Asymptotic analysis for the generalized Langevin equation

2011 Nonlinearity 24 1629

(http://iopscience.iop.org/0951-7715/24/5/013)

IOP PUBLISHING NONLINEARITY

Nonlinearity **24** (2011) 1629–1653 doi:10.1088/0951-7715/24/5/013

Asymptotic analysis for the generalized Langevin equation

M Ottobre and G A Pavliotis

Department of Mathematics, Imperial College London, London SW7 2AZ, UK

Received 24 September 2010, in final form 27 January 2011
Published 15 April 2011
Online at stacks.iop.org/Non/24/1629

Recommended by J-P Eckmann

Abstract

Various qualitative properties of solutions to the generalized Langevin equation (GLE) in a periodic or a confining potential are studied in this paper. We consider a class of quasi-Markovian GLEs, similar to the model that was introduced in Eckmann J-P *et al* 1999 *Commun. Math. Phys.* **201** 657–97. Ergodicity, exponentially fast convergence to equilibrium, short time asymptotics, a homogenization theorem (invariance principle) and the white noise limit are studied. Our proofs are based on a careful analysis of a hypoelliptic operator which is the generator of an auxiliary Markov process. Systematic use of the recently developed theory of hypocoercivity (Villani C 2009 *Mem. Am. Math. Soc.* **202** iv, 141) is made.

Mathematics Subject Classification: 82C31, 60H10, 35K10, 60J60, 60F17

1. Introduction

In this paper, we study various qualitative properties of solutions to the generalized Langevin equation (GLE) in $\mathbb{R}^d$

$$\ddot{q} = -\nabla V(q) - \int_0^t \gamma(t-s)\dot{q}(s)\,\mathrm{d}s + F(t), \tag{1}$$

where $V(q)$ is a smooth potential (confining or periodic), $F(t)$ a mean zero stationary Gaussian process with autocorrelation function $\gamma(t)$, in accordance to the fluctuation–dissipation theorem

$$\langle F(t) \otimes F(s) \rangle = \beta^{-1}\gamma(t-s)I. \tag{2}$$

Here β stands for the inverse temperature and I for the identity matrix. The dots in (1) denote differentiation with respect to time. The GLE equation (1), together with the fluctuation–dissipation theorem (2) appears in various applications such as surface diffusion [1] and polymer dynamics [46]. It also serves as one of the standard models of non-equilibrium statistical mechanics, describing the dynamics of a 'small' Hamiltonian system (the distinguished particle) coupled to one or more heat baths which are modelled as linear wave equations with initial conditions which are distributed according to appropriate Gibbs measures [44]. In this class of models the coupling between the distinguished particle and the heat bath is taken to be linear and is governed by a coupling function $\rho(x)$. The full Hamiltonian of the 'particle + heat bath' model is

$$H(q, p, \phi, \pi) = H_{DP}(p, q) + \mathcal{H}(\phi, \pi) + \lambda q \int \rho(x) \partial_q \phi(x)\, \mathrm{d}x, \tag{3}$$

where $H_{DP}(q, p)$ denotes the Hamiltonian of the distinguished particle whose position and momentum are denoted by q and p, respectively, and $\mathcal{H}(\phi, \pi)$ is the Hamiltonian density of the wave equation where ϕ and π are the canonically conjugate field variables. The linear coupling in (3) is motivated by the dipole approximation from classical electrodynamics. By integrating out the heat bath variables and using our assumptions on the initial conditions we obtain (1), together with (2). The memory kernel $\gamma(t)$ in (1) is given by the coupling function through the formula

$$\gamma(t) = \int |\widehat{\rho}(k)|^2 \mathrm{e}^{\mathrm{i}kt}\, \mathrm{d}k, \tag{4}$$

where $\widehat{\rho}(k)$ denotes the Fourier transform of $\rho(x)$ [24, 44].

The GLE has also attracted attention in recent years in the context of mode reduction and coarse-graining for high-dimensional dynamical systems [14]. One of the models that has been studied extensively within the framework of mode elimination is the Kac–Zwanzig model [13, 49] and its variants [3, 19, 28, 30]. In this model, the heat bath is modelled as a finite-dimensional system of N harmonic oscillators with random frequencies and random initial conditions distributed according to a Gibbs distribution at inverse temperature β. The heat bath can be coupled either linearly or nonlinearly with the distinguished particle [29]. Just as with model (3), we can integrate out the heat bath variables explicitly. Passing then to the thermodynamic limit $N \to +\infty$, we obtain the GLE (1). The form of the memory kernel $\gamma(t)$ depends on the choice of the distribution of the spring constants of the harmonic oscillators in the heat bath [14]. The Kac–Zwanzig model and its variants have proved to be very useful for testing various methodologies and techniques such as transition state theory [2, 20].

The GLE (1) is a stochastic integrodifferential equation which is equivalent to the original infinite-dimensional Hamiltonian system with random initial conditions. The infinite-dimensionality of the original Hamiltonian dynamics with random initial conditions (or, equivalently, the non-Markovianity of the finite-dimensional stochastic dynamics (1)) renders the analysis of this dynamical system very difficult. This problem was studied in detail by Jaksic and Pillet in a series of papers [24–26]. In these works, existence and uniqueness of solutions as well as the ergodic properties of (1) were studied in detail. In particular, it was shown that the process $\{q, p = \dot{q}\}$ is mixing with respect to the measure

$$\nu_\beta(\mathrm{d}q\mathrm{d}p) = \frac{1}{\mathcal{Z}_\beta} \mathrm{e}^{-\beta H_{DP}(q,p)}\, \mathrm{d}q\mathrm{d}p,$$

where $\mathcal{Z}_\beta$ is the normalization constant. To the best of our knowledge, no information concerning the rate of convergence to equilibrium for the non-Markovian dynamics (1) is known for general classes of memory kernels. Ergodic theory for a quite general class of non-Markovian processes has been developed recently, see [16] and the references therein.

A class of memory kernels for which more detailed information on the long time asymptotics of the GLE (1) can be obtained was considered by Eckmann, Hairer, Pillet and Rey-Bellet in a series of papers [9–11, 45]. Based on a generalization of Doob's theorem on stationary, Markovian, Gaussian processes [8], it was observed in these works that when the memory kernel $\gamma(t)$ has a rational spectral density, then the GLE (1) is equivalent to a *finite-dimensional* Markovian system. This system is obtained by adding a finite number of additional degrees of freedom which account for the memory in the system. These auxiliary variables satisfy linear stochastic differential equations. As an example, we mention the case where $\widehat{\rho}(k)$ in (4) can be written as

$$|\widehat{\rho}(k)|^2 = \frac{1}{|p(k)|^2},$$

where $p(k) = \sum_{m=1}^{M} c_m(-\mathrm{i}k)^m$ is a polynomial with real coefficients and roots in the upper half plane. Then the Gaussian process with spectral density $|\widehat{\rho}(k)|^2$ is the solution of the SDE

$$p\left(-\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\right)x(t) = \frac{\mathrm{d}W}{\mathrm{d}t},$$

where W is a standard one-dimensional Brownian motion (see [44, proposition 2.3]). A related finite-dimensional approximation of the infinite-dimensional dynamics (1) was introduced by Mori in [36], see also [15] and the references therein. Mori's technique is based on a continued fraction expansion of the Laplace transform of the memory function $\gamma(t)$.

Motivated by the above, in this paper we will consider finite-dimensional approximations of the GLE. The general form of the Markovian approximation of (1) when $d = 1$ can be written as [28]

$$\dot{Q}_m(t) = P_m(t), \qquad Q_m(0) = q(0), \tag{5a}$$
$$\dot{P}_m(t) = -\partial_q V(Q_m(t)) + \lambda^{\mathrm{T}} z(t), \qquad P_m(0) = p(0), \tag{5b}$$
$$\dot{z}(t) = -P_m(t)\lambda - \mathcal{A}\, z(t) + \mathcal{G}\dot{W}(t), \qquad z(0) \sim \mathcal{N}(0, I), \tag{5c}$$

where $z : \mathbb{R}^+ \mapsto \mathbb{R}^m$, $\lambda \in \mathbb{R}^m$, $\mathcal{A}, \mathcal{G} \in \mathbb{R}^{m\times m}$ and $W(t)$ is an m-dimensional Brownian motion. The fluctuation–dissipation theorem, which takes the form $\mathcal{G}\mathcal{G}^{\mathrm{T}} = \beta^{-1}(\mathcal{A} + \mathcal{A}^{\mathrm{T}})$, is assumed to be satisfied.

In this paper we will consider (5) with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ and $\mathcal{A}$ diagonal with $\mathcal{A}_{ii} = \alpha_i > 0$. This amounts to approximating the memory kernel by a sum of exponentials,

$$\gamma_m(t) = \sum_{i=1}^{m} \lambda_i^2 \mathrm{e}^{-\alpha_i|t|}. \tag{6}$$

It is expected that the results proved in this paper are also valid in the more general case (5). As remarked in [28], when $\mathcal{A}$ is invertible, the more standard Mori approximation [36] is equivalent to (6) after an appropriate orthogonal transformation.

For this particular choice of λ and $\mathcal{A}$ the SDEs (5) become (we drop the subscripts m for notational simplicity)

$$\dot{Q}(t) = P(t), \qquad Q(0) = q(0), \tag{7a}$$
$$\dot{P}(t) = -\nabla_q V(Q(t)) + \sum_{j=1}^{m} \lambda_j z_j(t), \qquad P(0) = p(0), \tag{7b}$$
$$\dot{z}_j(t) = -\lambda_j P(t) - \alpha_j z_j(t) + \sqrt{2\alpha_j\beta^{-1}}\dot{W}_j, \qquad z_j(0) \sim \mathcal{N}(0, \beta^{-1}) \tag{7c}$$

for $j = 1, \ldots, m$. The process $\{Q(t),\ P(t),\ z(t)\} \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{md}$ is Markovian with generator $-\mathcal{L}$ given by

$$-\mathcal{L} = p\cdot\nabla_q - \nabla_q V(q)\cdot\nabla_p + \sum_{j=1}^{m}\lambda_j z_j(t)\cdot\nabla_p + \sum_{j=1}^{m}\left(-\lambda_j p\cdot\nabla_{z_j} - \alpha_j z_j\cdot\nabla_{z_j} + \alpha_j\beta^{-1}\Delta_{z_j}\right). \tag{8}$$

This operator can be written in 'sum of squares' form

$$\mathcal{L} = B + \sum_{i=1}^{m} \sum_{j=1}^{d} A_{ij}^{*} A_{ij},$$

where $A_{ij} = -\sqrt{\beta^{-1}\alpha_i}\partial_{z_{i_j}}$, $A_{ij}^{*} = -\sqrt{\beta\alpha_i} z_{i_j} + \sqrt{\beta^{-1}\alpha_i}\partial_{z_{i_j}}$ [1] and

$$B = -p \cdot \nabla_q + \nabla_q V \cdot \nabla_p - \sum_{j=1}^{m} \lambda_j \left(z_j \cdot \nabla_p - p \cdot \nabla_{z_j}\right)$$

(for more details on this notation we refer the reader to section 2.1). This is a degenerate second-order elliptic differential operator of hypoelliptic type [23]. Convergence to equilibrium for models of the form (7) has been studied using functional analytic techniques [9, 11]. Similar results have also been proved using Markov chain techniques [34, 45]. In this paper we present an alternative proof of exponentially fast convergence to equilibrium in relative entropy using the recently developed theory of hypocoercivity [48].

The main objective of this paper is the rigorous analysis of the Markovian approximation (7) to the GLE (1). Our main results can be summarized as follows.

1. We prove ergodicity and exponentially fast convergence to equilibrium for (7), theorems 2.1, 2.2 and 2.3.
2. We obtain sharp estimates on derivatives of the Markov semigroup associated with the SDE (7), theorem 2.4.
3. We prove a homogenization theorem (invariance principle) when the potential $V(q)$ in (7) is periodic and we obtain estimates on the diffusion coefficient, theorem 2.5. In order to prove these results we prove compactness of the resolvent of the generator of the SDE (7), proposition 2.1.
4. We study the white noise limit of the GLE (1), i.e. the limit as the noise $F(t)$ in (1) (in the Markovian approximation (7)) converges to a white noise process. We show that in this limit the solution of (7) converges strongly to the solution of the Langevin equation

$$\ddot{q} = -\nabla V(q) - \gamma \dot{q} + \sqrt{2\gamma\beta^{-1}}\dot{W} \tag{9}$$

and we obtain a formula for the friction coefficient γ in terms of the coefficients $\{\lambda_j, \alpha_j\}_{j=1}^{m}$, theorem 2.6.

