CrossMark

Homogenization for a Class of Generalized Langevin Equations with an Application to Thermophoresis

Soon Hoe Lim[1,2] · **Jan Wehr[3,4]**

Abstract

We study a class of systems whose dynamics are described by generalized Langevin equations with state-dependent coefficients. We find that in the limit, in which all the characteristic time scales vanish at the same rate, the position variable of the system converges to a homogenized process, described by an equation containing additional drift terms induced by the noise. The convergence results are obtained using the main result in Hottovy et al. (Commun Math Phys 336(3):1259–1283, 2015), whose version is proven here under a weaker spectral assumption on the damping matrix. We apply our results to study thermophoresis of a Brownian particle in a non-equilibrium heat bath.

1 Introduction

From physical sciences to social sciences, many phenomena are modeled by noisy dynamical systems. In many such systems, several widely separated time scales are present. The system obtained in the homogenization limit, in which the fast time scales go to zero, is simpler than the original one, while often retaining the essential features of its dynamics [1–4]. On the other hand, the different fast time scales may compete and this competition is reflected in the homogenized equations.

✉ Soon Hoe Lim
  shoelim@math.arizona.edu

  Jan Wehr
  wehr@math.arizona.edu

1 Program in Applied Mathematics, University of Arizona, Tucson, USA

2 Present Address: Nordita, KTH Royal Institute of Technology and Stockholm University, Roslagstullsbacken 23, 106 91 Stockholm, Sweden

3 Department of Mathematics and Program in Applied Mathematics, University of Arizona, Tucson, USA

4 Present Address: Department of Mathematics, University of Arizona, Tucson, AZ 85721, USA

Of particular interest is the model of a Brownian particle interacting with the environment [5]. The usual model for such system neglects the memory effects, representing the interaction of the particle with the environment as a sum of an instantaneous damping force and a white noise. Although such an idealized model generally gives a good approximate description of the dynamics of the particle, there are situations where the memory effects play an important role, for instance when the particle is subject to a hydrodynamic interaction [6], or when the particle is an atom embedded in a condensed-matter heat bath [7]. In addition, the stochastic forcing introduced by the environment is often more accurately modeled by a colored noise than by white noise.

In this paper, we study a class of generalized Langevin equations (GLEs), with state-dependent drift and diffusion coefficients, driven by a colored noise. They provide a realistic description of the dynamics of a classical Brownian particle in an inhomogeneous environment; their solutions are not Markov processes. We are interested in the limiting behavior of the particle when the characteristic time scales become small and in how competition of the time scales, as well as inhomogeneity of the environment, impact its limiting dynamics. The main mathematical result of this paper is Theorem 2, in which we derive the homogenization limit for a general class of non-Markovian systems. Special cases are studied in some details to obtain more explicit results. Their physical relevance is illustrated by an application to thermophoresis models.

The paper is organized as follows. In Sect. 2, we introduce and discuss a class of GLEs, as well as its two sub-classes, to be studied in this paper. In Sect. 3, we revisit the Smoluchowski–Kramers limit for a class of SDEs with state-dependent drift and diffusion coefficients, under a weaker assumption on the spectrum of the damping matrix than that used in earlier work [8]. Using this result of Sect. 3, we study homogenization for the GLEs in Sect. 4. We specialize the study to two sub-classes of models in Sect. 5. In Sect. 6, we apply the results obtained in the previous sections to study thermophoresis of a Brownian particle in a non-equilibrium heat bath. We end the paper by giving the conclusions and final remarks in Sect. 7. The appendices provide some technical results used in the main paper, as well as physical motivation for the form of the GLEs studied here. In Appendix A we provide a variant of a (heuristic) derivation of the equations studied in this paper from Hamiltonian model of a particle interacting with a system of harmonic oscillators. Appendix B contains a sketch of the proof of Theorem 1.

2 Generalized Langevin Equations (GLEs)

2.1 GLEs as Non-Markovian Models

We consider a class of non-Markovian Langevin equations, with state-dependent coefficients, that describe the dynamics of a particle moving in a force field and interacting with the environment. Let $\boldsymbol{x}_t \in \mathbb{R}^d$, $t \geq 0$, be the position of the particle. The evolution of position, $\boldsymbol{x}_t$, is given by the solution to the following stochastic integro-differential equation (SIDE):

$$m\ddot{\boldsymbol{x}}_t = \boldsymbol{F}(\boldsymbol{x}_t) - \boldsymbol{g}(\boldsymbol{x}_t) \int_0^t \boldsymbol{\kappa}(t-s)\boldsymbol{h}(\boldsymbol{x}_s)\dot{\boldsymbol{x}}_s ds + \boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{\xi}_t, \tag{1}$$

with the initial conditions (here the initial time is chosen to be $t = 0$):

$$\boldsymbol{x}_0 = \boldsymbol{x}, \quad \dot{\boldsymbol{x}}_0 = \boldsymbol{v}. \tag{2}$$

The initial conditions x and v are random variables independent of the process $\{\xi_t : t \geq 0\}$. Our motivation to study the SIDE (1) is that study of microscopic dynamics leads naturally to equations of this form (see Appendix A).

Here and throughout the paper, overdot denotes derivative with respect to time t, the superscript * denotes conjugate transposition of matrices or vectors and E denotes expectation. In the SIDE (1), $m > 0$ is the mass of the particle, the matrix-valued functions $g : \mathbb{R}^d \to \mathbb{R}^{d \times q}$, $h : \mathbb{R}^d \to \mathbb{R}^{q \times d}$ and $\sigma : \mathbb{R}^r \to \mathbb{R}^{d \times r}$ are the state-dependent coefficients of the equation, and $F : \mathbb{R}^d \to \mathbb{R}^d$ is a force field acting on the particle. Here d, q and r are, possibly distinct, positive integers. The second term on the right hand side of (1) represents the drag experienced by the particle and the last term models the noise.

The matrix-valued function $\kappa : \mathbb{R} \to \mathbb{R}^{q \times q}$ is a memory function which is *Bohl*, i.e. the matrix elements of $\kappa(t)$ are finite linear combinations of the functions of the form $t^k e^{\alpha t} \cos(\omega t)$ and $t^k e^{\alpha t} \sin(\omega t)$, where k is an integer and α and ω are real numbers. For properties of Bohl functions, we refer to Chapter 2 of [9]. The noise process ξ_t is an r-dimensional mean zero stationary real-valued Gaussian vector process having a Bohl covariance function, $R(t) := E\xi_t \xi_0^* = R^*(-t)$, and, therefore, its spectral density, $S(\omega) := \int_{-\infty}^{\infty} R(t) e^{-i\omega t} dt$, is a rational function [10].

The SIDE (1) is a non-Markovian Langevin equation, since its solution at time t depends on the entire past. Two of its terms are different than those in the usual Langevin equations. One of them is the drag term, which here involves an integral over the particle's past velocities with a memory kernel $\kappa(t - s)$. It describes the state-dependent dissipation which comprises the back-action effects of the environment up to current time. The other term, involving a Gaussian colored noise ξ_t, is a multiplicative noise term, also arising from interaction of the particle with the environment. Therefore, (1) is a generalized Langevin equation (GLE), which in its most basic form was first introduced by Mori in [11] and subsequently used to model many systems in statistical physics [12–14].

As remarked by van Kampen in [15], "Non-Markov is the rule, Markov is the exception". Therefore, it is not surprising that non-Markovian equations (including those of form (1)) find numerous applications and thus have been studied widely in the mathematical, physical and engineering literature. See, for instance [16,17] for surveys of non-Markovian processes, [18–20] for physical applications and [21] for asymptotic analysis.

Note that the Gaussian process ξ_t which drives the SIDE (1) is not assumed to be Markov. The assumptions we made on its covariance will allow us to present it as a projection of a Markov process in a (typically higher-dimensional) space. This approach, which originated in stochastic control theory [22], is called *stochastic realization*. We describe it in detail below.

Let $\Gamma_1 \in \mathbb{R}^{d_1 \times d_1}$, $M_1 \in \mathbb{R}^{d_1 \times d_1}$, $C_1 \in \mathbb{R}^{q \times d_1}$, $\Sigma_1 \in \mathbb{R}^{d_1 \times q_1}$, $\Gamma_2 \in \mathbb{R}^{d_2 \times d_2}$, $M_2 \in \mathbb{R}^{d_2 \times d_2}$, $C_2 \in \mathbb{R}^{r \times d_2}$, $\Sigma_2 \in \mathbb{R}^{d_2 \times q_2}$ be constant matrices, where d_1, d_2, q_1, q_2, q and r are positive integers. In this paper, we study the class of SIDE (1), with the memory function defined in terms of the triple (Γ_1, M_1, C_1) of matrices as follows:

$$\kappa(t) = C_1 e^{-\Gamma_1 |t|} M_1 C_1^*. \tag{3}$$

The noise process is the mean zero, stationary Gaussian vector process, whose covariance will be expressed in terms of the triple (Γ_2, M_2, C_2). More precisely, we define it as:

$$\xi_t = C_2 \beta_t, \tag{4}$$

where β_t is the solution to the Itô SDE:

$$d\beta_t = -\Gamma_2 \beta_t dt + \Sigma_2 dW_t^{(q_2)}, \tag{5}$$

with the initial condition, $\boldsymbol{\beta}_0$, normally distributed with zero mean and covariance $\boldsymbol{M}_2$. Here, $\boldsymbol{W}_t^{(q_2)}$ denotes a q_2-dimensional Wiener process and is independent of $\boldsymbol{\beta}_0$. Throughout the paper the dimension of the Wiener process will be specified by the superscript.

