Delta t\,\widehat{r}_{\nabla V,\Delta t}.$$

The latter equality can be checked by applying $P_{\Delta t} - \mathrm{Id}$ on both sides and using (31). $\qquad\square$

We can now turn to the proof of Theorem 8.

Proof of Theorem 8. We rewrite the increment $\delta_{\Delta t}(q^m, G^m, U^m)$ as the sum of a discrete martingale $\delta_{\Delta t}(q^m, G^m, U^m) - \overline{\delta}_{\Delta t}(q^m)$ and the average increment $\overline{\delta}_{\Delta t}(q^m)$. We also use Lemma 9 to rewrite $\overline{\delta}_{\Delta t}(q^m)$ as

$$\overline{\delta}_{\Delta t}(q^m) = P_{\Delta t}N_{\Delta t}(q^m) - N_{\Delta t}(q^m) = \Big(P_{\Delta t}N_{\Delta t}(q^m) - N_{\Delta t}(q^{m+1})\Big) + \Big(N_{\Delta t}(q^{m+1}) - N_{\Delta t}(q^m)\Big).$$

Therefore,

$$\sum_{m=0}^{n-1} \delta_{\Delta t}(q^m, G^m, U^m) = N_{\Delta t}(q^n) - N_{\Delta t}(q^0) + \sum_{m=0}^{n-1} M_{\Delta t}^m,$$

with

$$M_{\Delta t}^m = \delta_{\Delta t}(q^m, G^m, U^m) - N_{\Delta t}(q^m + \delta_{\Delta t}(q^m, G^m, U^m)) - \overline{\delta}_{\Delta t}(q^m) + \overline{N}_{\Delta t}(q^m),$$

where $\overline{N}_{\Delta t}(q) = (P_{\Delta t}N_{\Delta t})(q) = \mathbb{E}_{G,U}[N_{\Delta t}(q + \delta_{\Delta t}(q, G, U))]$. Note that $(M_{\Delta t}^m)_{m \geqslant 0}$ are stationary, independent martingale increments when $q^0 \sim \mu$. In view of Lemma 9, $N_{\Delta t} \in L^\infty(\mathcal{M})$, so that, for a given $\xi \in \mathbb{R}^d$,

$$\xi^T \mathscr{D}_{\Delta t} \xi = \frac{1}{\Delta t} \mathbb{E}\left[\left(\xi^T M_{\Delta t}^0\right)^2\right] = \frac{1}{\Delta t} \int_{\mathcal{M}} \left(\mathbb{E}_{G,U}\left(\left[B_{\Delta t,\xi}(q, \delta_{\Delta t}(q, G, U))\right]^2\right) - \left(\overline{B}_{\Delta t,\xi}(q)\right)^2\right) \mu(dq), \quad (75)$$

with $B_{\Delta t,\xi}(q, \delta) = \xi^T(\delta - N_{\Delta t}(q + \delta))$ and where the expectation in the first equality is with respect to G, U and $q^0 \sim \mu$. We now use (71) to compute the right-hand side. Note first that, in view of (74) (setting $f(\delta) = g(\delta) = B_{\Delta t,\xi}(q, \delta)$, the first argument q in $B_{\Delta t,\xi}$ being a parameter),

$$\xi^T \mathscr{D}_{\Delta t} \xi = \frac{1}{\Delta t} \int_{\mathcal{M}} \left(\mathbb{E}_{G,U}\left(\left[B_{0,\xi}(q, \delta_{\Delta t}(q, G, U))\right]^2\right) - \left(\overline{B}_{0,\xi}(q)\right)^2\right) \mu(dq) + \Delta t^\alpha \xi^T \widetilde{\mathscr{D}}_{\Delta t} \xi$$

$$= 2a_1 \int_{\mathcal{M}} \left(|\xi|^2 - 2\nabla(\xi^T N_0) \cdot \xi + \left|\nabla\left(\xi^T N_0\right)\right|^2\right) d\mu + \Delta t \int_{\mathcal{M}} C_{B_{0,\xi},B_{0,\xi}} d\mu + \Delta t^\alpha \xi^T \widehat{\mathscr{D}}_{\Delta t} \xi,$$