The rest of the paper is organized as follows. In section 2 we state our main results and we introduce the notation that we will be using. In section 3 we prove exponentially fast convergence to equilibrium. In section 4 we prove estimates on the derivatives of the Markov semigroup generated by $-\mathcal{L}$ defined in (8). In section 5 we prove the homogenization theorem and the compactness of the resolvent of $\mathcal{L}$. In section 6 we study the white noise limit. For the reader's convenience, background material on the theory of hypocoercivity is summarized in appendix A. Finally, the proof of geometric ergodicity of the process (7) using Markov chain techniques is presented in appendix B.

2. Statement of main results

We will use the notation $X := \mathbb{T}^d \times \mathbb{R}^d \times \mathbb{R}^{dm}$ and $Y := \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{dm}$. We will also denote the process $\{q(t), p(t), z(t)\}$ by $\boldsymbol{x}(t)$. When we study the dynamics (7) in X the potential

[1] A_{ij}^{*} is the adjoint of A_{ij} in the L^2 space weighted by the invariant measure of the system. See section 2.1.

$V(q)$ is periodic, whereas when $\boldsymbol{x}(t) \in Y$ the potential will be taken to be confining. The precise assumptions on the potential are given in assumption 2.1.

Using a slight modification of the argument used in the proof of [21, proposition 5.5] (see also [48, theorem A.5]) we can prove that $-\mathcal{L}$ defined in (8) generates a contraction semigroup, see proposition 3.1.

Our first result concerns the ergodicity of the SDE (7) in X or in Y. To prove the ergodicity of the SDE in Y we need to make the following assumptions on the potential.

Assumption 2.1.

(i) *$V(q) \in C^\infty(\mathbb{R}^d)$ is a confining potential.*

(ii) *There exist strictly positive constants β, σ such that $\langle \nabla_q V, q\rangle \geqslant \sigma V(q) + \beta \|q\|^2$ where $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ denote the Euclidean inner product and norm, respectively.*

(iii) *There exists a constant c such that $\|\nabla^2 V\| \leqslant c$, where $\|\cdot\|$ denotes the Frobenius–Perron matrix norm and ∇^2 the Hessian.*

The density $\rho_\beta(q, p, z)$ of the invariant measure $\mu_\beta(\mathrm{d}q\,\mathrm{d}p\,\mathrm{d}\boldsymbol{z})$ of the process (7), which is the unique solution of the stationary Fokker–Planck equation, is known:

$$\rho_\beta(q, p, \boldsymbol{z}) = \frac{1}{\mathcal{Z}_\beta} \mathrm{e}^{-\beta(\frac{1}{2}|p|^2 + V(q) + \frac{1}{2}\|\boldsymbol{z}\|^2)}, \tag{10}$$

where $\mathcal{Z}_\beta$ is the normalization constant. This invariant measure is unique and the law of the process (7) converges exponentially fast to μ_β (*geometric ergodicity*).

Theorem 2.1 (Ergodicity). *The solution of (7) with $\boldsymbol{x}(t) \in X$ and $V(q) \in C^\infty(\mathbb{T}^d)$ is geometrically ergodic. The same holds true when $\boldsymbol{x}(t) \in Y$, provided that the potential $V(q) \in C^\infty(\mathbb{R}^d)$ satisfies assumption 2.1.*

The proof of this theorem, which is based on Markov chain-type arguments and which is similar to the proof presented in [45], see also [34], can be found in appendix B.

We can prove exponentially fast convergence to equilibrium using tools from the theory of hypocoercivity [48]. We will use the notation $\mathcal{K} := \mathrm{Ker}(\mathcal{L})$ and H^1_ρ for the weighted Sobolev space H^1 with respect to μ_β on either X or Y.

Theorem 2.2. *Let $-\mathcal{L}$ be the generator of the process $\boldsymbol{x}(t) \in X$, the solution of (7) and assume that $V(q) \in C^\infty(\mathbb{T}^d)$. Then there exist constants $C, \lambda > 0$ such that*

$$\|\mathrm{e}^{-t\mathcal{L}}\|_{H^1_\rho/\mathcal{K} \to H^1_\rho/\mathcal{K}} \leqslant C\mathrm{e}^{-\lambda t}.$$

The same holds true when $\boldsymbol{x}(t) \in Y$, provided that the potential satisfies assumptions 2.1(i) and 2.1(iii).

Using the tools from [48] we can prove exponentially fast convergence to equilibrium in relative entropy. The relative entropy (or Kullback information) between two probability measures μ and ν with smooth densities f and ρ, respectively, is defined as

$$H_\rho(f) = \int f \log\left(\frac{f}{\rho}\right) \mathrm{d}\boldsymbol{x}.$$

We will measure the distance in relative entropy between the law of the process $\boldsymbol{x}(t)$ at time t and the equilibrium distribution . Since the operator $\frac{\partial}{\partial t} + \mathcal{L}$ is hypoelliptic, the law of the process $\boldsymbol{x}(t)$ in (7) has a smooth density with respect to Lebesgue which we will denote by f_t.

Theorem 2.3 (Convergence to equilibrium). *Let f_t be the density of the law of the process $\boldsymbol{x}(t)$ at time t and assume that $H_\rho(f_0) < +\infty$ and $V(q) \in C^\infty(\mathbb{T}^d)$. Then there exist constants $C, \alpha > 0$ such that*

$$H_\rho(f_t) \leqslant C\mathrm{e}^{-\alpha t} H_\rho(f_0).$$

The same holds true when $\boldsymbol{x}(t) \in Y$, assuming that $H_\rho(f_0) < +\infty$ and provided that the potential $V(q)$ satisfies assumption 2.1(i) and 2.1(iii).

Remark 2.1. In view of the Csiszar–Kullback (Pinsker) inequality

$$\frac{1}{2}\|f_t - \rho\|_{L^1}^2 \leqslant H_\rho(f_t), \tag{11}$$

Theorem 2.3 implies that, for initial data with finite relative entropy, we have exponentially fast convergence to equilibrium in L^1. For more details on inequality (11), we refer the reader to [4, 32, 47].

The proofs of theorems 2.2 and 2.3 are presented in section 3.

Estimates on the Markov semigroup associated with the Langevin equation and its derivatives can be proved using an appropriate Lyapunov function with time dependent coefficients [17, 22]. In this paper we use similar techniques to obtain estimates on the Markov semigroup and its derivatives for the Markovian approximation to the GLE, equation (7). We introduce $C_k,\ k = 0, 1, 2$ with $C_0 = A,\ C_1 = [A, B]$ and $C_2 = [C_1, B]$ (see section 2.1). We will use the notation $L^2_\rho := L^2(\cdot\,; \mu_\beta(\mathrm{d}\boldsymbol{x}))$ where $\cdot$ is either X or Y.

Theorem 2.4 (Estimates on derivatives of the Markov semigroup). *Let $-\mathcal{L}$ be the generator of the process $\boldsymbol{x}(t) \in X$, the solution of (7) with $V(q) \in C^\infty(\mathbb{T}^d)$. Then the Markov semigroup $\mathrm{e}^{-t\mathcal{L}}$ satisfies the bounds*

$$\|C_k \mathrm{e}^{-t\mathcal{L}}\|_{L^2_\rho \to L^2_\rho} \leqslant \frac{c}{t^{\frac{1+2k}{2}}}, \qquad k = 0, 1, 2 \quad \text{and} \quad t \in (0, 1], \tag{12}$$

for some (explicitly computable) positive constant c. The same holds true when $\boldsymbol{x}(t) \in Y$, provided that the potential $V(q)$ satisfies assumption 2.1(i) and (iii).

Remark 2.2. Theorem 2.4 is a short time asymptotics result. As noticed in [22], using estimate (12) together with the semigroup property and the contractivity of the semigroup we obtain that $\forall h \in L^2_\rho$ and $\forall t > 0$

$$\begin{aligned}
\|C_k \mathrm{e}^{-t\mathcal{L}} h\|_{L^2_\rho} &= \left\| C_k \mathrm{e}^{-\frac{1}{2}\mathcal{L}} \left(\mathrm{e}^{-(t-\frac{1}{2})\mathcal{L}} h \right) \right\|_{L^2_\rho} \\
&\leqslant c\, 2^{\frac{1+2k}{2}} \|\mathrm{e}^{-(t-1/2)\mathcal{L}} h\|_{L^2_\rho} \\
&\leqslant c \|h\|_{L^2_\rho},
\end{aligned}$$

hence

$$\|C_k \mathrm{e}^{-t\mathcal{L}} h\|_{L^2_\rho} \leqslant c \left(1 + \frac{1}{t^{\frac{1+2k}{2}}} \right) \|h\|_{L^2_\rho} \qquad k = 0, 1, 2, \quad t > 0, \quad h \in L^2_\rho.$$

Remark 2.3. This result can also be obtained by applying theorem A.3. Malliavin calculus-based arguments show that estimate (12) is sharp.

When the potential $V(q)$ is periodic, the particle position, appropriately rescaled, converges weakly to a Brownian motion with a diffusion coefficient which can be calculated in terms of the solution of an appropriate Poisson equation. Results of this form have been known for a long time for the Smoluchowski (overdamped) equation [42, chapter 13] as well as for the Langevin dynamics [18, 40]. In this paper we prove a similar result for the GLE. We will use the notation $\phi^e := \phi \cdot e,\ p^e := p \cdot e$, where e denotes an arbitrary unit vector in $\mathbb{R}^d$.

When $V(q)$ is 1-periodic then $q(t)$ enters in the definition of the process $\boldsymbol{x}(t) = \{q(t), p(t), \boldsymbol{z}(t)\}$ only mod 1, so we may replace $q(t)$ by $\underline{q}(t) \in \mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}$. The Markov process $\underline{\boldsymbol{x}}(t) = \{\underline{q}(t), p(t), \boldsymbol{z}(t)\}$ has state space $\mathbb{T}^d \times R^d \times \mathbb{R}^{md}$ and, according to theorem 2.1, it is an ergodic Markov process with invariant measure given by (10). For this process we prove the homogenization theorem. To simplify the notation we shall drop the underbar from $\underline{\boldsymbol{x}}(t)$ and $\underline{q}(t)$.

Theorem 2.5 (Homogenization). *Let $\boldsymbol{x}(t)$ be the solution of (7) with $V(q) \in C^\infty(\mathbb{T}^d)$ with stationary initial conditions. Then the rescaled process $q^e_\epsilon(t) := e \cdot \epsilon q(t/\epsilon^2)$ converges weakly on $C([0,T],\mathbb{R})$ to a Brownian motion with diffusion coefficient D with*

$$D^e := De \cdot e = \beta^{-1} \sum_{j=1}^{m} \alpha_j \|\nabla_{z_j} \phi^e\|^2, \tag{13}$$

where $\phi^e \in L^2_\rho$ is the unique, smooth, mean zero, periodic in q solution of the Poisson equation

$$\mathcal{L}\phi^e = p^e \tag{14}$$

on X. Furthermore, the following estimates hold

$$0 < D^e \leqslant \frac{4}{\beta} \sum_{i=1}^{m} \frac{\alpha_i}{\lambda_i^2}. \tag{15}$$

The proof of this theorem is based on a careful study of the Poisson equation (14). The well-posedness of this equation follows from the compactness of the resolvent of $\mathcal{L}$.

Proposition 2.1. *Let $-\mathcal{L}$ be the generator of the process $\boldsymbol{x}(t) \in X$, the solution of (7) and assume that $V(q) \in C^\infty(\mathbb{T}^d)$. Then $\mathcal{L}$ has compact resolvent in $L^2_\rho/\mathcal{K}$. The same holds true when $\boldsymbol{x}(t) \in Y$, provided that the assumptions of theorem 2.2 are satisfied.*

Let $q(t)$ be the solution of the Langevin equation (9) and let $q^\gamma(t) := q(\gamma t)$. It is well known that this rescaled process converges strongly in the overdamped limit $\gamma \to +\infty$ to the solution of the Smoluchowski equation [37, chapter 10]

$$\dot{q} = -\nabla V(q) + \sqrt{2\beta^{-1}}\,\dot{W}. \tag{16}$$

Similar results have also been proven in infinite dimensions [6, 7]. In this paper, we prove a result of this type for the convergence of solutions to the Markovian approximation of the GLE to the Langevin equation in the strong topology and obtain a formula for the friction coefficient that appears in the limiting Langevin equation.