For $i = 1, 2$, the matrix $\boldsymbol{\Gamma}_i$ is *positive stable*, i.e. all its eigenvalues have positive real parts and $\boldsymbol{M}_i = \boldsymbol{M}_i^* > 0$ satisfies the following Lyapunov equation:

$$\boldsymbol{\Gamma}_i \boldsymbol{M}_i + \boldsymbol{M}_i \boldsymbol{\Gamma}_i^* = \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_i^*. \tag{6}$$

It follows from positive stability of $\boldsymbol{\Gamma}_i$ that this equation indeed has a unique solution [23]. The covariance matrix, $\boldsymbol{R}(t) \in \mathbb{R}^{r \times r}$, of the noise process is therefore expressed in terms of the matrices $(\boldsymbol{\Gamma}_2, \boldsymbol{M}_2, \boldsymbol{C}_2)$ as follows:

$$\boldsymbol{R}(t) = \boldsymbol{C}_2 e^{-\boldsymbol{\Gamma}_2 |t|} \boldsymbol{M}_2 \boldsymbol{C}_2^*, \tag{7}$$

and therefore the triple $(\boldsymbol{\Gamma}_2, \boldsymbol{M}_2, \boldsymbol{C}_2)$ completely specifies the probability distribution of $\boldsymbol{\xi}_t$. It is worth mentioning that the triples that specify the memory function in (3) and the noise process in (4) are only unique up to the following transformations:

$$(\boldsymbol{\Gamma}_i' = \boldsymbol{T}_i \boldsymbol{\Gamma}_i \boldsymbol{T}_i^{-1}, \boldsymbol{M}_i' = \boldsymbol{T}_i \boldsymbol{M}_i \boldsymbol{T}_i^*, \boldsymbol{C}_i' = \boldsymbol{C}_i \boldsymbol{T}_i^{-1}), \tag{8}$$

where $i = 1, 2$ and the $\boldsymbol{T}_i$ are invertible matrices of appropriate dimensions.

The triple $(\boldsymbol{\Gamma}_2, \boldsymbol{M}_2, \boldsymbol{C}_2)$ above is called a *(weak) stochastic realization* of the covariance matrix $\boldsymbol{R}(t)$ in the well established theory of stochastic realization, which is concerned with solving the inverse problem of stationary covariance generation (see [24,25]). Any zero mean stationary Gaussian process, $\boldsymbol{\xi}_t'$, having a Bohl covariance function, can be realized as a projection of a Gaussian Markov process in the above way. Let us remark that Gaussian processes with Bohl covariance functions are precisely those with rational spectral density [10].

Our approach allows us to consider the most general Gaussian noises that can be realized in a finite-dimensional state space in the above way (i.e. as a linear transformation of a Gaussian Markov process). In fact, the condition on the covariance function to have entries in the Bohl class is necessary and sufficient for solvability of the problem of stochastic realization of stationary Gaussian processes. We refer to the propositions and theorems on page 303-308 of [10] for a brief exposition of stochastic realization problems.

Remark 1 Physically, the choice of the matrices $\boldsymbol{\Gamma}_2$, $\boldsymbol{M}_2$, $\boldsymbol{C}_2$ specifies the characteristic time scales (eigenvalues of $\boldsymbol{\Gamma}_2^{-1}$) present in the environment, introduces the initial state of a stationary Markovian Gaussian noise and selects the parts of the prepared Markovian noise that are (partially) observed, respectively. In other words, we have assumed that the noise in the SIDE (1) is realized or "experimentally prepared" by the above triple of matrices.

For our homogenization study of Eq. (1) we need the *effective damping constant*,

$$\boldsymbol{K}_1 := \int_0^\infty \boldsymbol{\kappa}(t) dt = \boldsymbol{C}_1 \boldsymbol{\Gamma}_1^{-1} \boldsymbol{M}_1 \boldsymbol{C}_1^* \in \mathbb{R}^{q \times q}, \tag{9}$$

and the *effective diffusion constant*,

$$\boldsymbol{K}_2 := \int_0^\infty \boldsymbol{R}(t) dt = \boldsymbol{C}_2 \boldsymbol{\Gamma}_2^{-1} \boldsymbol{M}_2 \boldsymbol{C}_2^* \in \mathbb{R}^{r \times r}, \tag{10}$$

to be invertible (see Sect. 2.2). This is equivalent to the matrices $\boldsymbol{C}_i$ having full rank. Homogenization for a class of systems with vanishing effective damping and/or diffusion constant [26] will be explored in our future work.

With the above definitions of memory kernel and noise process, the SIDE (1) becomes:

$$m\ddot{x}_t = F(x_t) - g(x_t) \int_0^t C_1 e^{-\Gamma_1(t-s)} M_1 C_1^* h(x_s)\dot{x}_s ds + \sigma(x_t)C_2\beta_t, \qquad (11)$$

where β_t is the solution to the SDE (5). To illustrate the results of the general study in important special cases (which will also be used later in applications), we consider two representative classes of SIDE (1). The driving Gaussian colored noise is Markovian in the first class and non-Markovian in the second. We set $d = d_1 = d_2 = q_1 = q_2 = q = r$ in the following examples.

(i) *Example of a SIDE driven by a Markovian colored noise.* The memory kernel is given by an exponential function, i.e.

$$\kappa(t-s) = \kappa_1(t-s) := Ae^{-A|t-s|}, \qquad (12)$$

where $A \in \mathbb{R}^{d \times d}$ is a constant diagonal matrix with positive eigenvalues. The driving noise is the Ornstein–Uhlenbeck (OU) process, $\xi_t = \eta_t \in \mathbb{R}^d$, i.e. a mean zero stationary Gaussian process which is the solution to the SDE:

$$d\eta_t = -A\eta_t dt + A dW_t^{(d)}. \qquad (13)$$

In order for the process η_t to be stationary, the initial condition has to be distributed according to the (unique) stationary measure of the Markov process defined by the above equation, i.e. $\eta_0 = \eta$ is normally distributed with zero mean and covariance $A/2$. The mean and the covariance of η_t are given by:

$$E[\eta_t] = 0, \quad E[\eta_t \eta_s^*] = \frac{1}{2}\kappa_1(t-s), \quad s, t \geq 0. \qquad (14)$$

The resulting SIDE reads:

$$m\ddot{x}_t = F(x_t) - g(x_t) \int_0^t \kappa_1(t-s)h(x_s)\dot{x}_s ds + \sigma(x_t)\eta_t. \qquad (15)$$

Let us note that Ornstein–Uhlenbeck processes are the only stationary, ergodic, Gaussian, Markov processes with continuous covariance functions [27]. When all diagonal entries of A go to infinity, the OU process approaches the white noise. For details on OU processes, see for instance [27] and Sect. 2 of [28].

(ii) *Example of a SIDE driven by a non-Markovian colored noise.* The memory kernel is given by an oscillatory function whose amplitude is exponentially decaying, i.e. $\kappa(t-s) := \kappa_2(t-s)$, a diagonal matrix with the diagonal entries:

$$(\kappa_2)_{ii}(t-s) := \begin{cases} \frac{1}{\tau_{ii}} e^{-\omega_{ii}^2 \frac{|t-s|}{2\tau_{ii}}} \left[\cos\left(\frac{\omega_{ii}^0}{\tau_{ii}}(t-s)\right) + \frac{\omega_{ii}^1}{2}\sin\left(\frac{\omega_{ii}^0}{\tau_{ii}}|t-s|\right) \right], & \text{if } |\omega_{ii}| < 2 \\ \frac{1}{\tau_{ii}} e^{-\omega_{ii}^2 \frac{|t-s|}{2\tau_{ii}}} \left[\cosh\left(\frac{\tilde{\omega}_{ii}^0}{\tau_{ii}}(t-s)\right) + \frac{\tilde{\omega}_{ii}^1}{2}\sinh\left(\frac{\tilde{\omega}_{ii}^0}{\tau_{ii}}|t-s|\right) \right], & \text{if } |\omega_{ii}| > 2, \end{cases} \qquad (16)$$

where for $i = 1, \ldots, d$, τ_{ii} is a positive constant, ω_{ii} is a real constant, $\omega_{ii}^0 := \omega_{ii}\sqrt{1 - \omega_{ii}^2/4}$, $\tilde{\omega}_{ii}^0 := \omega_{ii}\sqrt{\omega_{ii}^2/4 - 1}$, $\omega_{ii}^1 := \omega_{ii}/\sqrt{1 - \omega_{ii}^2/4}$, and $\tilde{\omega}_{ii}^1 := \omega_{ii}/\sqrt{\omega_{ii}^2/4 - 1}$.

Let τ be constant diagonal matrix with the positive eigenvalues $(\tau_{jj})_{j=1}^d$, Ω be constant diagonal matrix with the real eigenvalues $(\omega_{jj})_{j=1}^d$, Ω_0 be constant $d \times d$ diagonal matrix

with the eigenvalues $\omega_{jj}\sqrt{1-\omega_{jj}^2/4}$ (if $|\omega_{jj}| < 2$) and $i\omega_{jj}\sqrt{\omega_{jj}^2/4-1}$ (if $|\omega_{jj}| > 2$), and $\boldsymbol{\Omega}_1$ be constant $d \times d$ diagonal matrix with the eigenvalues $\omega_{jj}/\sqrt{1-\omega_{jj}^2/4}$ (if $|\omega_{jj}| < 2$) and $-i\omega_{jj}/\sqrt{\omega_{jj}^2/4-1}$ (if $|\omega_{jj}| > 2$), where i is the imaginary unit.