where $C_{f,g}$ is defined in (72). Next, using the definition of $f_0(\delta) = B_{0,\xi}(q, \delta) = \xi^T \delta - \mathcal{L}^{-1}\left(\xi^T \nabla V\right)(q + \delta)$ (where again q is a parameter), a simple computation shows that $\nabla f_0(\delta) = \xi - \left[\nabla \mathcal{L}^{-1}(\xi^T \nabla V)\right](q + \delta)$ and $\Delta f_0(\delta) = -\left[\Delta \mathcal{L}^{-1}(\xi^T \nabla V)\right](q + \delta)$. Therefore, $(\mathcal{L}_Q f_0)(q) = -\xi^T \nabla V(q) + \nabla V(q)^T \nabla(\xi^T N_0)(2q - Q) + \Delta(\xi^T N_0)(2q - Q)$ so that $(\mathcal{L}_q f_0)(q) = 0$ in view of the definition of N_0. Using the identity

$$\mathcal{L}^2(\varphi\psi) = 2(\mathcal{L}\varphi)(\mathcal{L}\psi) + \varphi\mathcal{L}^2\psi + (\mathcal{L}^2\varphi)\psi + 2\nabla\varphi \cdot \nabla(\mathcal{L}\psi) + 2\nabla(\mathcal{L}\varphi) \cdot \nabla\psi + 2\mathcal{L}(\nabla\varphi \cdot \nabla\psi),$$

obtained by iterating $\mathcal{L}(\varphi\psi) = \varphi\mathcal{L}\psi + (\mathcal{L}\varphi)\psi + 2\nabla\varphi \cdot \nabla\psi$, we also compute

$$C_{B_{0,\xi},B_{0,\xi}}(q) = 2\mathcal{L}\left(\left[\xi - \nabla(\xi^T N_0)\right]^2\right),$$

which has average 0 with respect to μ. In conclusion,

$$\xi^T \mathscr{D}_{\Delta t} \xi = 2a_1 \int_{\mathcal{M}} \left(|\xi|^2 - 2\nabla(\xi^T N_0) \cdot \xi + \left|\nabla\left(\xi^T N_0\right)\right|^2\right) d\mu + \Delta t^\alpha \xi^T \widehat{\mathscr{D}}_{\Delta t} \xi.$$

The result is finally obtained by manipulations similar to the ones used to establish [8, Eq. (32)]. $\qquad\square$

4.7   Proof of Lemma 4

The result crucially relies on the expansion in powers of Δt of $\alpha_{\Delta t}(q, \Phi_{\Delta t}(q, G))$ (defined in (38)), where $\Phi_{\Delta t}(q^n, G^n)$ encodes the proposal (37):

$$\Phi_{\Delta t}(q, G) = q + \Delta t F(q) + \sqrt{2\Delta t}B(q)G.$$

For notational convenience, we introduced the symmetric, definite, positive matrix $B(q) = M^{1/2}(q)$. We also write remainder terms as $\mathrm{O}(\Delta t^\delta)$. The equality $c(q, G) = \mathrm{O}(\Delta t^\delta)$ should be understood as: there exists $K_c > 0$ and $p_c \in \mathbb{N}$ such that

$$|c(q, G)| \leqslant K_c \Delta t^\delta \left(1 + |G|^{p_c}\right).$$

In particular, $\mathbb{E}_G\,|c(q,G)|^r \leqslant C_r\Delta t^{\delta r}$ for any $r \geqslant 0$.

Let us now evaluate the various terms in $\alpha_{\Delta t}(q,\Phi_{\Delta t}(q,G))$. First,

$$V(\Phi_{\Delta t}(q,G)) - V(q) = \sqrt{2\Delta t}\nabla V(q)^T B(q)G + \mathrm{O}(\Delta t).$$

Consider next the terms corresponding to $\det M$. Since

$$M(\Phi_{\Delta t}(q,G)) = M(q) + \sqrt{2\Delta t}DM(q)\cdot\big(B(q)G\big) + \mathrm{O}(\Delta t),$$

it holds

$$\det M(\Phi_{\Delta t}(q,G)) = \Big(\det M(q)\Big)\det\Big(\mathrm{Id} + \sqrt{2\Delta t}M^{-1}(q)\big[DM(q)\cdot(B(q)G)\big] + \mathrm{O}(\Delta t)\Big)$$
$$= \Big(\det M(q)\Big)\Big(1 + \sqrt{2\Delta t}\,\mathrm{Tr}\Big(M^{-1}(q)\big[DM(q)\cdot(B(q)G)\big]\Big) + \mathrm{O}(\Delta t)\Big),$$

so that

$$\frac{1}{2}\Big(\ln\big(\det M(\Phi_{\Delta t}(q,G))\big) - \ln\big(\det M(q)\big)\Big) = \frac{1}{2}\sqrt{2\Delta t}\,\mathrm{Tr}\Big(M^{-1}(q)\big[DM(q)\cdot(B(q)G)\big]\Big) + \mathrm{O}(\Delta t).$$