Consider (1) with the rescaled noise process

$$F^\epsilon(t) := \frac{1}{\sqrt{\epsilon}} F(t/\epsilon), \tag{17}$$

which is a mean zero stationary Gaussian process with autocorrelation function

$$\gamma^\epsilon(t) = \frac{1}{\epsilon}\gamma(t/\epsilon). \tag{18}$$

For the memory kernel (6), $\gamma^\epsilon(t)$ becomes

$$\gamma^\epsilon(t) = \sum_{j=1}^{m} \frac{\lambda_j^2}{\epsilon} \mathrm{e}^{-\frac{\alpha_j}{\epsilon}|t|}. \tag{19}$$

Consequently, the rescaled noise process (17) is obtained by rescaling the coefficients in (7) according to $\lambda_j \to \frac{\lambda_j}{\sqrt{\epsilon}}$, $\alpha_j \to \frac{\alpha_j}{\epsilon}$. Under this rescaling the SDEs become

$$\mathrm{d}q(t) = p(t)\mathrm{d}t, \tag{20a}$$

$$\mathrm{d}p(t) = -\nabla_q V(q)\,\mathrm{d}t + \frac{1}{\sqrt{\epsilon}} \sum_{i=1}^{m} \lambda_i z_i(t)\,\mathrm{d}t, \tag{20b}$$

$$\mathrm{d}z_i(t) = -\frac{\lambda_i}{\sqrt{\epsilon}} p(t)\,\mathrm{d}t - \frac{\alpha_i}{\epsilon} z_i(t)\,\mathrm{d}t + \sqrt{\frac{2\alpha_i\beta^{-1}}{\epsilon}}\,\mathrm{d}W_i, \qquad i = 1, \ldots, m. \tag{20c}$$

Theorem 2.6 (The white noise limit). *Let $\{q(t), p(t), z(t)\} \in X$ be the solution of (20) with $V(q) \in C^\infty(\mathbb{T}^d)$ and initial conditions having finite moments of all orders. Then the process $\{q(t), p(t)\}$ converges strongly, as $\epsilon \to 0$, to the solution of the Langevin equation*

$$\begin{cases} \mathrm{d}Q(t) = P(t)\,\mathrm{d}t, \\ \mathrm{d}P(t) = \left(-\nabla_q V(Q(t)) - \sum_{i=1}^m \frac{\lambda_i^2}{\alpha_i} P(t)\right)\,\mathrm{d}t + \sum_{i=1}^m \sqrt{\frac{2\beta^{-1}\lambda_i^2}{\alpha_i}}\,\mathrm{d}W_i, \end{cases} \tag{21}$$

with the same initial conditions as q and p. Furthermore, for any $n \geqslant 1$, the following estimate holds:

$$\|q(t) - Q(t)\|_{n,\infty} + \|p(t) - P(t)\|_{n,\infty} \leqslant C\epsilon^{\frac{1}{2}}, \tag{22}$$

where $\|f(t)\|_{n,\infty} := (\mathbb{E}\sup_{t\in[0,T]} |f(t)|^n)^{1/n}$. The same result holds true when $\{q(t), p(t), z(t)\} \in Y$ provided that the potential $V(q)$ satisfies assumption 2.1(iii). Consequently, the process $\{q(t), p(t)\}$ converges weakly to the solution of the Langevin equation

$$\begin{cases} \mathrm{d}Q(t) = P(t)\,\mathrm{d}t, \\ \mathrm{d}P(t) = \left(-\nabla_q V(Q(t)) - \gamma P\right)\,\mathrm{d}t + \sqrt{2\gamma\beta^{-1}}\,\mathrm{d}W, \end{cases} \tag{23}$$

where the friction coefficient γ is given by the formula

$$\gamma = \sum_{j=1}^m \frac{\lambda_i^2}{\alpha_i}. \tag{24}$$

Remark 2.4. Note that the friction coefficient in (24) is precisely

$$\gamma = \int_0^{+\infty} \gamma_m(t)\,\mathrm{d}t,$$

with $\gamma_m(t)$ defined in (6), which is the formula for the friction coefficient that is commonly used in statistical physics.

2.1. Notation

For $x(t) = (q, p, z) \in Y := \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{dm}$ or $x(t) \in X := \mathbb{T}^d \times \mathbb{R}^d \times \mathbb{R}^{dm}$ consider the operator $\mathcal{L}$ defined in (8):

$$\begin{aligned} -\mathcal{L} = {} & p\cdot\nabla_q - \nabla_q V(q)\cdot\nabla_p + \left(\sum_{j=1}^m \lambda_j z_j\right)\cdot\nabla_p \\ & + \sum_{j=1}^m \left(-\alpha_j z_j\cdot\nabla_{z_j} - \lambda_j p\cdot\nabla_{z_j} + \beta^{-1}\alpha_j\Delta_{z_j}\right), \end{aligned} \tag{25}$$

with kernel $\mathcal{K} := \mathrm{Ker}\mathcal{L}$. The density of the invariant measure of the process $x(t)$ is

$$\rho_\beta(p, q, z) = \frac{1}{\mathcal{Z}_\beta} \mathrm{e}^{-\beta(V(q)+\frac{1}{2}|p|^2+\frac{1}{2}|z|^2)}, \qquad \mathcal{Z}_\beta = \int \mathrm{e}^{-\beta(V(q)+\frac{1}{2}|p|^2+\frac{1}{2}|z|^2)}\,\mathrm{d}p\,\mathrm{d}q\,\mathrm{d}z, \tag{26}$$

where $|\cdot|$ denotes either the Euclidean or the matrix norm. In (25), ∇ is the gradient (or the derivative when $d = 1$) and Δ the Laplacian. ∇^2 denotes the Hessian and if O is an operator then O^* is its adjoint in $L^2_\rho := L^2(\,\cdot\,; \mu_\beta(\mathrm{d}x))$. Define

$$B = -p\cdot\nabla_q + \nabla_q V\cdot\nabla_p - \sum_{j=1}^m \lambda_j\left(z_j\cdot\nabla_p - p\cdot\nabla_{z_j}\right). \tag{27}$$

We easily check that $B^* = -B$. To simplify the notation, we set $\beta = \alpha_j = 1$. When $m = 1$ then $A_i = -\partial_{z_i}$ (derivative with respect to the ith component of z) so that $A_i^* = -z_i + \partial_{z_i}$ and we can write

$$\mathcal{L} = B + \sum_{i=1}^{d} A_i^* A_i =: B + A^* A, \tag{28}$$

where A is intended to be the row vector of operators $(A_1, \dots, A_d)$ (the same for A^*). More precisely, if $m = 1$ then: $A : L^2_\rho \longrightarrow L^2_\rho \otimes \mathbb{R}^d$, $B : L^2_\rho \longrightarrow L^2_\rho$, $[A^*, A] : L^2_\rho \longrightarrow L^2_\rho$, being $[A^*, A] := \sum_{j=1}^{d}[A_j^*, A_j]$; on the other hand $[A, A^*] : L^2_\rho \longrightarrow L^2_\rho \otimes \mathbb{R}^d \otimes \mathbb{R}^d$ is a matrix of operators whose ijth component is given by $[A, A^*]_{ij} := [A_i, A_j^*]$; in an analogous way $[A, A] : L^2_\rho \longrightarrow L^2_\rho \otimes \mathbb{R}^d \otimes \mathbb{R}^d$ is a matrix of operators with $[A, A]_{ij} := [A_i, A_j]$; finally $C := [A, B]$, $C : L^2_\rho \longrightarrow L^2_\rho \otimes \mathbb{R}^d$ is a vector of operators, $C_i = [A_i, B]$, $i = 1 \dots d$, and the same holds for $C_2 := [C, B]$, $C_2 : L^2_\rho \longrightarrow L^2_\rho \otimes \mathbb{R}^d$.

When $m > 1$ then (28) becomes

$$\mathcal{L} = B + \sum_{i=1}^{m} \sum_{j=1}^{d} A_{ij}^* A_{ij} \tag{29}$$

with $A_{ij} = -\partial_{z_{i_j}}$ i.e. the partial derivative with respect to the jth component of z_i, and $A_{ij}^* = -z_{i_j} + \partial_{z_{i_j}}$. We will use the notation

$$\mathcal{L} = B + A^* A, \tag{30}$$

meaning either (28) or (29). For a detailed account on the use of this notation, see [48, pp 14–15].

As for the norms, unless otherwise specified, $\|\cdot\|$ indicates the norm of L^2_ρ, $\|\cdot\|_1^2 = \|A\cdot\|^2 + \|C\cdot\|^2 + \|C_2\cdot\|^2$ is a kind of homogeneous $H^1(Y; \mu_\beta(\mathrm{d}\boldsymbol{x})) =: H^1_\rho$ norm and $\|\cdot\|^2_{H^1_\rho} = \|\cdot\|^2 + \|A\cdot\|^2 + \|C\cdot\|^2 + \|C_2\cdot\|^2$ is the usual inhomogeneous one. The inner products in these Hilbert spaces are denoted by $(\cdot,\cdot)$, $(\cdot,\cdot)_1$ and $(\cdot,\cdot)_{H^1_\rho}$, respectively.

3. Convergence to equilibrium

In this section we present the proofs of theorems 2.2 and 2.3. As a preliminary result we show that $-\mathcal{L}$ given by (8) generates a contraction semigroup.

Proposition 3.1. *Let $-\mathcal{L}$ be the generator of the process $\boldsymbol{x}(t) \in X$, the solution of (7) and assume that $V(q) \in C^\infty(\mathbb{R}^d)$. Then $-\mathcal{L}$ generates a contraction semigroup.*

Proof. The proof is almost identical to the proof of proposition 5.5 in [21] and we will be very brief[2]. Let $\mathcal{L} = B + A^*A$. To simplify the notation, we will set all the constants equal to 1 and will also consider the case $d = m = 1$. Clearly, $\mathcal{L}$ is an accretive operator. Furthermore, its domain of definition is dense in L^2_ρ. Thus, we can consider its closure, which we will still denote by $\mathcal{L}$. We define $T = \mathcal{L} + 2I$. From the Lumer–Phillips theorem, [43, theorem X 48], to prove that $\mathcal{L}$ generates a contraction semigroup it is sufficient to show that the range of T is dense in L^2_ρ. For this it is sufficient to show that if

$$(f, Tu) = 0 \qquad \forall u \in C_0^\infty, \tag{31}$$

[2] Note, however, that rather than transforming $\mathcal{L}$ into a Schrödinger operator and working in a flat L^2 space, we work with the generator in its original form in the weighted L^2 space.

then $f = 0$. Note that equation (31) is equivalent to $(A^*A - B + 2I)f = 0$ in the distributional sense. Hence, by hypoellipticity (see equation (32)), this implies that f is a C^∞ function. Following the proof of [21, proposition 5.5], we introduce a family of cut-off functions

$$\zeta_k(q, p, z) := \zeta\left(\frac{q}{k}\right)\zeta\left(\frac{p}{\alpha(k)}\right)\zeta\left(\frac{z}{\omega(k)}\right), \qquad \forall k \in \mathbb{N}_+,$$

where ζ is a C^∞ function satisfying $\zeta \in [0, 1]$, $\zeta = 1$ on $B(0, 1)$ and $\operatorname{supp}\zeta \in B(0, 2)$, $\alpha(k)$ and $\omega(k)$ are positive functions which we will choose later on. With calculations analogous to those presented in [21, proposition 5.5] we have that for any $u \in C^\infty$,

$$\begin{aligned}
&(f, T(\zeta_k^2 u)) - (\partial_z(\zeta_k f), \partial_z(\zeta_k u)) \\
&\quad = (\partial_z f, \partial_z(\zeta_k^2 u)) + (f, B(\zeta_k^2 u)) + 2(f, \zeta_k^2 u) - (\partial_z(\zeta_k f), \partial_z(\zeta_k u)) \\
&\quad = (\zeta_k \partial_z f, u\partial_z\zeta_k) - (f\partial_z\zeta_k, u\partial_z\zeta_k) - (f\partial_z\zeta_k, \zeta_k\partial_z u) + 2(f, \zeta_k^2 u) + (f, B(\zeta_k^2 u)).
\end{aligned}$$

Let now f be the solution of (31) and choose $u = f$ in the above identity to obtain

$$2\|\partial_z(\zeta_k f)\|^2 + \|\partial_z(\zeta_k f)\|^2 = \|(\partial_z\zeta_k) f\|^2 - (f, B(\zeta_k^2 f)).$$

We use now the identity $(f, B(\zeta_k^2 f)) = (\zeta_k f^2, B\zeta_k)$, which follows from the antisymmetry of B, to deduce

$$2\|\zeta_k f\|^2 \leqslant \|f\partial_z\zeta_k\|^2 - (\zeta_k f^2, B\zeta_k).$$

Setting $\tilde{C}(k) := \sup_{|q|\leqslant 2k} |\partial_q V(q)|$, we then have

$$2\|\zeta_k f\|^2 \leqslant \frac{1}{\omega^2(k)}\|f\|^2 + \|f\|^2 + \frac{\tilde{C}(k)}{\alpha(k)}\|f\|^2 + \frac{k}{\alpha(k)}\|f\|^2 + \frac{k}{\omega(k)}\|f\|^2.$$

We now choose $\alpha(k)$ and $\omega(k)$ such that, as $k \to \infty$, $\omega(k) \to \infty$, $\tilde{C}(k)/\alpha(k) \to 0$ and $k/\alpha(k), k/\omega(k) \to 0$. So, letting $k \to \infty$, from the above inequality we obtain $\|f\|^2 = 0$, hence $f = 0$. $\square$

3.1. Hypocoercivity

Background material on hypocoercivity is presented in appendix A. In this section we only give the definition of hypocoercivity. To this end, let $\mathcal{T}$ be an unbounded operator on a Hilbert space $\mathcal{H}$ with kernel $\mathcal{K}$. Let $\tilde{\mathcal{H}}$ be another Hilbert space continuously and densely embedded in $\mathcal{K}^\perp$.