The driving noise is given by the harmonic noise process, $\boldsymbol{\xi}_t = \boldsymbol{h}_t \in \mathbb{R}^d$, i.e. a mean zero stationary Gaussian process which is the solution to the SDE system:

$$\tau d\boldsymbol{h}_t = \boldsymbol{u}_t dt, \tag{17}$$

$$\tau d\boldsymbol{u}_t = -\boldsymbol{\Omega}^2 \boldsymbol{u}_t dt - \boldsymbol{\Omega}^2 \boldsymbol{h}_t dt + \boldsymbol{\Omega}^2 d\boldsymbol{W}_t^{(d)}, \tag{18}$$

with the initial conditions, $\boldsymbol{h}_0$ and $\boldsymbol{u}_0$, distributed according to the (unique) stationary measure of the above SDE system. The mean and the covariance of $\boldsymbol{h}_t$ are given by:

$$E[\boldsymbol{h}_t] = \boldsymbol{0}, \quad E[\boldsymbol{h}_t \boldsymbol{h}_s^*] = \frac{1}{2}\kappa_2(t-s), \quad s, t \geq 0. \tag{19}$$

Note that $\boldsymbol{h}_t$ is not a Markov process (but the process $(\boldsymbol{h}_t, \boldsymbol{u}_t)$ is).

The resulting SIDE reads:

$$m\ddot{\boldsymbol{x}}_t = \boldsymbol{F}(\boldsymbol{x}_t) - \boldsymbol{g}(\boldsymbol{x}_t)\int_0^t \kappa_2(t-s)\boldsymbol{h}(\boldsymbol{x}_s)\dot{\boldsymbol{x}}_s ds + \boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{h}_t. \tag{20}$$

The harmonic noise is an approximation of the white noise, smoother than the Ornstein–Uhlenbeck process. It can be shown that in the limit $\omega_{ii} \to \infty$ (for all i) the process $\boldsymbol{h}_t$ converges to the Ornstein–Uhlenbeck process whose ith component process has correlation time τ_{ii}, whereas in the limit $\tau_{ii} \to 0$ (for all i) the process $\boldsymbol{h}_t$ converges to the white noise. For detailed properties of harmonic noise process, see for instance [29] or the Appendix in [30]. We remark that the harmonic noise is one of the simplest examples of a non-Markovian process and its use as the driving noise in the SIDE (1) is a natural choice that models the environment as a bath of damped harmonic oscillators [31].

Remark 2 Note that in the SIDEs for the above two sub-classes, the dimension of the driving Wiener process, $\boldsymbol{W}_t^{(d)}$, is the same as that of the colored noise processes $\boldsymbol{\eta}_t$ and $\boldsymbol{h}_t$, as well as the processes, $\boldsymbol{x}_t$ and $\boldsymbol{v}_t$. One could as well consider realizing the noise processes using a driving Wiener process of different dimension. Our choice of working with the same dimensions is for the sake of convenience as it will help to simplify the exposition.

Remark 3 Without loss of generality (due to (8)), we have taken $\boldsymbol{A}$ and $\boldsymbol{\Omega}$ to be diagonal.

Remark 4 In cases of particular interest in statistical physics, the triples $(\boldsymbol{\Gamma}_i, \boldsymbol{M}_i, \boldsymbol{C}_i)$ coincide, up to the transformations in (8) (for $i = 1, 2$), $\boldsymbol{h} = \boldsymbol{g}^*$ and $\boldsymbol{g}$ is proportional to $\boldsymbol{\sigma}$, with the proportionality factor equals $k_B T$, where k_B denotes the Boltzmann constant and $T > 0$ is temperature of the environment (see Appendix A). In this case, we have $d_1 = d_2$ and $q = r$. In particular, for the two sub-classes above we have

$$E[\boldsymbol{\eta}_t^0 (\boldsymbol{\eta}_s^0)^*] = \frac{1}{2}\kappa_1(t-s) \tag{21}$$

for the first sub-class and

$$E[\boldsymbol{h}_t^0 (\boldsymbol{h}_s^0)^*] = \frac{1}{2}\kappa_2(t-s) \tag{22}$$

for the second sub-class. In such cases, the SIDEs describe a particle interacting with an equilibrium heat bath at a temperature T, whose dynamics satisfy the fluctuation–dissipation relation [13,32].

2.2 Homogenization of SIDEs: Discussion and Statement of the Problem

There are three characteristic time scales defining the non-Markovian dynamics described by the SIDE (1):

(i) the inertial time scale, λ_m, proportional to m, whose physical significance is the relaxation time of the particle velocity process $v_t := \dot{x}_t$. The limit $\lambda_m \to 0$ is equivalent to the limit $m \to 0$;
(ii) the memory time scale, λ_κ, defined as the inverse rate of exponential decay of the memory kernel $\kappa(t - s)$;
(iii) the noise correlation time scale, λ_ξ.

For the purpose of general multiscale analysis, we set $m = m_0 \epsilon^\mu$, $\lambda_\kappa = \tau_\kappa \epsilon^\theta$ and $\lambda_\xi = \tau_\xi \epsilon^\nu$, where $\epsilon > 0$ is a parameter which will be taken to zero, $m_0, \tau_\kappa, \tau_\xi$ are (fixed) proportionality constants, and μ, θ, ν are positive constants (exponents), specifying the orders at which the time scales $\lambda_m, \lambda_\kappa, \lambda_\xi$ vanish respectively. We consider a family of SIDEs, parametrized by ϵ, with the inertial time scale λ_m proportional to $m_0 \epsilon^\mu$, memory time scale $\lambda_\kappa = \tau_\kappa \epsilon^\theta$ and noise correlation time scale $\lambda_\xi = \tau_\xi \epsilon^\nu$, to be defined in the following.

We replace m with $m_0 \epsilon^\mu$, Γ_1 with $\Gamma_1/(\tau_\kappa \epsilon^\theta)$, M_1 with $M_1/(\tau_\kappa \epsilon^\theta)$, and x_t with x_t^ϵ in (11). Also, we substitute Γ_2 with $\Gamma_2/(\tau_\xi \epsilon^\nu)$, Σ_2 with $\Sigma_2/(\tau_\xi \epsilon^\nu)$, and β_t with β_t^ϵ in (5). The SIDE (11) then becomes:

$$m_0 \epsilon^\mu \ddot{x}_t^\epsilon = F(x_t^\epsilon) - \frac{g(x_t^\epsilon)}{\tau_\kappa \epsilon^\theta} \int_0^t C_1 e^{-\frac{\Gamma_1}{\tau_\kappa \epsilon^\theta}(t-s)} M_1 C_1^* h(x_s^\epsilon) \dot{x}_s^\epsilon ds + \sigma(x_t^\epsilon) C_2 \beta_t^\epsilon, \quad (23)$$

with the initial conditions, $x_0^\epsilon = x$ and $v_0^\epsilon = v$, where β_t^ϵ is the Ornstein-Uhlenbeck process, with correlation time $\tau_\xi \epsilon^\nu$, satisfying the SDE:

$$d\beta_t^\epsilon = -\frac{\Gamma_2}{\tau_\xi \epsilon^\nu} \beta_t^\epsilon dt + \frac{\Sigma_2}{\tau_\xi \epsilon^\nu} dW_t^{(q_2)}, \quad (24)$$

with the initial condition, β_0^ϵ, normally distributed with zero mean and covariance of $M_2/(\tau_\xi \epsilon^\nu)$.

We will also perform similar analysis on the two sub-classes of SIDE, in which case:

(i) the SIDE (15) becomes (with $m := m_0 \epsilon^\mu$, A in the formula for κ_1 replaced by $A/(\tau_\kappa \epsilon^\theta)$, A in (13) replaced by $A/(\tau_\eta \epsilon^\nu)$, x_t replaced by x_t^ϵ and η_t replaced by η_t^ϵ):

$$m_0 \epsilon^\mu \ddot{x}_t^\epsilon = F(x_t^\epsilon) - \frac{g(x_t^\epsilon)}{\tau_\kappa \epsilon^\theta} \int_0^t A e^{-\frac{A}{\tau_\kappa \epsilon^\theta}(t-s)} h(x_s^\epsilon) \dot{x}_s^\epsilon ds + \sigma(x_t^\epsilon) \eta_t^\epsilon, \quad (25)$$

where η_t^ϵ is the Ornstein–Uhlenbeck process with the correlation time $\tau_\eta \epsilon^\nu$, i.e. it is a process satisfying the SDE:

$$d\eta_t^\epsilon = -\frac{A}{\tau_\eta \epsilon^\nu} \eta_t^\epsilon dt + \frac{A}{\tau_\eta \epsilon^\nu} dW_t^{(d)}. \quad (26)$$

(ii) the SIDE (20) becomes (with $m := m_0 \epsilon^\mu$, $\tau_{ii} := \tau_\kappa \epsilon^\theta$ in the formula for $(\kappa_2)_{ii}$ in (16), x_t replaced by x_t^ϵ, h_t, u_t replaced by h_t^ϵ, u_t^ϵ respectively and $\tau := \tau_h \epsilon^\nu I$ in (17), (18)):

$$m_0 \epsilon^\mu \ddot{\boldsymbol{x}}_t^\epsilon = \boldsymbol{F}(\boldsymbol{x}_t^\epsilon) - \frac{\boldsymbol{g}(\boldsymbol{x}_t^\epsilon)}{\tau_\kappa \epsilon^\theta} \int_0^t e^{-\boldsymbol{\Omega}^2 \frac{(t-s)}{2\tau_\kappa \epsilon^\theta}} \left[\cos\left(\frac{\boldsymbol{\Omega}_0}{\tau_\kappa \epsilon^\theta}(t-s) \right) \right.$$

$$\left. + \frac{\boldsymbol{\Omega}_1}{2} \sin\left(\frac{\boldsymbol{\Omega}_0}{\tau_\kappa \epsilon^\theta}(t-s) \right) \right] \boldsymbol{h}(\boldsymbol{x}_s^\epsilon) \dot{\boldsymbol{x}}_s^\epsilon ds + \boldsymbol{\sigma}(\boldsymbol{x}_t^\epsilon) \boldsymbol{h}_t^\epsilon, \tag{27}$$

where $\boldsymbol{h}_t^\epsilon$ is the harmonic noise process with the correlation time $\tau_h \epsilon^\nu$, i.e. it is a process satisfying the SDE system:

$$d\boldsymbol{h}_t^\epsilon = \frac{1}{\tau_h \epsilon^\nu} \boldsymbol{u}_t^\epsilon dt, \tag{28}$$

$$d\boldsymbol{u}_t^\epsilon = -\frac{\boldsymbol{\Omega}^2}{\tau_h \epsilon^\nu} \boldsymbol{u}_t^\epsilon dt - \frac{\boldsymbol{\Omega}^2}{\tau_h \epsilon^\nu} \boldsymbol{h}_t^\epsilon dt + \frac{\boldsymbol{\Omega}^2}{\tau_h \epsilon^\nu} d\boldsymbol{W}_t^{(d)}. \tag{29}$$

Both SIDEs have the initial conditions $\boldsymbol{x}_0^\epsilon = \boldsymbol{x}$, $\dot{\boldsymbol{x}}_0^\epsilon = \boldsymbol{v}$. The initial conditions $\boldsymbol{\eta}_0^\epsilon$ (respectively, $\boldsymbol{h}_0^\epsilon$ and $\boldsymbol{u}_0^\epsilon$) are distributed according to the stationary measure of the SDE that the process $\boldsymbol{\eta}_t^\epsilon$ (respectively, $\boldsymbol{h}_t^\epsilon$ and $\boldsymbol{u}_t^\epsilon$) satisfies.