We finally turn to the remaining term, which, using the short-hand notation $q' = \Phi_{\Delta t}(q,G)$, we decompose as

$$|q-q'-\Delta tF(q')|^2_{M(q')} - |q'-q-\Delta tF(q)|^2_{M(q)} = \Big(|q-q'-\Delta tF(q')|^2_{M(q)} - |q'-q-\Delta tF(q)|^2_{M(q)}\Big)$$
$$+ \Big(|q-q'-\Delta tF(q')|^2_{M(q')} - |q-q'-\Delta tF(q')|^2_{M(q)}\Big). \tag{76}$$

The first term on the right-hand side of (76) is the difference between two vectors in the same norm, while the second term is the variation of the norm a given vector when the matrix inducing the scalar product changes. We rely on the expansion

$$q - q' - \Delta tF(q') = q - \Phi_{\Delta t}(q,G) - \Delta tF(\Phi_{\Delta t}(q,G))$$
$$= -\sqrt{2\Delta t}B(q)G - \Delta t\Big(F(q) + F(\Phi_{\Delta t}(q,G))\Big)$$
$$= -\sqrt{2\Delta t}B(q)G - 2\Delta tF(q) + \mathrm{O}(\Delta t^{3/2}),$$

so that

$$|q-q'-\Delta tF(q')|^2_{M(q)} - |q'-q-\Delta tF(q)|^2_{M(q)} = 4\sqrt{2}\,\Delta t^{3/2}G^T B(q)M^{-1}(q)F(q) + \mathrm{O}(\Delta t^2).$$

For the second term, we use $M(q+x)^{-1} = M(q)^{-1} - M^{-1}(q)\big[DM(q)\cdot x\big]M^{-1}(q) + \mathrm{O}(|x|^2)$:

$$|q-q'-\Delta tF(q')|^2_{M(q')} - |q-q'-\Delta tF(q')|^2_{M(q)}$$
$$= \big(q-q'-\Delta tF(q')\big)^T\big(M(q')^{-1} - M(q)^{-1}\big)\big(q-q'-\Delta tF(q')\big)$$
$$= -(2\Delta t)^{3/2}G^T B(q)M(q)^{-1}\Big(DM(q)\cdot B(q)G\Big)M(q)^{-1}B(q)G + \mathrm{O}(\Delta t).$$

The combination of all terms gives

$$\alpha_{\Delta t}(q,\Phi_{\Delta t}(q,G)) = \sqrt{2\Delta t}\,\xi_{1/2}(q,G) + \mathrm{O}(\Delta t),$$

with

$$\xi_{1/2}(q,G) = \nabla V(q)^T B(q)G + \frac{1}{2}\,\mathrm{Tr}\Big(M^{-1}(q)\big[DM(q)\cdot(B(q)G)\big]\Big) + G^T B(q)M^{-1}(q)F(q)$$
$$-\frac{1}{2}G^T B(q)M^{-1}(q)\Big(DM(q)\cdot B(q)G\Big)M^{-1}(q)B(q)G$$
$$= (\mathrm{div}\,M)^T M^{-1/2}G + \frac{1}{2}\,\mathrm{Tr}\Big(M^{-1}(q)\big[DM(q)\cdot(B(q)G)\big]\Big)$$
$$-\frac{1}{2}G^T M^{-1/2}(q)\Big(DM(q)\cdot B(q)G\Big)M^{-1/2}(q)G.$$

Therefore,
$$A_{\Delta t}^{\mathrm{MH}}\big(q, \Phi_{\Delta t}(q, G)\big) = 1 - \sqrt{2\Delta t}\, \min\big(0, \xi_{1/2}(q, G)\big) + \mathrm{O}(\Delta t), \tag{77}$$
while, in view of (54),
$$A_{\Delta t}^{\mathrm{Barker}}\big(q, \Phi_{\Delta t}(q, G)\big) = \frac{1}{2} - \frac{1}{4}\sqrt{2\Delta t}\,\xi_{1/2}(q, G) + \Delta t\,\xi_2(q, G) + \mathrm{O}(\Delta t^{3/2}). \tag{78}$$

Since $\xi_{1/2}(q, G)$ is odd in G, its expectation with respect to G vanishes. The term $\xi_2(q, G)$ involves only even powers of G.