Definition 3.1 (Hypocoercivity). *Assume $\mathcal{T}$ generates a continuous semigroup. Then $\mathcal{T}$ is said to be λ-hypocoercive on $\tilde{\mathcal{H}}$ if there exists a constant $\kappa > 0$ such that*

$$\| \mathrm{e}^{-\mathcal{T}t} h \|_{\tilde{\mathcal{H}}} \leqslant \kappa \mathrm{e}^{-\lambda t} \| h \|_{\tilde{\mathcal{H}}} \qquad \forall h \in \tilde{\mathcal{H}} \quad \text{and} \quad t \geqslant 0.$$

We say that an unbounded linear operator S on $\mathcal{H}$ is *relatively bounded* with respect to the (linear unbounded) operators $T_1, \ldots, T_n$ if $\mathcal{D}(S) \subset (\cap \mathcal{D}(T_j))$ and $\exists$ a constant $\alpha > 0$ s.t.

$$\forall h \in \mathcal{D}(S), \qquad \|Sh\| \leqslant \alpha(\|T_1 h\| + \cdots + \|T_n h\|).$$

The basic idea employed in the proof of exponentially fast convergence to equilibrium for hypocoercive diffusions is to appropriately construct a scalar product on H^1_ρ by adding lower order terms and then use the fact that hypocoercivity is invariant under a change of equivalent norms, whereas coercivity does not enjoy such invariance. Finally, we note that $\mathcal{S}$, the class of Schwartz functions, is dense in $D(A) \cap D(B)$ as well as in L^2_ρ. This guarantees that all the operations performed with these (unbounded) operators are well defined.

Set $m = 1 = d$, $\alpha = \lambda = \beta = 1$. The first two commutators are

$$C_1 = C = [A, B] = \partial_p \qquad \text{and} \qquad C_2 = [C, B] = \partial_z - \partial_q. \tag{32}$$

Hence the operator is hypoelliptic [23]. Furthermore,

$$[A, A] = 0 \qquad [A, C] = 0 \qquad [A, C_2] = 0, \tag{33a}$$

$$[A, A^*] = I \qquad [C, A^*] = 0 \qquad [C_2, A^*] = -I, \tag{33b}$$

$$[C_2, B] = -\partial^2 V \partial_p - \partial_p, \tag{33c}$$

$$[C, C^*] = I \qquad [C_2^*, C_2] = -I - \partial_q^2 V, \tag{33d}$$

where I is the identity operator.

3.2. *Proof of theorem 2.2*

Proof. We will use theorem A.2 . To this end, set

$$P = A^*A + C^*C + C_2^*C_2$$

and note that $\mathrm{Ker}(P) = \mathcal{K} =: \mathrm{Ker}\mathcal{L}$ contains only constants; in fact

$$\mathrm{Ker}(P) = \mathrm{Ker}(A^*A) \cap \mathrm{Ker}(C^*C) \cap \mathrm{Ker}(C_2^*C_2) = \mathrm{Ker}(A) \cap \mathrm{Ker}(C) \cap \mathrm{Ker}(C_2).$$

To show that $\mathcal{K} = \mathrm{Ker}(A^*A) \cap \mathrm{Ker}(C^*C) \cap \mathrm{Ker}(C_2^*C_2)$: the inclusion $\supseteq$ is obvious. For the other inclusion: if $h \in \mathcal{K}$ then $\|Ah\|^2 + \|Ch\|^2 + \|C_2h\|^2 = 0 \Rightarrow Ah = Ch = C_2h = 0$.

Theorem A.2 requires two sets of hypotheses to be fulfilled. Hypotheses 1, 2 and 3 in theorem A.2 are quantitative assumptions, which are satisfied in our case with $N = 2$, $C_0 = A$, $C_1 = C$, $R_1 = R_2 = 0$, $R_3 = [C_2, B]$ (this is to have $C_3 = 0$) and thanks to assumption 2.1(iii). Hypothesis 4 requires, in our case, for the operator P to be κ-*coercive* on $\mathcal{K}^\perp \cong L^2_\rho/\mathcal{K}$. The coercivity of P is equivalent to

$$\|Ah\|^2 + \|Ch\|^2 + \|C_2h\|^2 \geqslant \kappa\|h\|^2,$$

that is, more explicitly,

$$\|\nabla_z h\|^2 + \|\nabla_p h\|^2 + \|\left(\nabla_z - \nabla_q\right) h\|^2 \geqslant \kappa\|h\|^2.$$

Using the fact that $\|a - b\|^2 \geqslant \frac{\|a\|^2}{3} - \frac{\|b\|^2}{2}$, we have

$$\|\nabla_z h\|^2 + \|\nabla_p h\|^2 + \|\left(\nabla_z - \nabla_q\right) h\|^2 \geqslant \tfrac{1}{3}\left(\|\nabla_z h\|^2 + \|\nabla_p h\|^2 + \|\nabla_q h\|^2\right)$$

so we just need

$$\|\nabla_z h\|^2 + \|\nabla_p h\|^2 + \|\nabla_q h\|^2 \geqslant \kappa\|h\|^2$$

to hold true. Since μ_β is a product measure, we only need to verify that[3]

$$\int |\nabla_q h|^2 \mathrm{e}^{-V(q)}\,\mathrm{d}q \geqslant \mu \int (h - \langle h\rangle)^2 \mathrm{e}^{-V(q)}\,\mathrm{d}q$$

holds true for some constant μ, where the notation $\langle h\rangle := \int h\mathrm{e}^{-V(q)\mathrm{d}q}$ has been used. It is a standard result that if $V(q) \in C^2(\mathbb{R}^d)$ is such that $\mathrm{e}^{-V(q)}/\mathcal{Z}$ is a probability density and

$$\frac{|\ \nabla V(q)\ |^2}{2} - \Delta V(q) \overset{|q|\to\infty}{\longrightarrow} +\infty \tag{34}$$

then $\mathrm{e}^{-V(q)}/\mathcal{Z}$ satisfies a Poincaré inequality (see, e.g. [48, theorem A.1]). From assumption 2.1(iii), condition (34) is satisfied. We can conclude that there exist a scalar

[3] To simplify the notation we have set $\beta = 1$.

product $((\cdot,\cdot))$ inducing a norm equivalent to the inhomogeneous norm of H^1_ρ and a constant $\hat\lambda > 0$ such that $\mathcal{L}$ is coercive in this norm:

$$\forall h \in L^2_\rho/\mathcal{K}, \qquad ((h, \mathcal{L}h)) \geqslant \hat\lambda((h,h)).$$

This implies that $\mathcal{L}$ is hypocoercive in this norm, hence it is hypocoercive on $L^2_\rho/\mathcal{K}$ endowed with the $\|\cdot\|_{H^1_\rho}$ norm:

$$\|e^{-t\mathcal{L}}h_0\|_{H^1_\rho} \leqslant C e^{-\lambda t}\|h_0\|_{H^1_\rho}. \tag{35}$$

□

Remark 3.1. The orthogonal space to $\mathcal{K}$ is the same with respect to both the $(\cdot,\cdot)_1$ and the $(\cdot,\cdot)_{H^1_\rho}$ norms; moreover, since P is coercive, these two norms are equivalent.

Remark 3.2. Theorem A.3 in appendix A allows us to state a similar result when the initial datum is in L^2_ρ. In fact, using remark 3.1,

$$\|e^{-t\mathcal{L}}h\|_{H^1_\rho} \leqslant \frac{c}{t^{\frac{5}{2}}}\|h\|, \qquad t \in (0,1]. \tag{36}$$

So, putting together (35) and (36) we obtain, for $0 < t_0 < t$, $t_0 < 1$:

$$\begin{aligned}\|e^{-t\mathcal{L}}h_0\|_{H^1_\rho} &= \|e^{-(t-t_0)\mathcal{L}}e^{-t_0\mathcal{L}}h_0\|_{H^1_\rho} = \|e^{-(t-t_0)\mathcal{L}}h_{t_0}\|_{H^1_\rho} \\ &\leqslant c e^{-\lambda(t-t_0)}\|h_{t_0}\|_{H^1_\rho} \leqslant c e^{-\lambda(t-t_0)}\|e^{-t_0\mathcal{L}}h_0\|_{H^1_\rho} \\ &\leqslant c\frac{e^{-\lambda(t-t_0)}}{t_0^{\frac{5}{2}}}\|h_0\|,\end{aligned} \tag{37}$$

where the notation $e^{-t_0\mathcal{L}}h_0 =: h_{t_0}$ has been used.

Remark 3.3. The proof is identical when $m, d > 1$. In this case we can think of A as a matrix of operators, see (29).

3.3. Proof of theorem 2.3

Proof. For simplicity we present the proof of this result in one dimension, i.e. $d = 1$, and for $m = 1$; we also set $\alpha = \beta = 1$. The extension to arbitrary dimensions is straightforward.

Let f_t denote the density of the law of the process $x(t)$, i.e. the solution of the Fokker–Plank equation

$$\partial_t f_t + \mathcal{L}' f_t = 0,$$

where $\mathcal{L}'$ denotes the (flat) L^2 adjoint of $\mathcal{L}$, namely

$$\mathcal{L}' = p\partial_q - \partial_q V\partial_p + z\partial_p - p\partial_z - \partial_z(z\cdot) - \partial_z^2.$$

We set $f_t = \rho h_t$. Then h_t satisfies the equation

$$\partial_t h_t = Bh_t - A^*Ah_t. \tag{38}$$

We apply theorem A.4 to the operator $\mathcal{F} = -B + A^*A$ with

$$A = -\partial_z, \qquad C_1 = -\partial_p, \qquad C_2 = -\partial_q, \qquad Z_2 = I, \qquad R_2 = -\partial_z.$$

Furthermore assumption 2.1(i) and (iii) together with the Holley–Strook perturbation lemma imply that $\mathcal{Z}^{-1}e^{-V(q)}$ satisfies a logarithmic Sobolev inequality (LSI).

Hypotheses 1, 2 and 4 are automatically satisfied. We put $C_2 = \partial_q$ and we added the remainder R_2 in order to fulfill hypothesis 4. Hypothesis 3 is satisfied on account of assumption 2.1 (iii). Now consider the relative entropy $H_\rho(f)$,

$$H_\rho(f) = \int f \log\left(\frac{f}{\rho}\right) \mathrm{d}q\,\mathrm{d}p\,\mathrm{d}r = \int h \log h \,\mathrm{d}\rho, \qquad f = \rho h \tag{39}$$

and the Fisher information $I_\rho(f)$

$$I_\rho(f) = \int f|\nabla \log(h)|^2 \,\mathrm{d}q\,\mathrm{d}p\,\mathrm{d}r = \int h|\nabla \log h|^2 \,\mathrm{d}\rho, \qquad f = \rho h. \tag{40}$$

Then if the initial datum has finite relative entropy, we obtain that

$$H_\rho(f_t) = \mathcal{O}(\mathrm{e}^{-t\alpha}) \tag{41}$$

for some $\alpha > 0$ and for $t > 0$. If the initial datum has also finite Fisher information then

$$I_\rho(f_t) = \mathcal{O}(\mathrm{e}^{-t\alpha}), \tag{42}$$

as well. □

Remark 3.4. We note that (42), together with the LSI, implies (41).

Remark 3.5. In view of the LSI, it is interesting to note that, by applying theorem A.5, we get the following bounds:

$$\int h_t |C_k \log h_t|^2 \,\mathrm{d}\rho \leqslant \frac{c}{t^{2k+1}} \int h_0 \log h_0 \,\mathrm{d}\rho, \tag{43}$$

for $k = 0, 1, 2$ and c an explicitly computable positive constant.

4. Bounds on the derivatives of the Markov semigroup

Throughout this section we will use the notation $u = \mathrm{e}^{-t\mathcal{L}}u_0$. We introduce the Lyapunov function

$$F(t) = a_0 t\|Au\|^2 + a_1 t^3\|Cu\|^2 + a_2 t^5\|C_2 u\|^2 + b_0 t^2(Au, Cu) + t^4 b_1(Cu, C_2 u) + b_2\|u\|^2, \quad t \in (0, 1], \tag{44}$$

where a_j, b_j, $j = 0, 1, 2$ are positive constants to be chosen.