In this paper we set $\mu = \theta = \nu$, which is the case when all the characteristic time scales are of comparable magnitude in the limit as $\epsilon \to 0$. Our main goal is to derive a limiting equation for the (slow) $\boldsymbol{x}^\epsilon$-component of the process solving Eqs. (23), (24), including the special cases (25), (26) and (27)–(29), in the limit as $\epsilon \to 0$, in a strong pathwise sense.

We explain the motivation behind the above rescalings. The rescaling of the memory kernels $\boldsymbol{\kappa}(t - s)$, $\boldsymbol{\kappa}_1(t - s)$, $\boldsymbol{\kappa}_2(t - s)$ is such that in the limit $\epsilon \to 0$ the rescaled memory kernels converge to $\boldsymbol{K}_1 \delta(t)$ formally, where $\delta(t)$ is the Dirac-delta function and $\boldsymbol{K}_1$ is the effective damping constant defined in (9). On the other hand, the noise processes $\boldsymbol{\xi}_t^\epsilon = \boldsymbol{C}_2 \boldsymbol{\beta}_t^\epsilon$, $\boldsymbol{\eta}_t^\epsilon$ and $\boldsymbol{h}_t^\epsilon$ converge to white noise processes in the limit $\epsilon \to 0$.

3 Smoluchowski–Kramers Limit of SDE's Revisited

Let $(\boldsymbol{x}_t^m, \boldsymbol{v}_t^m) \in \mathbb{R}^n \times \mathbb{R}^n$, where $t \in [0, T]$, be a family of solutions (parametrized by a positive constant m) to the following SDEs:

$$d\boldsymbol{x}_t^m = \boldsymbol{v}_t^m dt, \tag{30}$$

$$m d\boldsymbol{v}_t^m = \boldsymbol{F}(\boldsymbol{x}_t^m) dt - \boldsymbol{\gamma}(\boldsymbol{x}_t^m) \boldsymbol{v}_t^m dt + \boldsymbol{\sigma}(\boldsymbol{x}_t^m) d\boldsymbol{W}_t^{(k)}. \tag{31}$$

In the SDEs above, $m > 0$ is the mass of the particle, $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{\gamma} : \mathbb{R}^n \to \mathbb{R}^{n \times n}$, $\boldsymbol{\sigma} : \mathbb{R}^n \to \mathbb{R}^{n \times k}$, and $\boldsymbol{W}^{(k)}$ is a k-dimensional Wiener process on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ satisfying the usual conditions [33]. The initial conditions are given by $\boldsymbol{x}_0^m = \boldsymbol{x}$, $\boldsymbol{v}_0^m = \boldsymbol{v}^m$. The above SDE system models diffusive phenomena in cases where the damping coefficient $\boldsymbol{\gamma}$ and diffusion coefficient $\boldsymbol{\sigma}$ are state-dependent.

The Smoluchowski–Kramers limit (or the small mass limit) of the system (30), (31) has been studied in [8,34–37]. The main result in [37] says that, under certain assumptions, the $\boldsymbol{x}^m$-component of the solution to (30), (31) converges (in a strong pathwise sense), as $m \to 0$, to the solution of a homogenized SDE that contains in particular the so-called noise-induced drift, not present in the pre-limit SDEs (see Theorem 1 for a precise statement). The presence of such noise-induced drift in the homogenized equation is a consequence of the state-dependence of the damping coefficient (and therefore also the diffusion coefficient if the system satisfies a fluctuation–dissipation relation). For an overview of the noise-induced drift phenomena, we refer to the review article [38].

In all the works mentioned previously, the spectral assumption made on the matrix $\boldsymbol{\gamma}$ was that the symmetrized damping matrix $\frac{1}{2}(\boldsymbol{\gamma} + \boldsymbol{\gamma}^*)$ is uniformly positive definite (i.e. its smallest eigenvalue is positive uniformly in $\boldsymbol{x}$). The same results can be obtained under a weaker assumption that matrix $\boldsymbol{\gamma}$ is *uniformly positive stable*, i.e. all real parts of the eigenvalues of $\boldsymbol{\gamma}$ are positive uniformly in $\boldsymbol{x}$ [39].

Notation Here and in the following, we use Einstein summation convention on repeated indices. The Euclidean norm of a vector $\boldsymbol{w}$ is denoted by $|\boldsymbol{w}|$ and the (induced operator) norm of a matrix A by $\|A\|$. For an $\mathbb{R}^{n_2 \times n_3}$-valued function $\boldsymbol{f}(\boldsymbol{y}) := ([f]_{jk}(\boldsymbol{y}))_{j=1,\dots,n_2;k=1,\dots,n_3}$, $\boldsymbol{y} := ([y]_1, \dots, [y]_{n_1}) \in \mathbb{R}^{n_1}$, we denote by $(\boldsymbol{f})_{\boldsymbol{y}}(\boldsymbol{y})$ the $n_1 n_2 \times n_3$ matrix:

$$(\boldsymbol{f})_{\boldsymbol{y}}(\boldsymbol{y}) = (\nabla_{\boldsymbol{y}}[f]_{jk}(\boldsymbol{y}))_{j=1,\dots,n_2;k=1,\dots,n_3}, \tag{32}$$

where $\nabla_{\boldsymbol{y}}[f]_{jk}(\boldsymbol{y})$ denotes the gradient vector $(\frac{\partial[f]_{jk}(\boldsymbol{y})}{\partial[y]_1}, \dots, \frac{\partial[f]_{jk}(\boldsymbol{y})}{\partial[y]_{n_1}}) \in \mathbb{R}^{n_1}$ for every j, k. The symbol $\mathbb{E}$ denotes expectation with respect to $\mathbb{P}$.

We make the following assumptions.

Assumption 1 *For every $\boldsymbol{x} \in \mathbb{R}^n$, the functions $\boldsymbol{F}(\boldsymbol{x})$ and $\boldsymbol{\sigma}(\boldsymbol{x})$ are continuous, bounded and Lipschitz in $\boldsymbol{x}$, whereas the functions $\boldsymbol{\gamma}(\boldsymbol{x})$ and $(\boldsymbol{\gamma})_{\boldsymbol{x}}(\boldsymbol{x})$ are continuously differentiable, bounded and Lipschitz in $\boldsymbol{x}$. Moreover, $(\boldsymbol{\gamma})_{\boldsymbol{xx}}(\boldsymbol{x})$ is bounded for every $\boldsymbol{x} \in \mathbb{R}^n$.*

Assumption 2 *The matrix $\boldsymbol{\gamma}$ is uniformly positive stable, i.e. all real parts of the eigenvalues of $\boldsymbol{\gamma}(\boldsymbol{x})$ are bounded below by some constant $2\kappa > 0$, uniformly in $\boldsymbol{x} \in \mathbb{R}^n$.*

Assumption 3 *The initial condition $\boldsymbol{x}_0^m = \boldsymbol{x}_0$ is a random variable independent of m and has finite moments of all orders, i.e. $\mathbb{E}|\boldsymbol{x}|^p < \infty$ for all $p > 0$. The initial condition $\boldsymbol{v}_0^m$ is a random variable that possibly depends on m and we assume that for every $p > 0$, $\mathbb{E}|m\boldsymbol{v}^m|^p = O(m^\alpha)$ as $m \to 0$, for some $\alpha \geq p/2$.*

Assumption 4 *The global solutions, defined on $[0, T]$, to the pre-limit SDEs (30), (31) and to the limiting SDE (33) a.s. exist and are unique for all $m > 0$ (i.e. there are no explosions).*

We now state the result.