To conclude the proof, we write

$$\frac{P_{\Delta t} - \mathrm{Id}}{\Delta t}\psi(q) = \mathbb{E}_G\left[A_{\Delta t}\big(q, \Phi_{\Delta t}(q, G)\big)\frac{\psi(\Phi_{\Delta t}(q, G)) - \psi(q)}{\Delta t}\right]$$

$$= a\,\mathbb{E}_G\left[\frac{\psi(\Phi_{\Delta t}(q, G)) - \psi(q)}{\Delta t}\right] + \mathbb{E}_G\left[\big(A_{\Delta t}\big(q, \Phi_{\Delta t}(q, G)\big) - a\big)\frac{\psi(\Phi_{\Delta t}(q, G)) - \psi(q)}{\Delta t}\right]$$

$$= a\,\mathcal{L}\psi(q) + \mathbb{E}_G\left[\frac{A_{\Delta t}\big(q, \Phi_{\Delta t}(q, G)\big) - a}{\sqrt{\Delta t}}\,\frac{\psi(\Phi_{\Delta t}(q, G)) - \psi(q)}{\sqrt{\Delta t}}\right].$$

In the Metropolis case, $a = 1$ and, by the symmetry property $\xi_{1/2}(q, G)G = -\xi_{1/2}(q, -G)G$ (similar to (60)),

$$\mathbb{E}_G\left[\frac{A_{\Delta t}^{\mathrm{MH}}\big(q, \Phi_{\Delta t}(q, G)\big) - a}{\sqrt{\Delta t}}\,\frac{\psi(\Phi_{\Delta t}(q, G)) - \psi(q)}{\sqrt{\Delta t}}\right] = -2\mathbb{E}_G\left[\min\big(0, \xi_{1/2}(q, G)\big)\,G^T B(q)\nabla\psi(q)\right] + \mathrm{O}(\sqrt{\Delta t})$$

$$= -\mathbb{E}_G\left[\xi_{1/2}(q, G)\,G^T B(q)\nabla\psi(q)\right] + \mathrm{O}\left(\sqrt{\Delta t}\right);$$

while, in the Barker case, $a = 1/2$ and

$$\mathbb{E}_G\left[\frac{A_{\Delta t}^{\mathrm{Barker}}\big(q, \Phi_{\Delta t}(q, G)\big) - a}{\sqrt{\Delta t}}\,\frac{\psi(\Phi_{\Delta t}(q, G)) - \psi(q)}{\sqrt{\Delta t}}\right] = -\frac{1}{2}\mathbb{E}_G\left[\xi_{1/2}(q, G)\,G^T B(q)\nabla\psi(q)\right] + \mathrm{O}\left(\Delta t\right).$$

Let us emphasize that the remainder indeed is of order Δt and not $\sqrt{\Delta t}$ since $\xi_{1/2}(q, G)$ involves only odd powers of G while $\xi_2(q, G)$ and the Δt term in the expansion of $\psi(\Phi_{\Delta t}(q, G)) - \psi(q)$ in powers of Δt involve only even powers of G. The claimed result now follows from the following lemma.

Lemma 10. *For any $v \in \mathbb{R}^d$, it holds $\mathbb{E}_G\left[\xi_{1/2}(q, G)\left(G^T v\right)\right] = 0$.*

Proof. Note first that, for a given vector $b \in \mathbb{R}^d$ and a given tensor $\mathcal{A}$ of order 3,

$$\mathbb{E}_G\left[\left(G^T b\right)\left(G^T v\right)\right] = b^T v, \qquad \mathbb{E}_G\left[\left(\mathcal{A} : G^{\otimes 3}\right)\left(G^T v\right)\right] = \sum_{i,j=1}^d v_i\left(\mathcal{A}_{jji} + \mathcal{A}_{jij} + \mathcal{A}_{ijj}\right).$$

In view of the expression of $\xi_{1/2}$, we introduce

$$\mathcal{A} : G^{\otimes 3} = G^T M^{-1/2}(q)\Big(DM(q)\cdot B(q)G\Big)M^{-1/2}(q)G,$$

and

$$b^T G = 2(\mathrm{div}M)^T M^{-1/2}G + \mathrm{Tr}\left(M^{-1}(q)\left[DM(q)\cdot\big(B(q)G\big)\right]\right).$$