Lemma 4.1. *There exist constants a_j, b_j, $j = 0, 1, 2$ such that the time derivative $\partial_t F$ of the Lyapunov function along the semigroup is negative.*

Proof. We will calculate the time derivative of each term in (44) separately and using the explicit relations (33):

$$\begin{aligned}
&\partial_t\|u\|^2 = -2(\mathcal{L}u, u) = -2\|Au\|^2,\\
&\partial_t(Au, Au) = -2(Cu, Au) - 2\|A^*Au\|^2 = -2(Cu, Au) - 2\|Au\|^2 - 2\|A^2u\|^2,\\
&\partial_t(Cu, Cu) = -2\|ACu\|^2 - 2(C_2u, Cu),\\
&\partial_t(C_2u, C_2u) = ((2 + \partial_q^2 V)C_2u, Cu) - 2\|AC_2u\|^2 + 2(Au, C_2u),\\
&\partial_t(Au, Cu) = -2(A^2u, ACu) - (Au, Cu) - \|Cu\|^2 - (Au, C_2u),\\
&\partial_t(Cu, C_2u) = -\|C_2u\|^2 - 2(ACu, AC_2u) + 2\|Cu\|^2 + (Cu, Au).
\end{aligned}$$

Putting everything together we obtain

$$\begin{aligned}\partial_t F(t) = & -2a_0 t\|A^2u\|^2 - 2a_1 t^3\|ACu\|^2 - 2a_2 t^5\|AC_2u\|^2 && (45a)\\ & -2b_0 t^2(A^2u, ACu) - 2b_1 t^4(ACu, AC_2u) && (45b)\\ & +(-2a_0 t + a_0 - 2b_2)\|Au\|^2 + (3a_1t^2 + 2b_1t^4 - b_0t^2)\|Cu\|^2 && (45c)\\ & +(5a_2t^4 - b_1t^4)\|C_2u\|^2 + (2b_0t - 2a_0t - b_0t^2 + b_1t^4)(Au, Cu) && (45d)\\ & +(4b_1t^3 - 2a_1t^3 + 2a_2t^5)(Cu, C_2u) + (2a_2t^5 - b_0t^2)(Au, C_2u). && (45e)\end{aligned}$$

Now we estimate the sum of the first and of the second line (i.e. the sum of all the terms where A^2, AC and AC_2 appear). For $t \in (0, 1]$ we have

$$\begin{aligned}(45a) + (45b) \leqslant & -2a_0t\|A^2u\|^2 + 2b_0t^2\|A^2u\|\|ACu\|\\ & + 2b_1t^4\|ACu\|\|AC_2u\| - 2a_1t^3\|ACu\|^2 - 2a_2t^5\|AC_2u\|^2\\ \leqslant & -2a_0t\|A^2u\|^2 + b_0^2t\|A^2u\|^2 + t^3\|ACu\|^2 - 2a_1t^3\|ACu\|^2\\ & + b_1^2t^3\|ACu\|^2 + t^5\|AC_2u\|^2 - 2a_2t^5\|AC_2u\|^2.\end{aligned}$$

Similarly for the sum of the remaining terms (those with A, C and C_2) we have

$$\begin{aligned}(45c) + (45d) + (45e) \leqslant & (-2a_0t + a_0 - 2b_2)\|Au\|^2\\ & +(2b_0t + 2a_0t + b_0t^2 + b_1t^4)\|Au\|\|Cu\| + (2a_2t^5 + b_0t^2)\|Au\|\|C_2u\|\\ & +(3a_1t^2 + 2b_1t^4 - b_0t^2)\|Cu\|^2 + (5a_2t^4 - b_1t^4)\|C_2u\|^2\\ & +(4b_1t^3 + 2a_1t^3 + 2a_2t^5)\|Cu\|\|C_2u\|\\ \leqslant & (-2a_0t + a_0 - 2b_2)\|Au\|^2 + a_0^2\|Au\|^2 + \|Cu\|^2\\ & +\frac{3}{2}b_0^2\|Au\|^2 + \frac{3}{2}t^2\|Cu\|^2 + \frac{1}{2}b_1^2\|Au\|^2 + \frac{t^4}{2}\|Cu\|^2\\ & +a_2^2t^5\|Au\|^2 + t^5\|C_2u\|^2 + \frac{t^2}{2}b_0^2\|Au\|^2 + \frac{t^2}{2}\|C_2u\|^2\\ & +(3a_1t^2 + 2b_1t^4 - b_0t^2)\|Cu\|^2 + (5a_2t^4 - b_1t^4)\|C_2u\|^2\\ & +2b_1^2t^3\|Cu\|^2 + t^3\|C_2u\|^2 + a_1^2t^3\|Cu\|^2 + t^3\|C_2u\|^2\\ & +a_2^2t^5\|Cu\|^2 + t^5\|C_2u\|^2.\end{aligned}$$

Choosing the constants in such a way that $\boxed{b_2 \gg a_0 \gg b_0 \gg a_1 \gg b_1 \gg a_2 > 1/c}$, where c is a constant depending on the bound on the second derivative of the potential, we obtain that $\partial_t F < 0 \; \forall t \in (0, 1]$. □

Proof of theorem 2.4. We use the previous lemma to deduce

$$a_0t\|Au\|^2 + a_1t^3\|Cu\|^2 + a_2t^5\|C_2u\|^2 + b_0t^2(Au, Cu) + t^4b_1(Cu, C_2u) + b_2\|u\|^2 < b_2\|u_0\|^2. \tag{46}$$

This, in turn, implies that

$$\begin{aligned}&\|\nabla_z u\|^2 = \|Au\|^2 < \frac{\kappa}{t}\|u_0\|^2,\\ &\|\nabla_p u\|^2 = \|Cu\|^2 < \frac{\kappa}{t^3}\|u_0\|^2,\\ &\frac{\|\nabla_q u\|^2}{3} - \frac{\|\nabla_z u\|^2}{2} \leqslant \|\nabla_q u - \nabla_z u\|^2 = \|C_2u\|^2 < \frac{\kappa}{t^5}\|u_0\|^2\\ &\Rightarrow \|\nabla_q u\|^2 \leqslant \frac{\kappa}{t^5}\|u_0\|^2,\end{aligned}$$

where κ is an explicitly computable positive constant. The previous inequalities are justified by the fact that

$$
\begin{aligned}
&a_0 t\|Au\|^2 + a_1 t^3\|Cu\|^2 + a_2 t^5\|C_2 u\|^2 + b_0 t^2 (Au, Cu) + t^4 b_1 (Cu, C_2 u) \\
&\quad \geqslant \left(a_0 t - \frac{b_0^2}{2} t\right)\|Au\|^2 + \left(a_1 t^3 - \frac{t^3}{2} - t^3\frac{b_1^2}{2}\right)\|Cu\|^2 + \left(a_2 t^5 - \frac{t^5}{2}\right)\|C_2 u\|^2
\end{aligned}
$$

and the second line is positive thanks to the choice of the constants we made. □

Remark 3.3 holds also in this case.

Remark 4.1. From estimates (12), similar estimates on $A^\star \mathrm{e}^{-t\mathcal{L}^\bullet}$, $\mathrm{e}^{-t\mathcal{L}^\star} A^\bullet$, $C^\star \mathrm{e}^{-t\mathcal{L}^\bullet}$, $\mathrm{e}^{-t\mathcal{L}^\star} C^\bullet$, $C_2^\star \mathrm{e}^{-t\mathcal{L}^\bullet}$ and $\mathrm{e}^{-t\mathcal{L}^\star} C_2^\bullet$ follow, where $\star$ and $\bullet$ stand for either the L^2_ρ-adjoint or nothing. In fact

(i) $(A\mathrm{e}^{-t\mathcal{L}} f, g) = (f, \mathrm{e}^{-t\mathcal{L}^*} A^* g) \leqslant \|A\mathrm{e}^{-t\mathcal{L}} f\|\|g\| \leqslant \frac{\kappa}{\sqrt{t}}\|f\|\|g\|$
$\Rightarrow (f, \mathrm{e}^{-t\mathcal{L}^*} A^* g) \leqslant \frac{\kappa}{\sqrt{t}}\|f\|\|g\|$, choose $f = \mathrm{e}^{-t\mathcal{L}^*} A^* g$ and the result on $\mathrm{e}^{-t\mathcal{L}^*} A^*$ follows.

(ii) Using $[A, A^*] = I$ we have $\|A^* \mathrm{e}^{-t\mathcal{L}} u_0\|^2 = \|A^* u\|^2 = \|Au\|^2 + \|u\|^2$, hence the estimate for $A^* \mathrm{e}^{-t\mathcal{L}}$. Taking the adjoint as in (i) we get the result for $\mathrm{e}^{-t\mathcal{L}^*} A$.

(iii) For $A\mathrm{e}^{-t\mathcal{L}^*}$ we can just repeat the proof we wrote for $A\mathrm{e}^{-t\mathcal{L}}$, since the only thing that changes when considering $\mathcal{L}^*$ is the sign of B, which does not play any role in the proof. Now, by acting as in (i) and (ii), we obtain the results for $\mathrm{e}^{-t\mathcal{L}} A^*$, $A^* \mathrm{e}^{-t\mathcal{L}^*}$ and $\mathrm{e}^{-t\mathcal{L}} A$.

5. The homogenization theorem

In this section we prove theorem 2.5. The proof of this theorem is based on standard techniques, namely the central limit theorem for additive functionals of Markov processes [27, 31, 40], which in turn is based on the martingale central limit theorem [12, theorem 7.1.4]. In order to apply these techniques we need to study the Poisson equation

$$\mathcal{L} u = f. \tag{47}$$

The boundary conditions for (47) are that $u \in L^2_\rho$ and it is periodic in q.

Proposition 5.1. *Let $f \in L^2_\rho \cap C^\infty(X)$ with $\int_X f \mu_\beta(\mathrm{d}x) = 0$. Then the Poisson equation (47) has a unique smooth mean zero solution $u \in L^2_\rho \cap C^\infty(X)$.*

The proof of theorem 2.5 follows now from the above proposition.

Proof of theorem 2.5. To simplify the notation we present the proof for $d = 1$. When $d > 1$ the same proof applies to the one-dimensional projections $q^e := q \cdot e$. In this case the diffusion coefficient D is replaced by the projections of the diffusion tensor $D^e := De \cdot e$.

We consider the process $x(t)$ on X with stationary initial conditions. For non-stationary initial conditions we need to combine the analysis presented below with the exponential convergence to equilibrium, theorem 2.2. Since $p \in L^2_\rho \cap C^\infty(X)$ and centred with respect to the invariant measure $\mu_\beta(\mathrm{d}x)$, proposition 5.1 applies and there exists a unique mean zero solution $\phi \in L^2_\rho \cap C^\infty(X)$ to the problem

$$\mathcal{L}\phi = p. \tag{48}$$

We use Itô's formula to obtain

$$\mathrm{d}\phi = \mathcal{L}\phi \,\mathrm{d}t + \sum_{j=1}^{m} \sqrt{2\alpha_j \beta^{-1}}\, \partial_{z_j}\phi \,\mathrm{d}W_j.$$

We combine this, together with (48) and the equations of motion to deduce

$$\begin{aligned} q_\epsilon(t) &:= \epsilon q(t/\epsilon^2) \\ &= \epsilon q(0) + \epsilon \int_0^{t/\epsilon^2} p(s)\,\mathrm{d}s \\ &= \epsilon q(0) - \epsilon \left[\phi(q(t/\epsilon^2), p(t/\epsilon^2), z(t/\epsilon^2)) - \phi(q(0), p(0), z(0))\right] \\ &\quad + \epsilon \sum_{j=1}^m \int_0^{t/\epsilon^2} \sqrt{2\alpha_j \beta^{-1}} \partial_{z_j} \phi \,\mathrm{d}W_j(s) \\ &=: \epsilon R^\epsilon + M^\epsilon. \end{aligned}$$

Our stationarity assumption, together with the fact that $\phi \in L^2_\rho$, imply that

$$\mathbb{E}|R^\epsilon|^2 \leqslant C.$$

To study the martingale term M^ϵ we use the martingale central limit theorem [12, theorem 7.1.4] or [42, theorem 3.33]. We have that $M^\epsilon(0) = 0$, that $M^\epsilon(t)$ has continuous sample paths and, by stationarity, that it has stationary increments. Furthermore, by ergodicity and the fact that the Brownian motions $W_i(t)$, $i = 1, \ldots, m$ are independent, we deduce that

$$\lim_{\epsilon \to 0} \langle M_t^\epsilon \rangle = 2 \sum_{i=1}^m \alpha_i \beta^{-1} \|\partial_{z_i} \phi\|^2 t \ a.s.$$

Note that in view of estimate (50), we have that $\|\partial_{z_i}\phi\| \leqslant C$. The above calculations imply that the rescaled process $q_\epsilon(t) := \epsilon q(t/\epsilon^2)$ converges weakly in $C([0,t];\mathbb{R})$ to a Brownian motion $\sqrt{2D}W(t)$ where

$$D = 2\beta^{-1} \sum_{i=1}^m \alpha_i \|\partial_{z_i} \phi\|^2. \tag{49}$$

Remark 5.1. Note that when $d > 1$ the convergence of the one-dimensional projections $q_\epsilon^e(t) := e \cdot \epsilon q(t/\epsilon^2)$ does not imply the convergence of the process $q_\epsilon(t) = \epsilon q(t/\epsilon^2)$ [18, remark 2.3]. The proof of the homogenization theorem in the multidimensional case, which is also based on the analysis of the Poisson equation, is very similar and it is omitted. Similar results for diffusion processes with periodic coefficients in arbitrary dimensions can be found in e.g. [5, 41].