Theorem 1 *Suppose that the SDE system (30), (31) satisfies Assumption 1–4. Let $(\boldsymbol{x}_t^m, \boldsymbol{v}_t^m) \in \mathbb{R}^n \times \mathbb{R}^n$ be its solution, with the initial condition $(\boldsymbol{x}, \boldsymbol{v}^m)$. Let $\boldsymbol{X}_t \in \mathbb{R}^n$ be the solution to the following Itô SDE with initial position $\boldsymbol{X}_0 = \boldsymbol{x}$:*

$$d\boldsymbol{X}_t = [\boldsymbol{\gamma}^{-1}(\boldsymbol{X}_t)\boldsymbol{F}(\boldsymbol{X}_t) + \boldsymbol{S}(\boldsymbol{X}_t)]dt + \boldsymbol{\gamma}^{-1}(\boldsymbol{X}_t)\boldsymbol{\sigma}(\boldsymbol{X}_t)d\boldsymbol{W}_t^{(k)}, \tag{33}$$

where $\boldsymbol{S}(\boldsymbol{X}_t)$ is the noise-induced drift whose ith component is given by

$$S_i(\boldsymbol{X}) = \frac{\partial}{\partial X_l}[(\gamma^{-1})_{ij}(\boldsymbol{X})]J_{jl}(\boldsymbol{X}), \quad i, j, l = 1, \dots, n, \tag{34}$$

and $\boldsymbol{J}$ is the unique matrix solving the Lyapunov equation

$$\boldsymbol{J}\boldsymbol{\gamma}^* + \boldsymbol{\gamma}\boldsymbol{J} = \boldsymbol{\sigma}\boldsymbol{\sigma}^*. \tag{35}$$

Then the process $\boldsymbol{x}_t^m$ converges, as $m \to 0$, to the solution $\boldsymbol{X}_t$, of the Itô SDE (33), in the following sense: for all finite $T > 0$,

$$\sup_{t \in [0,T]} |\boldsymbol{x}_t^m - \boldsymbol{X}_t| \to 0 \tag{36}$$

in probability, in the limit as $m \to 0$.

We end this section with a few remarks concerning the statements in Theorem 1.

Remark 5 (i) Because of the relaxed spectral assumption on $\boldsymbol{\gamma}$, a new idea has to be used to prove decay estimates for solutions of the velocity equation. Once this is done, Theorem 1 can be proven using the technique of [37] (note that Assumption 1 is essentially the same as the assumptions in Appendix A of [37], specialized to the present case). In Appendix B we give a sketch of the proof of Theorem 1, pointing out the necessary modifications. The reader is referred to [37] for more details.

(ii) Our assumption on the initial variable $\boldsymbol{v}_0^m$ implies that the initial average kinetic energy, $K(\boldsymbol{v}^m) := \mathbb{E}\frac{1}{2}m|\boldsymbol{v}^m|^2$, does not blow up (but can possibly vanish) as $m \to 0$. This is analogous to the Assumption 2.4 in [37] and it is more general than the one in [8].

(iii) With slightly more work and additional assumptions, one could prove the statement in Assumption 4 from Assumptions 1–3. See Appendix C in [37] for a result along these lines. However, such existence and uniqueness result is not the focus of this paper and, therefore, we choose to take the existence and uniqueness result for granted in Assumption 4.

(iv) We make no claim that Assumptions 2–4 are as weak or as general as possible. In particular, the boundedness assumption on the coefficients of the SDEs could be relaxed (for instance, using the techniques in [35]) and the initial condition $\boldsymbol{x}$ could have some dependence on m (see, for instance [37]) at the cost of more technicalities.

The main focus of our revisit here is to point out that the result in [8] still holds with a relaxed spectral assumption on the matrix $\boldsymbol{\gamma}$ and with the initial condition $\boldsymbol{v}_0^m$ possibly dependent on m—this will be important for applications in later sections (see also Remark 11).

4 Homogenization for Generalized Langevin Dynamics

In this section, we study homogenization for the system of equations (23), (24) (with $\mu = \theta = \nu$) by taking the limit as $\epsilon \to 0$, under appropriate assumptions.

Without loss of generality, we set $\mu = \theta = \nu = 1$. We cast (23), (24) as the system of SDEs for the Markov process $(\boldsymbol{x}_t^\epsilon, \boldsymbol{v}_t^\epsilon, \boldsymbol{z}_t^\epsilon, \boldsymbol{y}_t^\epsilon, \boldsymbol{\zeta}_t^\epsilon, \boldsymbol{\beta}_t^\epsilon)$ on the state space $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_2}$:

$$d\boldsymbol{x}_t^\epsilon = \boldsymbol{v}_t^\epsilon dt, \tag{37}$$

$$m_0\epsilon d\boldsymbol{v}_t^\epsilon = -\boldsymbol{g}(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_1\boldsymbol{y}_t^\epsilon dt + \boldsymbol{\sigma}(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_2\boldsymbol{\beta}_t^\epsilon dt + \boldsymbol{F}(\boldsymbol{x}_t^\epsilon)dt, \tag{38}$$

$$d\boldsymbol{z}_t^\epsilon = \boldsymbol{y}_t^\epsilon dt, \tag{39}$$

$$\tau_\kappa \epsilon d\boldsymbol{y}_t^\epsilon = -\boldsymbol{\Gamma}_1\boldsymbol{y}_t^\epsilon dt + \boldsymbol{M}_1\boldsymbol{C}_1^*\boldsymbol{h}(\boldsymbol{x}_t^\epsilon)\boldsymbol{v}_t^\epsilon dt, \tag{40}$$

$$d\boldsymbol{\zeta}_t^\epsilon = \boldsymbol{\beta}_t^\epsilon dt, \tag{41}$$

$$\tau_\xi \epsilon d\boldsymbol{\beta}_t^\epsilon = -\boldsymbol{\Gamma}_2\boldsymbol{\beta}_t^\epsilon dt + \boldsymbol{\Sigma}_2 d\boldsymbol{W}_t^{(q_2)}, \tag{42}$$

where we have defined the auxiliary process

$$\boldsymbol{y}_t^\epsilon := \frac{1}{\tau_\kappa \epsilon} \int_0^t e^{-\frac{\boldsymbol{\Gamma}_1}{\tau_\kappa \epsilon}(t-s)} \boldsymbol{M}_1\boldsymbol{C}_1^*\boldsymbol{h}(\boldsymbol{x}_s^\epsilon)\boldsymbol{v}_s^\epsilon ds. \tag{43}$$

Here, the initial conditions $\boldsymbol{x}_0^\epsilon = \boldsymbol{x}$, $\boldsymbol{v}_0^\epsilon = \boldsymbol{v}$, $\boldsymbol{z}_0^\epsilon = \boldsymbol{z}$ and $\boldsymbol{\zeta}_0^\epsilon = \boldsymbol{\zeta}$ are random variables. Note that $\boldsymbol{y}_0^\epsilon = \boldsymbol{0}$, and $\boldsymbol{\beta}_0^\epsilon$ is a zero mean Gaussian random variable with covariance $\boldsymbol{M}_2/\tau_\xi \epsilon$.

Let $W_t^{(q_2)}$ be an $\mathbb{R}^{q_2}$-valued Wiener process on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ satisfying the usual conditions [33] and $\mathbb{E}$ denotes expectation with respect to $\mathbb{P}$.

We use the notation introduced in Sect. 3 and make the following assumptions.

Assumption 5 *For every $x \in \mathbb{R}^d$, the vector-valued function $F(x)$ is continuous, bounded and Lipschitz in x, whereas the matrix-valued functions $g(x)$, $h(x)$, $\sigma(x)$, $(g)_x(x)$, $(h)_x(x)$ and $(\sigma)_x(x)$ are continuously differentiable, bounded and Lipschitz in x. Moreover, $(g)_{xx}(x)$, $(h)_{xx}(x)$ and $(\sigma)_{xx}(x)$ are bounded for every $x \in \mathbb{R}^d$.*

Assumption 6 *The initial conditions x, v, z, ζ are random variables independent of ϵ. We assume that they have finite moments of all orders, i.e. $\mathbb{E}|x|^p$, $\mathbb{E}|v|^p$, $\mathbb{E}|z|^p$, $\mathbb{E}|\zeta|^p < \infty$ for all $p > 0$.*

Assumption 7 *There are no explosions, i.e. almost surely, for any value of the parameter ϵ there exists a unique solution on the compact time interval $[0, T]$ to the pre-limit Eqs. (23), (24), and also to the limiting Eq. (45).*

The following convergence theorem is the main result of this paper. It provides a homogenized SDE for the particle's position in the limit as the inertial time scale, the memory time scale and the noise correlation time scale go to zero at the same rate in the case when the pre-limit dynamics are described by the family of Eqs. (23), (24) (with $\mu = \theta = \nu = 1$), or equivalently by the SDEs (37)–(42). In the following, $(D)_{ij}$ denotes the (i, j)-entry of the matrix D.

Theorem 2 *Let $x_t^\epsilon \in \mathbb{R}^d$ be the solution to the SDEs (37)–(42). Suppose that Assumptions 5–7 are satisfied and the effective damping and diffusion (constant) matrices, K_1, K_2, defined in (9) and (10) respectively, are invertible. Moreover, we assume that for every $x \in \mathbb{R}^d$,*

$$B_\lambda(x) := I + g(x)\tilde{\kappa}(\lambda\tau_\kappa)h(x)/\lambda m_0 \tag{44}$$

is invertible for all λ in the right half plane $\{\lambda \in \mathbb{C} : Re(\lambda) > 0\}$, where $\tilde{\kappa}(z) := C_1(zI + \Gamma_1)^{-1}M_1C_1^*$, for $z \in \mathbb{C}$, is the Laplace transform of the memory function.