Recall that M, M^{-1} and B are symmetric. The components of $\mathcal{A}$ and b respectively read

$$\mathcal{A}_{ijk} = \sum_{r,s,t=1}^d \left[M^{-1/2}\right]_{r,i}\left[M^{-1/2}\right]_{s,j}\left(\partial_{q_t} M_{rs}\right)B_{tk},$$

and

$$b_i = 2\sum_{r=1}^d \left[M^{-1/2}\right]_{i,r}\mathrm{div}(M_r) + \sum_{r,s,t=1}^d \left[M^{-1}\right]_{r,s}\left(\partial_{q_t} M_{rs}\right)B_{ti}.$$

34

Now, in view of the equality

$$\sum_{j=1}^{d} \left[M^{-1/2} \right]_{s,j} B_{tj} = \left[M^{-1/2} B \right]_{s,t} = \delta_{s,t},$$

it holds

$$\sum_{j=1}^{d} \mathcal{A}_{ijj} + \mathcal{A}_{jij} = 2 \sum_{r,s}^{d} \left[M^{-1/2} \right]_{r,i} (\partial_{q_t} M_{rt}) = 2 \sum_{r=1}^{d} \left[M^{-1/2} \right]_{i,r} \operatorname{div}(M_r).$$

Since

$$\sum_{j=1}^{d} \left[M^{-1/2} \right]_{r,j} \left[M^{-1/2} \right]_{s,j} = \left[M^{-1} \right]_{r,s},$$

we also have

$$\sum_{j=1}^{d} \mathcal{A}_{jji} = \sum_{r,s,t=1}^{d} \left[M^{-1} \right]_{r,s} (\partial_{q_t} M_{rs}) B_{ti}.$$

Therefore,

$$\sum_{j=1}^{d} \mathcal{A}_{ijj} + \mathcal{A}_{jij} + \mathcal{A}_{jji} = b_i,$$

which shows that $\mathbb{E}_G \left[\left(\mathcal{A} : G^{\otimes 3} - G^T b \right) \left(G^T v \right) \right] = 0$ and gives the expected result. $\qquad\square$

Acknowledgements

References

[1] A. Abdulle, D. Cohen, G. Vilmart, and K.C. Zygalakis. High weak order methods for stochastic differential equations based on modified equations. *SIAM J. Sci. Comput.*, 34(3):1800–1823, 2012.

[2] J. A. Barker. Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. *Aust. J. Phys.*, 18:119–133, 1965.

[3] N. Bou-Rabee, A. Donev, and E. Vanden-Eijnden. Metropolis integration schemes for self-adjoint diffusions. *Multiscale Model. Sim.*, 12(2):781–831, 2014.

[4] N. Bou-Rabee and M. Hairer. Nonasymptotic mixing of the MALA algorithm. *IMA J. Numer. Anal.*, 33:80–110, 2013.

[5] N. Bou-Rabee and E. Vanden-Eijnden. Pathwise accuracy and ergodicity of metropolized integrators for SDEs. *Commun. Pure Appl. Math.*, 63(5):655–696, 2009.

[6] N. Bou-Rabee and E. Vanden-Eijnden. A patch that imparts unconditional stability to explicit integrators for Langevin-like equations. *J. Comput. Phys.*, 231:2565–2580, 2012.

[7] A. Durmus, G. Roberts, G. Vilmart, and K. Zygalakis. In preparation.

[8] M. Fathi, A.-A. Homman, and G. Stoltz. Error analysis of the transport properties of Metropolized schemes. *ESAIM Proc.*, 2015. In press.

[9] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, 2006.

[10] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer, 2011.

[11] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[12] B. Leimkuhler, Ch. Matthews, and G. Stoltz. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA J. Numer. Anal.*, 2015. In press.

[13] T. Lelièvre, M. Rousset, and G. Stoltz. *Free-energy Computations: A Mathematical Perspective*. Imperial College Press, 2010.

[14] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Proc. Appl.*, 101(2):185–232, 2002.

[15] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1091, 1953.

[16] G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Scientific Computation. Springer, 2004.

[17] P. Peskun. Optimum Monte Carlo sampling using Markov chains. *Biometrika*, 60:607–612, 1973.

[18] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[19] P. J. Rossky, J. D. Doll, and H. L. Friedman. Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys.*, 69(10):4628–4633, 1978.

[20] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1987.