To prove estimate (15), we first show the upper bound and then the fact that the diffusion coefficient is bounded away from zero. We set $\phi = g_i + \frac{1}{\lambda_i} z_i$ and use the Poisson equation (48) to obtain

$$\mathcal{L} g_i = -\frac{\alpha_i}{\lambda_i} z_i,$$

from which we obtain the estimate

$$\begin{aligned} \alpha_i \beta^{-1} \|\partial_{z_i} g_i\|^2 \leqslant \sum_{j=1}^m \alpha_j \beta^{-1} \|\partial_{z_j} g_i\|^2 &= (\mathcal{L} g_i, g_i) \\ &= \frac{\alpha_i}{\beta \lambda_i} \int g_i \partial_{z_i} \rho \,\mathrm{d}\boldsymbol{x} = -\frac{\alpha_i}{\beta\lambda_i} \int \rho\, \partial_{z_i} g_i \mathrm{d}\boldsymbol{x} \\ &\leqslant \frac{\alpha_i}{\beta \lambda_i} \|\partial_{z_i} g_i\|. \end{aligned}$$

Consequently,

$$\|\partial_{z_i} g_i\| \leqslant \frac{1}{\lambda_i}. \tag{50}$$

From this we obtain the following estimate on the diffusion coefficient D:

$$
\begin{aligned}
D &= \sum_{i=1}^{m} \alpha_i \beta^{-1} \|\partial_{z_i}\phi\|^2 = \frac{1}{\beta}\sum_{i=1}^{m}\alpha_i \left\| \partial_{z_i} g_i + \frac{1}{\lambda_i}\right\|^2 \\
&\leqslant \frac{2}{\beta}\sum_{i=1}^{m}\alpha_i \left(\|\partial_{z_i} g_i\|^2 + \frac{1}{\lambda_i^2}\right) \\
&\leqslant \frac{4}{\beta}\sum_{i=1}^{m}\frac{\alpha_i}{\lambda_i^2}.
\end{aligned}
$$

The fact that $D > 0$ is easily seen by contradiction. Assume that $D = 0$. Then by (49), $\|\partial_{z_i}\phi\|^2 = 0 \; \forall i = 1, \ldots, m$. Hence $\phi = \phi(q, p)$ and

$$
\mathcal{L}\phi = -p\partial_q\phi + \partial_q V \partial_p \phi + \sum_{i=1}^{m} \lambda_i z_i \partial_p \phi = p.
$$

Multiplying both sides by $e^{z_i^2/2}$ and then integrating with respect to z_i we get

$$
\begin{aligned}
&-\int p\partial_q\phi\, e^{z_i^2/2}\,dz_i + \int \partial_q V \partial_p\phi\, e^{z_i^2/2}\,dz_i \\
&+\int \lambda_i z_i^2 e^{z_i^2/2}\,dz_i + \sum_{j\neq i}\int \lambda_i z_i z_j \partial_p \phi e^{z_i^2/2}\,dz_i \\
&= \int pz\, e^{z_i^2/2}\,dz_i,
\end{aligned}
$$

from which we conclude that $\lambda_i \partial_p\phi = 0$ for all $i = 1, \ldots, m$. Hence $\phi = \phi(q)$. By the same reasoning we get that $-p\partial_q\phi = p$, which does not have a periodic solution. □

We now prove proposition 2.1

Proof of proposition 2.1. To prove the compactness of the resolvent we use (37) (for example with $t_0 = 1/2$) and the fact that the resolvent $\mathcal{L}_\eta^{-1} := (\eta I + \mathcal{L})^{-1}$ can be represented as the Laplace transform of the semigroup:

$$
\begin{aligned}
\|\mathcal{L}_\eta^{-1} h\|_{H_\rho^1} &\leqslant \int_0^\infty dt e^{-\eta t}\|e^{-\mathcal{L}t}h\|_{H_\rho^1} && (51)\\
&\leqslant C\int_0^\infty dt e^{-\eta t}e^{-\lambda t}\|h\| \\
&\leqslant \frac{C}{\lambda}\|h\|, && (52)
\end{aligned}
$$

where C is a constant that does not depend on η, since $e^{-\eta t} \leqslant 1$. The compactness of the resolvent follows now from the compactness of the embedding of H_ρ^1 into L_ρ^2. □

Proof of proposition 5.1. This is a consequence of the compactness of the resolvent of $\mathcal{L}$, which allows us to use Fredholm's theorem. Recall that we are considering the Poisson equation $\mathcal{L}\phi = f$ where $f \in L_\rho \cap C^\infty(X)$ and centred with respect to the invariant measure $\mu_\beta(d\boldsymbol{x})$.

Set $\mathcal{L}_\eta u = \eta u + \mathcal{L}u$. Fredholm's theorem applies so either the solution of

$$
\left(\frac{1}{\eta} I - \mathcal{L}_\eta^{-1}\right) u = \tilde{h}, \qquad \tilde{h} = \mathcal{L}_\eta^{-1} f/\eta
$$

exists and is unique (and hence, by construction the solution to (47) is unique) or $(\frac{1}{\eta}I - \mathcal{L}_\eta^{-1})u = 0$ admits a non-zero solution. We can rule out the latter option because $(\frac{1}{\eta}I - \mathcal{L}_\eta^{-1})u = 0$ is equivalent to $\mathcal{L}u = 0$; since we know that Ker$\mathcal{L}$ contains only constants and we require the solution to have mean zero, we can conclude that the only solution of the equation $\mathcal{L}u = 0$ is $u = 0$.

6. The white noise limit

Throughout this section C denotes a generic constant and $c(t)$ denotes a generic positive increasing continuous function bounded on compacts [0,T]; both C and $c(t)$ are independent of ϵ, even though they can depend on the coefficients $\{\lambda_i, \alpha_i\}_{i=1,\dots,m}$ and they do depend on the exponent n in estimate (22). To simplify the notation we present the proof in one dimension, i.e. $d = 1$ and we set $\beta = 1$. The proof is exactly the same in arbitrary dimensions. Let $(Q(t), P(t)) \in \mathbb{R} \times \mathbb{R}$ be the solution to system (21), and $(q(t), p(t), z(t))$ be the solution to system (20), then

$$| q(t) - Q(t) | \leqslant \int_0^t | p(s) - P(s) | \, \mathrm{d}s.$$

From (20*c*)

$$\frac{1}{\sqrt{\epsilon}} \int_0^t \mathrm{d}s z_i(s) = -\frac{\sqrt{\epsilon}}{\alpha_i}(z_i(t) - z_i(0)) - \frac{\lambda_i}{\alpha_i}\int_0^t \mathrm{d}s p(s) + \sqrt{\frac{2}{\alpha_i}} W_i(t),$$

so that, setting $\theta_i = \lambda_i^2/\alpha_i$, we have

$$\begin{aligned} p(t) - P(t) = &\int_0^t \left(-\partial_q V(q(s)) + \partial_q V(Q(s))\right) \mathrm{d}s \\ &+ \sum_{i=1}^m \theta_i \int_0^t (P(s) - p(s)) \, \mathrm{d}s - \sqrt{\epsilon} \sum_{i=1}^m \frac{\lambda_i}{\alpha_i}(z_i(t) - z_i(0)). \end{aligned}$$

We use the Lipshitz continuity of $\partial_q V(q)$ together with Hölder's inequality to obtain

$$\begin{aligned} \eta_n(T) &:= E \sup_{t\in[0,T]} \left\{| q(t) - Q(t) |^n + | p(t) - P(t) |^n\right\} \\ &\leqslant CT^{n-1} \int_0^T E \sup_{s\in[0,t]} | q(s) - Q(s) |^n \, \mathrm{d}t \\ &\quad + C\left(\sum_{i=1}^m \theta_i^n\right) T^{n-1} \int_0^T E \sup_{s\in[0,t]} | p(s) - P(s) |^n \, \mathrm{d}t \\ &\quad + C\epsilon^{\frac{n}{2}} \sum_{i=1}^m \left(\frac{\lambda_i}{\alpha_i}\right)^n E \sup_{t\in[0,T]} | z_i(t) - z_i(0) |^n . \end{aligned}$$

From this we deduce

$$\eta_n(T) \leqslant Cc(T) \int_0^T \mathrm{d}t \eta_n(t) + C\epsilon^{\frac{n}{2}} \sum_{i=1}^m E \sup_{t\in[0,T]} | z_i(t) - z_i(0) |^n .$$

From Gronwall's lemma we then have

$$\eta(T) \leqslant C\epsilon^{\frac{n}{2}} \sum_{i=1}^m E \sup_{t\in[0,T]} |z_i(t) - z_i(0)|^n + Cc(T)\epsilon^{\frac{n}{2}} \int_0^T \mathrm{d}t \sum_{i=1}^m E \sup_{s\in[0,t]} |z_i(s) - z_i(0)|^n$$

and the result now follows from proposition 6.1.

Proposition 6.1. *With the same notation and assumptions as in theorem 2.6 the following estimate holds true:*

$$\sum_{i=1}^{m} E \sup_{t\in[0,T]} |z_i(t) - z_i(0)|^n \leqslant Cc(T) \left[\sum_{i=1}^{m} E|z_i(0)|^n + E|p(0)|^n + E|q(0)|^n + 1 \right],$$

where $c(t)$ is a positive increasing continuous function bounded on compacts $[0, T]$.

Proof. From (20c),

$$z_i(t) = \mathrm{e}^{-\frac{\alpha_i}{\epsilon}t} z_i(0) + \int_0^t \mathrm{e}^{-(t-s)\frac{\alpha_i}{\epsilon}} \left(-\frac{\lambda_i}{\sqrt{\epsilon}} p(s)\,\mathrm{d}s + \sqrt{\frac{2\alpha_i}{\epsilon}}\,\mathrm{d}W_i(s) \right). \tag{53}$$

So from (20a), (20b) and (53) we have

$$\begin{aligned} q(t) + p(t) = &-\int_0^t \mathrm{d}s \partial_q V(q(s)) + \int_0^t \mathrm{d}s p(s) + q(0) + p(0) \\ &+ \frac{1}{\sqrt{\epsilon}} \sum_{i=1}^{m} \lambda_i \left[\int_0^t \mathrm{d}s \mathrm{e}^{-\frac{\alpha_i}{\epsilon}s} z_i(0) \right. \\ &\left. + \frac{1}{\sqrt{\epsilon}} \int_0^t \mathrm{d}s \mathrm{e}^{-\frac{s\alpha_i}{\epsilon}} \int_0^s \mathrm{d}u \mathrm{e}^{\frac{u\alpha_i}{\epsilon}} \Big(-\lambda_i p(u)\,\mathrm{d}u + \sqrt{2\alpha_i}\,\mathrm{d}W_i(u) \Big) \right]. \end{aligned}$$

By integration by parts,

$$\begin{aligned} &\int_0^t \mathrm{d}s \mathrm{e}^{-\frac{s\alpha_i}{\epsilon}} \int_0^s \mathrm{d}u \mathrm{e}^{\frac{u\alpha_i}{\epsilon}} \left(-\lambda_i p(u)\,\mathrm{d}u + \sqrt{2\alpha_i}\,\mathrm{d}W_i \right) \\ &\qquad = \frac{\epsilon}{\alpha_i} \int_0^t (\mathrm{e}^{-(t-u)\frac{\alpha_i}{\epsilon}} + 1) \left(-\lambda_i p(u)\,\mathrm{d}u + \sqrt{2\alpha_i}\,\mathrm{d}W_i(u) \right), \end{aligned}$$

hence, using again the Hölder continuity of $V(q)$, we obtain

$$\begin{aligned} \xi_n(T) &:= E \sup_{t\in[0,T]} \left\{ |q(t)|^n + |p(t)|^n \right\} \\ &= Cc(T) \left[\int_0^T \mathrm{d}t \xi_n(t) + E\left(|q(0)|^n + |p(0)|^n \right) + C\epsilon^{\frac{n}{2}} \sum_{i=1}^{m} E|z_i(0)|^n + 1 \right] \end{aligned}$$

and by Gronwall's lemma

$$\xi_n(T) \leqslant C \left[E\left(|q(0)|^n + |p(0)|^n \right) + C\epsilon^{\frac{n}{2}} \sum_{i=1}^{m} E|z_i(0)|^n \right] (1 + c(T)),$$

which implies

$$E \sup_{t\in[0,T]} |p(t)|^n \leqslant \left[E\left(|q(0)|^n + |p(0)|^n \right) + C\epsilon^{\frac{n}{2}} \sum_{i=1}^{m} E|z_i(0)|^n \right] (1 + c(T)). \tag{54}$$

Since by (53) we have

$$E \sup_{t\in[0,T]} |z_i(t)|^n \leqslant C \left(E|z_i(0)|^n + E \sup_{t\in[0,T]} |p(t)|^n + 1 \right), \tag{55}$$

Proposition 6.1 follows from (54) and (55). □

Acknowledgments

The authors thank the anonymous referees for many useful comments on an earlier version of this paper. They are grateful to K Pravda-Starov and to G Stoltz for many useful suggestions. They acknowledge financial support from the EPSRC.