Denote $\theta(X) := g(X)K_1h(X) \in \mathbb{R}^{d\times d}$ for $X \in \mathbb{R}^d$. Then as $\epsilon \to 0$, the process x_t^ϵ converges to the solution, X_t, of the following Itô SDE:

$$dX_t = S(X_t)dt + \theta^{-1}(X_t)F(X_t)dt + \theta^{-1}(X_t)\sigma(X_t)C_2\Gamma_2^{-1}\Sigma_2dW_t^{(q_2)}, \tag{45}$$

with $S(X_t) = S^{(1)}(X_t) + S^{(2)}(X_t) + S^{(3)}(X_t)$, where the $S^{(k)}$ are the noise-induced drifts whose ith components are given by

$$S_i^{(1)} = m_0 \frac{\partial}{\partial X_l}\left[(\theta^{-1})_{ij}(X)\right](J_{11})_{jl}(X), \quad i, j, l = 1, \ldots, d, \tag{46}$$

$$S_i^{(2)} = -\tau_\kappa \frac{\partial}{\partial X_l}\left[(\theta^{-1}g)_{ij}(X)\right](C_1\Gamma_1^{-1}J_{21})_{jl}(X), \quad i, l = 1, \ldots, d; \ j = 1, \ldots, q, \tag{47}$$

$$S_i^{(3)} = \tau_\xi \frac{\partial}{\partial X_l}\left[(\theta^{-1}\sigma)_{ij}(X)\right](C_2\Gamma_2^{-1}J_{31})_{jl}(X), \quad i, l = 1, \ldots, d; \ j = 1, \ldots, r. \tag{48}$$

Here $J_{11} = J_{11}^* \in \mathbb{R}^{d\times d}$, $J_{21} = J_{12}^* \in \mathbb{R}^{d_1\times d}$ and $J_{31} = J_{13}^* \in \mathbb{R}^{d_2\times d}$ satisfy the following system of five matrix equations:

$$gC_1J_{12}^* + J_{12}C_1^*g^* = \sigma C_2J_{13}^* + J_{13}C_2^*\sigma^*, \tag{49}$$

$$m_0 J_{11} h^* C_1 M_1 + \tau_\kappa \sigma C_2 J_{23}^* = \tau_\kappa g C_1 J_{22} + m_0 J_{12} \Gamma_1^*, \tag{50}$$

$$\tau_\xi g C_1 J_{23} + m_0 J_{13} \Gamma_2^* = \sigma C_2 M_2, \tag{51}$$

$$M_1 C_1^* h J_{12} + J_{12}^* h^* C_1 M_1 = \Gamma_1 J_{22} + J_{22} \Gamma_1^*, \tag{52}$$

$$\tau_\xi M_1 C_1^* h J_{13} = \tau_\xi \Gamma_1 J_{23} + \tau_\kappa J_{23} \Gamma_2^*.. \tag{53}$$

The convergence is obtained in the following sense: for all finite $T > 0$,

$$\sup_{t \in [0,T]} |x_t^\epsilon - X_t| \to 0 \tag{54}$$

in probability, in the limit as $\epsilon \to 0$.

Remark 6 Invertibility of the matrices $B_\lambda(x)$ is a technical condition which will be used in the proof of the theorem. We are going to verify it in the special cases and applications discussed later (see Corollaries 3 and 5). In particular, it follows from the stronger spectral condition, namely that $g(x)\tilde{\kappa}(\mu)h(x)$ has no spectrum in the right half plane for any μ with $Re(\mu) > 0$. See also Remark 8.

Remark 7 It is straightforward to extend Theorems 1 and 2 to the case when the coefficients in the pre-limit equations depend explicitly on time, i.e. when $\gamma = \gamma(x_t^m, t)$, $\sigma = \sigma(x_t^m, t)$ and $F = F(x_t^m, t)$ in (30), (31), as well as when $g = g(x_t^\epsilon, t)$, $h = h(x_t^\epsilon, t)$, $\sigma = \sigma(x_t^\epsilon, t)$ and $F = F(x_t^\epsilon, t)$ in (37)–(42). In this case, we expect similar results to hold under additional assumptions, analogous to those studied in [37].

Proof We denote $\hat{x}_t^\epsilon := (x_t^\epsilon, z_t^\epsilon, \zeta_t^\epsilon) \in \mathbb{R}^{d+d_1+d_2}$ and $\hat{v}_t^\epsilon := (v_t^\epsilon, y_t^\epsilon, \eta_t^\epsilon) \in \mathbb{R}^{d+d_1+d_2}$ and rewrite the above SDE system in the form (30), (31):

$$d\hat{x}_t^\epsilon = \hat{v}_t^\epsilon dt, \tag{55}$$

$$\epsilon d\hat{v}_t^\epsilon = -\hat{\gamma}(x_t^\epsilon)\hat{v}_t^\epsilon dt + \hat{F}(x_t^\epsilon)dt + \hat{\sigma} dW_t^{(q_2)}, \tag{56}$$

with

$$\hat{\gamma}(x_t^\epsilon) = \begin{bmatrix} \mathbf{0} & \frac{g(x_t^\epsilon)C_1}{m_0} & -\frac{\sigma(x_t^\epsilon)C_2}{m_0} \\ -\frac{M_1 C_1^* h(x_t^\epsilon)}{\tau_\kappa} & \frac{\Gamma_1}{\tau_\kappa} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{\Gamma_2}{\tau_\xi} \end{bmatrix}, \quad \hat{F}(x_t^\epsilon) = \begin{bmatrix} \frac{F(x_t^\epsilon)}{m_0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \hat{\sigma} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \frac{\Sigma_2}{\tau_\xi} \end{bmatrix}, \tag{57}$$

where $\hat{\gamma} \in \mathbb{R}^{(d+d_1+d_2) \times (d+d_1+d_2)}$ is a 3 by 3 block matrix with each block a matrix of appropriate dimension, $\hat{F} \in \mathbb{R}^{d+d_1+d_2}$, $\hat{\sigma} \in \mathbb{R}^{(d+d_1+d_2) \times q_2}$ and the $\mathbf{0}$ appearing in $\hat{\gamma}$, $\hat{F}$ and $\hat{\sigma}$ is a zero vector or matrix of appropriate dimension.

We now want to apply Theorem 1 (with $m := \epsilon$, $n := d + d_1 + d_2$, γ replaced by $\hat{\gamma}$, F replaced by $\hat{F}$, σ replaced by $\hat{\sigma}$, etc.) to (55), (56).

It is straightforward to see that Assumption 5 implies Assumptions 1 and 7 implies Assumption 4.

To verify Assumption 3, note again that $y_0^\epsilon = \mathbf{0}$ and so by Assumption 6, we only need to show that for every $p > 0$, $\mathbb{E}|\epsilon \beta_0^\epsilon|^p = O(\epsilon^\alpha)$ as $\epsilon \to 0$, for some constant $\alpha \geq p/2$. To show this, we use the fact that for a mean zero Gaussian random variable, $X \in \mathbb{R}$, with variance σ^2,

$$\mathbb{E}|X|^p = \sigma^p \frac{2^{p/2} \Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}, \tag{58}$$

for every $p > 0$, where Γ denotes the gamma function [40]. Applying this to $\boldsymbol{\beta}_0^\epsilon$, we obtain, for every $p > 0$, $\mathbb{E}|\boldsymbol{\beta}_0^\epsilon|^p = O(1/\epsilon^{p/2})$ as $\epsilon \to 0$ and so $\mathbb{E}|\epsilon\boldsymbol{\beta}_0^\epsilon|^p = O(\epsilon^{p/2})$ as $\epsilon \to 0$. Therefore, Assumption 3 is verified.

It remains to verify Assumption 2, i.e. that $\hat{\boldsymbol{\gamma}}$ is positive stable. Note that $\boldsymbol{\Gamma}_2$ is positive stable by assumption and the triangular-block structure of $\hat{\boldsymbol{\gamma}}$ implies that one only needs to verify that the upper left 2 by 2 block matrix of $\hat{\boldsymbol{\gamma}}$:

$$L(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{0} & g(\boldsymbol{x})C_1/m_0 \\ -M_1 C_1^* h(\boldsymbol{x})/\tau_\kappa & \boldsymbol{\Gamma}_1/\tau_\kappa \end{bmatrix} \tag{59}$$

is positive stable, where $\boldsymbol{x} \in \mathbb{R}^d$.

We thus need to show that the resolvent set of $-L(\boldsymbol{x})$, $\rho(-L(\boldsymbol{x})) := \{\lambda \in \mathbb{C} : (\lambda I + L(\boldsymbol{x}))^{-1}$ exists$\}$, contains the right half plane $\{\lambda \in \mathbb{C} : Re(\lambda) > 0\}$ for every $\boldsymbol{x} \in \mathbb{R}^d$.

Let $\lambda \in \mathbb{C}$ such that $Re(\lambda) > 0$. We will use the following formula for blockwise inversion of a block matrix: provided that S and $P - QS^{-1}R$ are nonsingular, we have

$$\begin{bmatrix} P & Q \\ R & S \end{bmatrix}^{-1} = \begin{bmatrix} (P - QS^{-1}R)^{-1} & -(P - QS^{-1}R)^{-1}QS^{-1} \\ -S^{-1}R(P - QS^{-1}R)^{-1} & S^{-1} + S^{-1}R(P - QS^{-1}R)^{-1}QS^{-1} \end{bmatrix}, \tag{60}$$

where P, Q, R, S are matrix sub-blocks of arbitrary dimension.

Since the matrices $A_\lambda := \boldsymbol{\Gamma}_1/\tau_\kappa + \lambda I$ and $B_\lambda(\boldsymbol{x}) := I + g(\boldsymbol{x})\tilde{\kappa}(\lambda\tau_\kappa)h(\boldsymbol{x})/\lambda m_0$ are invertible for all λ in the right half plane by assumption, $\lambda I + L(\boldsymbol{x})$ is indeed invertible for every $\boldsymbol{x}$ and in fact using the above formula for the inverse of a block matrix, we have:

$$(\lambda I + L(\boldsymbol{x}))^{-1}$$
$$= \begin{bmatrix} B_\lambda^{-1}(\boldsymbol{x})/\lambda & -B_\lambda^{-1}(\boldsymbol{x})g(\boldsymbol{x})C_1 A_\lambda^{-1}/\lambda m_0 \\ A_\lambda^{-1}M_1 C_1^* h(\boldsymbol{x})B_\lambda^{-1}(\boldsymbol{x})/\lambda\tau_\kappa & A_\lambda^{-1}(I - M_1 C_1^* h(\boldsymbol{x})B_\lambda^{-1}(\boldsymbol{x})g(\boldsymbol{x})C_1 A_\lambda^{-1}/\lambda m_0\tau_\kappa) \end{bmatrix}. \tag{61}$$

Therefore, $\hat{\boldsymbol{\gamma}}$ is invertible and one can compute:

$$\hat{\boldsymbol{\gamma}}^{-1} = \begin{bmatrix} m_0\theta^{-1} & -\tau_\kappa\theta^{-1}gC_1\boldsymbol{\Gamma}_1^{-1} & \tau_\xi\theta^{-1}\sigma C_2\boldsymbol{\Gamma}_2^{-1} \\ m_0\boldsymbol{\Gamma}_1^{-1}M_1 C_1^* h\theta^{-1} & \tau_\kappa\boldsymbol{\Gamma}_1^{-1}(I - M_1 C_1^* h\theta^{-1}gC_1\boldsymbol{\Gamma}_1^{-1}) & \tau_\xi\boldsymbol{\Gamma}_1^{-1}M_1 C_1^* h\theta^{-1}\sigma C_2\boldsymbol{\Gamma}_2^{-1} \\ \boldsymbol{0} & \boldsymbol{0} & \tau_\xi\boldsymbol{\Gamma}_2^{-1} \end{bmatrix}, \tag{62}$$

where $\theta = gK_1 h$. The result follows by applying Theorem 1 to the SDE systems (55), (56). In particular, a rewriting of the resulting Lyapunov equation (35) gives the system of matrix equations (49)–(53). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 8 In the above proof, the condition of invertibility of $B_\lambda(\boldsymbol{x})$ is only used to guarantee the positive stability of the matrix $\hat{\boldsymbol{\gamma}}$. Therefore, the conclusion of the theorem holds also when the latter can be established in another way. This can indeed be done in a number of concrete examples (see, for instance, the matrix $\boldsymbol{\gamma}$ in Eq. (83), or the line (90) and the sentence below the line).

Remark 9 Our SIDEs belong to a special class of non-Markovian equations, the so-called *quasi-Markovian Langevin equations* [41]. For these equations one can introduce a finite number of auxiliary variables in such a way that the evolution of particle's position and velocity, together with these auxiliary variables, is described by a usual SDE system and is thus Markovian. We remark that such "Markovianization" procedure works here because the colored noise can be generated by a linear system of SDEs and the memory kernel satisfies

a linear system of ordinary differential equations—both with constant coefficients. If, on the other hand, the memory kernel decays as a power, then there is no finite dimensional extension of the space which would make the solution process Markovian [16].

The following corollary uses a linear change of variables in a given SIDE, to arrive at an alternative form of the corresponding homogenized SDE of the form (45).

Corollary 1 *For $i = 1, 2$, let $\boldsymbol{T}_i$ be arbitrary $d_i \times d_i$ constant invertible matrix, where d_1, d_2 are positive integers. For $t \geq 0$, denote $\boldsymbol{\Gamma}'_i = \boldsymbol{T}_i \boldsymbol{\Gamma}_i \boldsymbol{T}_i^{-1}$, $\boldsymbol{M}'_i = \boldsymbol{T}_i \boldsymbol{M}_i \boldsymbol{T}_i^*$, $\boldsymbol{C}'_i = \boldsymbol{C}_i \boldsymbol{T}_i^{-1}$, $(\boldsymbol{\beta}^\epsilon_t)' = \boldsymbol{T}_2 \boldsymbol{\beta}^\epsilon_t$, $\boldsymbol{\Sigma}'_i = \boldsymbol{T}_i \boldsymbol{\Sigma}_i$ and consider the equations:*

$$m_0 \epsilon^\mu \ddot{\boldsymbol{x}}^\epsilon_t = \boldsymbol{F}(\boldsymbol{x}^\epsilon_t) - \frac{\boldsymbol{g}(\boldsymbol{x}^\epsilon_t)}{\tau_\kappa \epsilon^\theta} \int_0^t \boldsymbol{C}'_1 e^{-\frac{\boldsymbol{\Gamma}'_1}{\tau_\kappa \epsilon^\theta}(t-s)} \boldsymbol{M}'_1 (\boldsymbol{C}'_1)^* \boldsymbol{h}(\boldsymbol{x}^\epsilon_s) \dot{\boldsymbol{x}}^\epsilon_s ds + \sigma(\boldsymbol{x}^\epsilon_t) \boldsymbol{C}'_2 (\boldsymbol{\beta}^\epsilon_t)',$$
(63)

$$\tau_\xi \epsilon^\nu d(\boldsymbol{\beta}^\epsilon_t)' = -\boldsymbol{\Gamma}'_2 (\boldsymbol{\beta}^\epsilon_t)' dt + \boldsymbol{\Sigma}'_2 d\boldsymbol{W}'_t,$$
(64)

where $\boldsymbol{W}'_t$ is a q_2-dimensional Wiener process and the initial condition, $(\boldsymbol{\beta}^\epsilon_0)'$, is normally distributed with zero mean and covariance of $\boldsymbol{M}'_2/(\tau_\xi \epsilon^\nu)$.

Suppose that Assumptions 5–7 are satisfied and the effective damping and diffusion constants, $\boldsymbol{K}'_i = \boldsymbol{C}'_i (\boldsymbol{\Gamma}'_i)^{-1} \boldsymbol{M}'_i (\boldsymbol{C}'_i)^* = \boldsymbol{K}_i$ ($i = 1, 2$), are invertible. Moreover, we assume that $\boldsymbol{I} + \boldsymbol{g}(\boldsymbol{x}) \tilde{\boldsymbol{\kappa}}'(\lambda \tau_\kappa) \boldsymbol{h}(\boldsymbol{x})/\lambda m_0$ is invertible for all λ in the right half plane $\{\lambda \in \mathbb{C} : Re(\lambda) > 0\}$ and $\boldsymbol{x} \in \mathbb{R}^d$, where $\tilde{\boldsymbol{\kappa}}'(z) := \boldsymbol{C}'_1 (z\boldsymbol{I} + \boldsymbol{\Gamma}'_1)^{-1} \boldsymbol{M}'_1 (\boldsymbol{C}'_1)^* = \tilde{\boldsymbol{\kappa}}(z)$ for $z \in \mathbb{C}$.

Let $\mu = \theta = \nu$ in (63), (64). Then as $\epsilon \to 0$, the process $\boldsymbol{x}^\epsilon_t$ converges, in the similar sense as in Theorem 2, to $\boldsymbol{X}_t$, where $\boldsymbol{X}_t$ is the solution of the SDE (45) with the $\boldsymbol{C}_i$, $\boldsymbol{\Gamma}_i$, $\boldsymbol{M}_i$, $\boldsymbol{\Sigma}_i$ replaced by $\boldsymbol{C}'_i$, $\boldsymbol{\Gamma}'_i$, $\boldsymbol{M}'_i$, $\boldsymbol{\Sigma}'_i$ respectively, and the driving Wiener process $\boldsymbol{W}^{(q_2)}_t$ replaced by $\boldsymbol{W}'_t$.

Corollary 1 is an easy consequence of Theorem 2.

Next, we discuss a particular, but very important, case when a *fluctuation–dissipation relation* holds. This is, for instance, the case when the pre-limit dynamics are (heuristically) derived from Hamiltonian dynamics (see Appendix A). We will further explore similar cases of fluctuation–dissipation relations for the two sub-classes.

Corollary 2 *Let $\boldsymbol{x}^\epsilon_t \in \mathbb{R}^d$ be the solution to the SDEs (37)–(42). Suppose that the assumptions of Theorem 2 holds. Moreover, we assume that:*

$$\tau_\kappa = \tau_\xi = \tau, \quad \sigma = \boldsymbol{g}, \quad \boldsymbol{h} = \boldsymbol{g}^*,$$
(65)

where τ is a positive constant, and

$$\boldsymbol{C}_1 = \boldsymbol{C}_2 := \boldsymbol{C}, \quad \boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_2 := \boldsymbol{\Gamma}, \quad \boldsymbol{M}_1 = \boldsymbol{M}_2 := \boldsymbol{M}, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 := \boldsymbol{\Sigma},$$
(66)

(so that $q = r$ and $d_1 = d_2$). Denote $\boldsymbol{K} := \boldsymbol{C} \boldsymbol{\Gamma}^{-1} \boldsymbol{M} \boldsymbol{C}^*$. Then as $\epsilon \to 0$, the process $\boldsymbol{x}^\epsilon_t$ converges to the solution, $\boldsymbol{X}_t$, of the following Itô SDE:

$$d\boldsymbol{X}_t = \boldsymbol{S}(\boldsymbol{X}_t)dt + [\boldsymbol{g}(\boldsymbol{X}_t)\boldsymbol{K}\boldsymbol{g}^*(\boldsymbol{X}_t)]^{-1}\boldsymbol{F}(\boldsymbol{X}_t)dt + [\boldsymbol{g}(\boldsymbol{X}_t)\boldsymbol{K}\boldsymbol{g}^*(\boldsymbol{X}_t)]^{-1}\boldsymbol{g}(\boldsymbol{X}_t)\boldsymbol{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}d\boldsymbol{W}^{(q_2)}_t,$$
(67)

where $\boldsymbol{S}(\boldsymbol{X}_t)$ is the noise-induced drift whose ith component is given by

$$S_i(\boldsymbol{X}) = m_0 \frac{\partial}{\partial X_l} \left[((\boldsymbol{g}\boldsymbol{K}\boldsymbol{g}^*)^{-1})_{ij}(\boldsymbol{X}) \right] (\boldsymbol{J}_{11})_{jl}(\boldsymbol{X}), \quad i, j, l = 1, \ldots, d,$$
(68)

where $\boldsymbol{J}_{11}$ solves the following system of three matrix equations:

$$m_0 \boldsymbol{J}_{11} \boldsymbol{g} \boldsymbol{C} \boldsymbol{M} + \tau \boldsymbol{g} \boldsymbol{C} (\boldsymbol{J}_{23} + \boldsymbol{J}^*_{23}) = \tau \boldsymbol{g} \boldsymbol{C} \boldsymbol{J}_{22} + \boldsymbol{g} \boldsymbol{C} \boldsymbol{M},$$
(69)

$$
MC^* g^* g C M (\Gamma^{-1})^* = \tau MC^* g^* g C J_{23} (\Gamma^{-1})^*
$$
$$
+ m_0 (\Gamma J_{23} + J_{23} \Gamma^*), \tag{70}
$$
$$
MC^* g^* g C M (\Gamma^{-1})^* + \Gamma^{-1} MC^* g^* g C M = \tau (MC^* g^* \Gamma^{-1} J_{23}^* C^* g^*
$$
$$
+ \Gamma^{-1} J_{23}^* C^* g^* g C M)
$$
$$
+ m_0 (\Gamma J_{22} + J_{22} \Gamma^*). \tag{71}
$$

The convergence is obtained in the same sense as in Theorem 2.