Appendix A. Hypocoercivity

In this appendix we recall some of the main results from the theory of hypocoercivity, as presented in [48]. Throughout this appendix we will use the notation introduced in section 2.1 and in definition 3.1.

Definition A.1 (Coercivity). *With the same notation of definition 3.1, the operator $\mathcal{T}$ is said to be λ-coercive on $\tilde{\mathcal{H}}$ if*

$$(\mathcal{T}h, h)_{\tilde{\mathcal{H}}} \geqslant \lambda \|h\|^2_{\tilde{\mathcal{H}}} \qquad \forall h \in \mathcal{K}^{\perp} \cap D(\mathcal{T}).$$

The following proposition gives an equivalent definition of coercivity.

Proposition A.1. *With the same notation as in definition A.1, $\mathcal{T}$ is λ-coercive on $\tilde{\mathcal{H}}$ iff*

$$\| \mathrm{e}^{-\mathcal{T}t} h \|_{\tilde{\mathcal{H}}} \leqslant \mathrm{e}^{-\lambda t} \| h \|_{\tilde{\mathcal{H}}} \qquad \forall h \in \tilde{\mathcal{H}} \quad \text{and} \quad t \geqslant 0.$$

Theorem A.2. *Let $\mathcal{L}$ be an operator of the form $\mathcal{L} = A^*A + B$, with $B^* = -B$, $\mathcal{K} = \mathrm{Ker}\mathcal{L}$ and assume there exists $N \in \mathbb{N}$ such that*

$$[C_{j-1}, B] = C_j + R_j \quad 1 \leqslant j \leqslant N+1, \qquad C_0 = A, \qquad C_{N+1} = 0. \tag{56}$$

Consider the following assumptions: for $k = 0, \ldots, N+1$

1. *$[A, C_k]$ is relatively bounded with respect to $\{C_j\}_{0\leqslant j\leqslant k}$ and $\{C_jA\}_{0\leqslant j\leqslant k-1}$.*
2. *$[C_k, A^*]$ is relatively bounded with respect to I and $\{C_j\}_{0\leqslant j\leqslant k}$ (here I indicates the identity operator on L^2_ρ).*
3. *R_k is relatively bounded with respect to $\{C_j\}_{0\leqslant j\leqslant k-1}$ and $\{C_jA\}_{0\leqslant j\leqslant k-1}$.*
4. *$\sum_{j=0}^{N} C_j^* C_j$ is κ-coercive for some $\kappa > 0$.*

If assumptions 1–3 are satisfied then there exists a scalar product $((\cdot,\cdot))$ on H^1_ρ defining a norm equivalent to the usual H^1_ρ norm and such that

$$\forall h \in H^1_\rho/\mathcal{K}, \qquad ((h, \mathcal{L}h)) \geqslant K \sum_{j=0}^{N} \|C_j h\|^2, \tag{57}$$

for some constant $K > 0$. Furthermore, if assumption 4 is satisfied, then there exists a constant $\lambda > 0$ such that

$$\forall h \in H^1_\rho/\mathcal{K}, \qquad ((h, \mathcal{L}h)) \geqslant \lambda((h, h)).$$

In particular, $\mathcal{L}$ is hypocoercive in $H^1_\rho/\mathcal{K}$, i.e.

$$\|\mathrm{e}^{-t\mathcal{L}}\|_{H^1_\rho/\mathcal{K} \to H^1_\rho/\mathcal{K}} \leqslant C\mathrm{e}^{-\lambda t}$$

for some $C, \lambda > 0$.

Theorem A.3. *With the same notation as in theorem A.2, if assumptions 1–3 are satisfied then*

$$\|C_k \mathrm{e}^{-t\mathcal{L}} h\| \leqslant \frac{C}{t^{k+\frac{1}{2}}} \|h\|, \qquad \forall k = 0, \ldots, N,$$

for all functions $h \in L^2_\rho$.

Theorem A.4. *Let $V(x) \in C^\infty(\mathbb{R}^d)$ such that $\mu(\mathrm{d}x) = \mathrm{e}^{-V(x)}\,\mathrm{d}x$ is a probability measure on $\mathbb{R}^d$ and assume that $\mathcal{L}$ generates a semigroup on a suitable space of positive functions. Let $\{A_j\}_{1\leqslant j\leqslant M}$ and B be first-order differential operators with smooth coefficients, with $B = -B^*$. Assume there exists $N \in \mathbb{N}$ such that*

$$[C_{j-1}, B] = C_j + R_j \qquad 1 \leqslant j \leqslant N+1, \qquad C_0 = A, C_{N+1} = 0.$$

If, for $0 \leqslant k \leqslant N+1$ the following assumptions are fulfilled

1. *$[A, C_k]$ is pointwise bounded with respect to A.*
2. *$[C_k, A^*]$ is pointwise bounded with respect to I and $\{C_j\}_{0\leqslant j\leqslant k}$.*
3. *R_k is pointwise bounded with respect to $\{C_j\}_{0\leqslant j\leqslant k-1}$.*
4. *$[A, C_k]^*$ is pointwise bounded relatively to I and A.*
5. *there exists a positive constant $\lambda > 0$ such that $\sum_k C_k^* C_k \geqslant \lambda I$ pointwise on $\mathbb{R}^d$ (I is the identity matrix on $\mathbb{R}^d$).*
6. *The probability measure μ satisfies a logarithmic Sobolev inequality.*

Then the Kullback information (39) and the Fisher information (40) decay exponentially fast to zero.

Theorem A.5. *With the same notation as in theorem A.4, let $V(x) \in C^\infty(\mathbb{R}^d)$ be such that $\mu(\mathrm{d}x) = \mathrm{e}^{-V(x)}\,\mathrm{d}x$ is a probability measure on $\mathbb{R}^n$ and assume that $\mathcal{L}$ generates a semigroup on a suitable space of positive functions. If assumptions 1–4 of theorem A.4 are fulfilled, then the following bounds hold*

$$\int h_t \mid C_k \log h_t \mid^2 \mathrm{d}\mu \leqslant \frac{C}{t^{2k+1}} \int h_0 \log h_0 \mathrm{d}\mu \qquad \forall k = 0, \ldots, N,$$

where $h_t = f_t/\rho$ and f_t is the density of the law of the process with generator $-\mathcal{L}$.

Appendix B. Ergodicity

This appendix is devoted to the proof of theorem 2.1. We apply Markov chain techniques [34, 35, 45] to prove ergodicity of the Markov process $x(t) := \{q(t),\ p(t),\ z(t)\}$ given by (7). To be more precise, we will study the ergodic properties of the following SDEs:

$$\dot{q} = p, \tag{58a}$$

$$\dot{p} = -\nabla_q V(q) + r, \tag{58b}$$

$$\dot{r} = -p - r + \dot{W}. \tag{58c}$$

We consider both the case $q \in \mathbb{R}^d$ and $q \in \mathbb{T}^d$ and $p, r \in \mathbb{R}^d$. $L = -\mathcal{L}$ is the generator of the process. The extension to the case $r \in \mathbb{R}^{md}$ is straightforward, so we shall not present it. Throughout this appendix $(\cdot, \cdot)$ denotes the Euclidean inner product. The main result of this appendix is theorem 2.1, which we include here for the reader's convenience.

Theorem B.1 (Ergodicity). *The solution of (58) with $\boldsymbol{x}(t) \in X$ and $V(q) \in C^\infty(\mathbb{T}^d)$ is geometrically ergodic. The same holds true when $\boldsymbol{x}(t) \in Y$, provided that the potential $V(q) \in C^\infty(\mathbb{R}^d)$ satisfies assumption 2.1.*

Following [34] and [33], let $P_t(x, A)$ be the transition kernel of the Markov process $\boldsymbol{x}(t)$. Consider the discretized process $\{x_n\}_{n\in\mathbb{N}}$, obtained by sampling at the rate $T > 0$ and with transition kernel $P(x, A) := P_T(x, A)$.

Lyapunov Condition. There exists a function $G(x) : \mathbb{R}^{3d} \to [1, \infty)$ such that $G(x) \to \infty$ as $\|x\| \to \infty$ and $LG(x) \leqslant -aG(x) + \tilde{d}$ for some $a, \tilde{d} > 0$.

Minorization condition. There exist $T > 0, \eta > 0$ and a probability measure ν, with $\nu(C^c) = 0$ and $\nu(C) = 1$ for some fixed compact set C in the phase space, such that

$$P_T(x, A) \geqslant \eta\nu(A) \qquad \forall A \in \mathcal{B}(\mathbb{R}^{3d}), \quad x \in C.$$

Consider now the set $\mathcal{G} = \{x \in \mathbb{R}^{3d} : G(x) \leqslant \frac{2\tilde{d}/a}{\gamma - \mathrm{e}^{-aT/2}}\}$ for some $\gamma \in (\mathrm{e}^{-aT/2}, 1)$, G, a and $\tilde{d}$ as in the Lyapunov condition. We will use the following result, the proof of which can be found in [35].

Theorem B.2. *If there exists a sampling rate $T > 0$ such that the resulting chain $\{x_n\}_{n\in\mathbb{N}}$ is a Markov chain satisfying the minorization condition on the set $\mathcal{G}$ and there exists a function G satisfying the Lyapunov condition, then the process is ergodic.*

Assumption ($\star$). Let $B_s(y) \in R^{3d}$ be the ball of radius s centred in y. For some fixed compact set C we have

- $P_t(x, A)$ has a density $p_t(x, y)$ which is continuous $\forall(x, y) \in C \times C$, more precisely

$$P_t(x, A) = \int_A p_t(x, y)\,\mathrm{d}y \qquad \forall A \in \mathcal{B}(\mathbb{R}^{3d}) \cap \mathcal{B}(C), \qquad \forall x \in C;$$

- $\forall\delta > 0$ one can find a $\bar{t} = \bar{t}(\delta)$ such that

$$P_{\bar{t}}(x, B_\delta(x^*)) > 0 \qquad \text{for some } x^* \in \mathrm{int}(C), \forall x \in C.$$

We have the following result.

Lemma B.1. *Assumption ($\star$) $\Longrightarrow$ Minorization Condition.*

We shall prove that, under the assumptions of theorems 2.1, B.2 applies, hence system (58) is ergodic.

Proof of theorem B.1. Consider first the case $x(t) \in X$. Let $V(q)$ be a $C^\infty(\mathbb{T}^d)$ potential, $V(q) > -k$ for some positive constant k. Consider the function

$$G(x) = \hat{C} + \frac{B}{2}\|p\|^2 + \frac{C}{2}\|r\|^2 + DV(q) + H(p, r),$$

where B, C, D, H and $\hat{C}$ are positive constants to be chosen. We have that

$$G(x) \geqslant \hat{C} + \frac{B}{2}\|p\|^2 + \frac{C}{2}\|r\|^2 - \frac{H}{2}\|p\|^2 - \frac{H}{2}\|r\|^2 - Dk, \tag{59}$$

so we need $B > H$, $C > H$ and $\hat{C} > Dk$. Moreover,

$$\begin{aligned} LG(x) &= D(\nabla_q V, p) - B(\nabla_q V, p) - H(\nabla_q V, r) + B(r, p) + H\|r\|^2 \\ &\quad - C(p, r) - H\|p\|^2 - C\|r\|^2 - H(p, r) + C \\ &\leqslant H\|r\|^2 + \frac{H}{4}\|\nabla_q V\|^2 - H\|p\|^2 - C\|r\|^2 + C + H\|r\|^2, \end{aligned}$$

where we have chosen $B = D = C + H$. On the other hand, since $V(q) \leqslant K$,

$$G(x) \geqslant -\frac{a}{2}B\|p\|^2 - \frac{a}{2}C\|r\|^2 - aKB - a\frac{H}{2}\|r\|^2 - a\frac{H}{2}\|p\|^2$$

so imposing also $2H - C \leqslant -\frac{a}{2}(C + H)$, $-H \leqslant -\frac{a}{2}(B + H)$ for some $a > 0$, the Lyapunov condition is satisfied. One possible choice is $a = 1/4$, $B = 13/16$, $C = 5/8$ and $H = 3/16$.

Note that from what we have just proven it follows that $\forall l \geqslant 1$ we have

$$LG(x)^l \leqslant -a_l G(x)^l + \tilde{d}_l, \tag{60}$$

for some suitable positive constants a_l and $\tilde{d}_l$. In fact,

$$\partial_{q_i} G(x)^l = lG(x)^{l-1}\partial_{q_i} G(x),$$
$$\partial_{p_i} G(x)^l = lG(x)^{l-1}\partial_{p_i} G(x)$$

and

$$\partial_{r_i}^2 G(x) = \partial_{r_i}\left[lG(x)^{l-1}\partial_{r_i} G(x)\right] = l(l-1)G(x)^{l-2}(\partial_{r_i} G)^2 + lG(x)^{l-1}\partial_{r_i}^2 G(x).$$

Furthermore, using (59), we obtain

$$l(l-1)G(x)^{l-2}(\partial_{r_i} G)^2 \leqslant c_l G(x)^{l-1},$$

so that

$$LG(x)^l \leqslant lG(x)^{l-1} LG(x) + c_l G(x)^{l-1}.$$

Hence, using what we have proven in the case $l = 1$, we obtain (60).

Consider now the case $x(t) \in Y$. We introduce the Lyapunov function

$$\begin{aligned} G(x) = \hat{C} + \frac{A}{2}\|q\|^2 + \frac{B}{2}\|p\|^2 + \frac{C}{2}\|r\|^2 + DV(q) \\ + E(p,q) + F(q,r) + H(p,r) + M(\nabla_q V, p). \end{aligned}$$

Consequently,

$$\nabla_q G = Aq + D\nabla_q V + Ep + Fr + M\nabla^2 V(q)\cdot p,$$
$$\nabla_p G = Bp + Eq + Hr + M\nabla_q V,$$
$$\nabla_r G = Cr + Fq + Hp.$$

Thus,

$$\begin{aligned} LG(x) = {} & A(p,q) + D(\nabla_q V, p) + E\|p\|^2 + F(p,r) - B(\nabla_q V, p) \\ & - E(\nabla_q V, q) - H(\nabla_q V, r) + B(p,r) + E(q,r) + H\|r\|^2 \\ & - C(p,r) - F(p,q) - H\|p\|^2 + M(p, \nabla^2 V(q)\cdot p) \\ & - C\|r\|^2 - F(r,q) - H(p,r) + C - M\|\nabla_q V\|^2 + M(r, \nabla_q V) \end{aligned}$$

From assumption 2.1(iii) it follows that there exist constants $\tilde{\beta}$ and $\tilde{\sigma}$ such that

$$\tilde{\sigma}\|q\|^2 - \tilde{\beta}\|\nabla_q V\|^2 \to +\infty \text{ as } \|q\|^2 \to +\infty.$$

Hence, it follows that G satisfies the Lyapunov condition. Also in this case, one can prove that the Lyapunov condition holds for $G(x)^l$, $l \geqslant 1$, as well.

As for assumption ($\star$), first of all let us note that, since the operator $\partial_t + \mathcal{L}$ is hypoelliptic, the transition probability has a density; because the SDE we consider has time independent coefficients the density is C^∞ provided that $V(q) \in C^\infty$ [38]. Moreover, studying the control problem associated with $\mathrm{d}x = b(x)\,\mathrm{d}t + \sigma\,\mathrm{d}w$, namely $\mathrm{d}X = b(X)\,\mathrm{d}t + \sigma\,\mathrm{d}U$ where $U(t)$ is a smooth control, and using the Stroock–Varadhan support Theorem, we can prove that $P_t(x, A) > 0\,\forall x \in \mathbb{R}^{3d}$, $t > 0$ and for any open $A \in \mathbb{R}^{3d}$.

Consider now the set $\mathcal{G}_l = \{g : \mathbb{R}^{3d} \to \mathbb{R}$, measurable $:|\ g(x)\ | \leqslant G(x)^l\}$. Then there exist constants $k = k(l)$ and $\lambda = \lambda(l)$, such that $\forall g \in \mathcal{G}_l$

$$|E^{x_0} g(x(t)) - \rho(g)| \leqslant kG(x_0)\mathrm{e}^{-\lambda t}, \qquad t \geqslant 0, \tag{61}$$

that is, the process is geometrically ergodic, see [33, theorem 3.2] or [35, theorem 15.0.1].

References

[1] Ying S C, Ala-Nissila T and Ferrando R 2002 Collective and single particle diffusion on surfaces *Adv. Phys.* **51** 949–1078
[2] Ariel G and Vanden-Eijnden E 2007 Testing transition state theory on Kac–Zwanzig model *J. Stat. Phys.* **126** 43–73
[3] Ariel G and Vanden-Eijnden E 2009 A strong limit theorem in the Kac–Zwanzig model *Nonlinearity* **22** 145–62
[4] Arnold A, Markowich P, Toscani G and Unterreiter A 2001 On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations *Commun. Partial Diff. Eqns* **26** 43–100
[5] Battacharya R 1985 A central limit theorem for diffusions with periodic coefficients *Ann. Probab.* **13** 385–96
[6] Cerrai S and Freidlin M 2006 On the Smoluchowski–Kramers approximation for a system with an infinite number of degrees of freedom *Probab. Theory Relat. Fields* **135** 363–94
[7] Cerrai S and Freidlin M 2006 Smoluchowski–Kramers approximation for a general class of SPDEs *J. Evol. Eqns* **6** 657–89
[8] Dym H and McKean H P 1976 *Gaussian Processes, Function Theory, and the Inverse Spectral Problem* (*Probability and Mathematical Statistics* vol 31) (New York: Academic) (Harcourt Brace Jovanovich Publishers)
[9] Eckmann J-P and Hairer M 2000 Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators *Commun. Math. Phys.* **212** 105–64
[10] Eckmann J-P, Pillet C-A and Rey-Bellet L 1999 Entropy production in nonlinear, thermally driven Hamiltonian systems *J. Stat. Phys.* **95** 305–31
[11] Eckmann J-P, Pillet C-A and Rey-Bellet L 1999 Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures *Commun. Math. Phys.* **201** 657–97
[12] Ethier S N and Kurtz T G 1986 *Markov Processes* (*Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*) (New York: Wiley)
[13] Ford G W, Kac M and Mazur P 1965 Statistical mechanics of assemblies of coupled oscillators *J. Math. Phys.* **6** 504–15
[14] Givon D, Kupferman R and Stuart A M 2004 Extracting macroscopic dynamics: model problems and algorithms *Nonlinearity* **17** R55–127
[15] Grigolini P 1982 A generalized Langevin equation for dealing with nonadditive fluctuations *J. Stat. Phys.* **27** 283–316
[16] Hairer M 2009 Ergodic properties of a class of non-Markovian processes *Trends in Stochastic Analysis* (*London Mathematical Society Lecture Note Series* vol 353) (Cambridge: Cambridge University Press) pp 65–98
[17] Hairer M and Pavliotis G A 2008 From ballistic to diffusive behavior in periodic potentials *J. Stat. Phys.* **131** 175–202
[18] Hairer M and Pavliotis G A 2004 Periodic homogenization for hypoelliptic diffusions *J. Stat. Phys.* **117** 261–79
[19] Hald O H and Kupferman R 2002 Asymptotic and numerical analyses for mechanical models of heat baths *J. Stat. Phys.* **106** 1121–84
[20] Hanggi P, Talkner P and Borkovec M 1990 Reaction-rate theory: fifty years after Kramers *Rev. Mod. Phys.* **62** 251–341
[21] Helffer B and Nier F 2005 *Hypoelliptic Estimates and Spectral Theory for Fokker–Planck Operators and Witten Laplacians* (*Lecture Notes in Mathematics* vol 1862) (Berlin: Springer)
[22] Hérau F 2007 Short and long time behavior of the Fokker–Planck equation in a confining potential and applications *J. Funct. Anal.* **244** 95–118
[23] Hörmander L 1967 Hypoelliptic second order differential equations *Acta Math.* **119** 147–71
[24] Jakšić V and Pillet C-A 1997 Ergodic properties of the non-Markovian Langevin equation *Lett. Math. Phys.* **41** 49–57
[25] Jakšić V and Pillet C-A 1997 Spectral theory of thermal relaxation *J. Math. Phys.* **38** 1757–80 (Quantum problems in condensed matter physics)
[26] Jakšić V and Pillet C-A 1998 Ergodic properties of classical dissipative systems: I *Acta Math.* **181** 245–82
[27] Kipnis C and Varadhan S R S 1986 Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions *Commun. Math. Phys.* **104** 1–19
[28] Kupferman R 2004 Fractional kinetics in Kac–Zwanzig heat bath models *J. Stat. Phys.* **114** 291–326
[29] Kupferman R and Stuart A M 2004 Fitting SDE models to nonlinear Kac–Zwanzig heat bath models *Physica* D **199** 279–316
[30] Kupferman R, Stuart A M, Terry J R and Tupper P F 2002 Long-term behaviour of large mechanical systems with random initial data *Stochastics Dyn.* **2** 533–62

[31] Landim C 2003 Central limit theorem for Markov processes *From Classical to Modern Probability* (*Progr. Probab.* vol 54) (Basel: Birkhäuser) pp 145–205

[32] Markowich P A and Villani C 2000 On the trend to equilibrium for the Fokker–Planck equation: an interplay between physics and functional analysis *Mat. Contemp.* **19** 1–29

[33] Mattingly J C, Stuart A M and Higham D J 2002 Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise *Stochastic Process. Appl.* **101** 185–232

[34] Mattingly J C and Stuart A M 2002 Geometric ergodicity of some hypo-elliptic diffusions for particle motions *Markov Process. Relat. Fields* **8** 199–214

[35] Meyn S P and Tweedie R L 1993 *Markov Chains and Stochastic Stability* (*Communications and Control Engineering Series*) (London: Springer)

[36] Mori H 1965 A continued-fraction representation of the time-correlation functions *Progr. Theor. Phys.* **34** 399–416

[37] Nelson E 1967 *Dynamical Theories of Brownian Motion* (Princeton, NJ: Princeton University Press)

[38] Nualart D 2006 *The Malliavin Calculus and Related Topics* (*Probability and its Applications*) 2nd edn (Berlin: Springer)

[39] Ottobre M, Pavliotis G A and Pravda-Starov K 2010 Exponential return to equilibrium for hypoelliptic quadratic systems (in preparation)

[40] Papanicolaou G C and Varadhan S R S 1985 Ornstein–Uhlenbeck process in a random potential *Commun. Pure Appl. Math.* **38** 819–34

[41] Pardoux E 1999 Homogenization of linear and semilinear second order parabolic pdes with periodic coefficients: a probabilistic approach *J. Funct. Anal.* **167** 498–520

[42] Pavliotis G A and Stuart A M 2008 *Multiscale Methods* (*Texts in Applied Mathematics* vol 53) (New York: Springer) (Averaging and homogenization)

[43] Reed M and Simon B 1980 *Methods of Modern Mathematical Physics: I. Functional Analysis* (New York: Academic)

[44] Rey-Bellet L 2006 Open classical systems *Open Quantum Systems: II* (*Lecture Notes in Mathematics* vol 1881) (Berlin: Springer) pp 41–78

[45] Rey-Bellet L and Thomas L E 2002 Exponential convergence to non-equilibrium stationary states in classical statistical mechanics *Commun. Math. Phys.* **225** 305–29

[46] Snook I 2007 *The Langevin and Generalized Langevin Approach to the Dynamics of Atomic, Polymeric and Colloidal Systems* (Amsterdam: Elsevier)

[47] Villani C 2003 *Topics in Optimal Transportation* (*Graduate Studies in Mathematics* vol 58) (Providence, RI: American Mathematical Society)

[48] Villani C 2009 Hypocoercivity *Mem. Am. Math. Soc.* **202** iv+141

[49] Zwanzig R 1973 Nonlinear generalized Langevin equations *J. Stat. Phys.* **9** 215–20