Equations (65) and (66) are a form of fluctuation–dissipation relation familiar from non-equilibrium statistical mechanics [13]. As stationary measures of systems satisfying fluctuation–dissipation relations are in equilibrium with respect to the underlying dynamics, this result is relevant for describing equilibrium properties of such systems in the small mass limit.

Remark 10 Therefore, if the fluctuation–dissipation relation holds, the noise-induced drift in the limiting SDE reduces to *a single term* (later we will see how this term simplifies in some special cases). This result may have interesting implications for nanoscale systems in equilibrium. We remark that the conditions for the fluctuation–dissipation relation in Corollary 2 can be written in other equivalent forms, up to the transformations in (8) and multiplication by a constant.

Proof The above corollary follows from applying Theorem 2. Indeed, by assumptions of the corollary, (49) simplifies to:

$$
gC(J_{12} - J_{13})^* + (J_{12} - J_{13})(gC)^* = 0. \tag{72}
$$

This implies that $J_{12} = J_{13}$ and, therefore, $S^{(2)}$ and $S^{(3)}$ cancel. Rewriting the resulting system of matrix equations in (49)–(53) give (69)–(71).   □

5 Homogenization for Models of the Two Sub-classes

We now return to the two sub-classes of SIDEs (11) introduced in Sect. 2. In this section, we study the effective dynamics described by SIDEs (25) and (27) in the limit as $\epsilon \to 0$. By specializing to these two sub-classes, the convergence result of Theorem 2, in particular the expressions in (45)–(53), can be made more explicit under certain assumptions on the matrix-valued coefficients and therefore the limiting equation obtained may be useful for modeling purposes.

5.1 SIDEs Driven by a Markovian Colored Noise

The following convergence result provides a homogenized SDE for the particle's position in the limit as the inertial time scale, the memory time scale and the noise correlation time scale vanish at the same rate in the case when the pre-limit dynamics are driven by an Ornstein–Uhlenbeck noise.

Corollary 3 *Let $d = d_1 = d_2 = q_1 = q_2 = q = r$. We set, in the SDEs (37)-(42): $\beta_t^\epsilon = \eta_t^\epsilon$, $\tau_\xi = \tau_\eta$, $W_t^{(q_2)} = W_t^{(d)} := W_t$ and*

$$
(\Gamma_1, M_1, C_1) = (A, A, I), \quad (\Gamma_2, M_2, C_2) = (A, A/2, I), \tag{73}
$$

to obtain SDEs equivalent to Eqs. (25) and (26) with $\mu = \theta = \nu = 1$. Let $\boldsymbol{x}_t^\epsilon \in \mathbb{R}^d$ be the solution to these equations, with the matrices $\boldsymbol{g}(\boldsymbol{x})$ and $\boldsymbol{h}(\boldsymbol{x})$ positive definite for every $\boldsymbol{x} \in \mathbb{R}^d$. Suppose that Assumptions 5–7 are satisfied and, moreover, that $\boldsymbol{g}(\boldsymbol{x})$, $\boldsymbol{h}(\boldsymbol{x})$ and the diagonal matrix $\boldsymbol{A}$ are commuting.

Then as $\epsilon \to 0$, the process $\boldsymbol{x}_t^\epsilon$ converges to the solution, $\boldsymbol{X}_t$, of the following Itô SDE:

$$dX_t = S(X_t)dt + (gh)^{-1}(X_t)F(X_t)dt + (gh)^{-1}(X_t)\sigma(X_t)dW_t, \tag{74}$$

with $\boldsymbol{S}(\boldsymbol{X}_t) = \boldsymbol{S}^{(1)}(\boldsymbol{X}_t) + \boldsymbol{S}^{(2)}(\boldsymbol{X}_t) + \boldsymbol{S}^{(3)}(\boldsymbol{X}_t)$, where the $\boldsymbol{S}^{(k)}$ are the noise-induced drifts whose ith components are given by

$$S_i^{(1)}(X) = m_0 \frac{\partial}{\partial X_l}[((gh)^{-1})_{ij}(X)](J_{11})_{jl}(X), \quad i, j, l = 1, \ldots, d, \tag{75}$$

$$S_i^{(2)}(X) = -\tau_\kappa \frac{\partial}{\partial X_l}[((Ah)^{-1})_{ij}(X)](J_{21})_{jl}(X), \quad i, j, l = 1, \ldots, d, \tag{76}$$

$$S_i^{(3)}(X) = \tau_\eta \frac{\partial}{\partial X_l}[((gh)^{-1}\sigma A^{-1})_{ij}(X)](J_{31})_{jl}(X), \quad i, j, l = 1, \ldots, d. \tag{77}$$

Here $\boldsymbol{J}_{11} = \boldsymbol{J}_{11}^*$, $\boldsymbol{J}_{21} = \boldsymbol{J}_{12}^*$ and $\boldsymbol{J}_{31} = \boldsymbol{J}_{13}^*$ are d by d block matrices satisfying the following system of matrix equations:

$$\tau_\eta g J_{23} + m_0 J_{13} A = \sigma A/2, \tag{78}$$

$$\tau_\eta Ah J_{13} = \tau_\eta A J_{23} + \tau_\kappa J_{23} A, \tag{79}$$

$$Ah J_{12} + J_{12}^* hA = A J_{22} + J_{22} A, \tag{80}$$

$$g J_{12}^* + J_{12} g = \sigma J_{13}^* + J_{13} \sigma^*, \tag{81}$$

$$m_0 J_{11} hA + \tau_\kappa \sigma J_{23}^* = \tau_\kappa g J_{22} + m_0 J_{12} A. \tag{82}$$

The convergence is obtained in the same sense as in Theorem 2.

Proof We will apply Theorem 2. As $\boldsymbol{K}_1 = 2\boldsymbol{K}_2 = \boldsymbol{I}$, clearly they are invertible. Also, being positive definite, $\boldsymbol{g}$ and $\boldsymbol{h}$ are invertible.

Since $\boldsymbol{g}, \boldsymbol{h}$ and $\boldsymbol{A}$ are positive definite and commuting matrices, the matrix $\boldsymbol{B}_\lambda(\boldsymbol{x})$, defined in (44), is invertible for all λ such that $Re(\lambda) > 0$. Indeed, in this case $\boldsymbol{B}_\lambda(\boldsymbol{x}) = \boldsymbol{I} + \boldsymbol{g}(\boldsymbol{x})(\lambda \tau_\kappa \boldsymbol{I} + \boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{h}(\boldsymbol{x})/\lambda m_0$. Since $\boldsymbol{g}, \boldsymbol{h}$ and $\boldsymbol{A}$ are positive definite and commuting, they have positive eigenvalues and can be simultaneously diagonalized. Therefore, all the eigenvalues of $\boldsymbol{B}_\lambda(\boldsymbol{x})$ are nonzero for every λ with $Re(\lambda) > 0$ and $\boldsymbol{x} \in \mathbb{R}^d$, so the invertibility condition is verified. Therefore, the block matrix:

$$\hat{\gamma}(x_t^\epsilon) = \begin{bmatrix} \mathbf{0} & \frac{g(x_t^\epsilon)}{m_0} & -\frac{\sigma(x_t^\epsilon)}{m_0} \\ -\frac{Ah(x_t^\epsilon)}{\tau_\kappa} & \frac{A}{\tau_\kappa} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{A}{\tau_\eta} \end{bmatrix}, \tag{83}$$

is positive stable (see Remark 8). The result then follows by applying Theorem 2. $\qquad\square$

For special one-dimensional systems, the form of the limiting equation can be made even more explicit.

Corollary 4 *In the one-dimensional case, we drop the boldface and write $X_t := X_t \in \mathbb{R}$, $g(X) := g(X)$, with $g : \mathbb{R} \to \mathbb{R}$, etc.. We assume that $h = g$ and $A := \alpha > 0$ is a constant. The homogenized equation is given by:*

$$dX_t = S(X_t)dt + g^{-2}(X_t)F(X_t)dt + g^{-2}(X_t)\sigma(X_t)dW_t, \tag{84}$$

with $S(X_t) = S^{(1)}(X_t) + S^{(2)}(X_t) + S^{(3)}(X_t)$, where the noise-induced drift terms $S^{(k)}(X_t)$ have the following explicit expressions that depend on the parameters m_0, τ_κ and τ_η:

$$S^{(1)}(X_t) = \left(\frac{1}{g^2(X_t)}\right)' \frac{\sigma(X_t)^2}{2g^2(X_t)} \left[\frac{\tau_\kappa^2 g^2(X_t) + m_0\alpha(\tau_\kappa + \tau_\eta)}{\tau_\eta^2 g^2(X_t) + m_0\alpha(\tau_\kappa + \tau_\eta)}\right], \tag{85}$$

$$S^{(2)}(X_t) = -\left(\frac{1}{g(X_t)}\right)' \frac{\sigma(X_t)^2 \tau_\kappa (\tau_\kappa + \tau_\eta)}{2g(X_t)[\tau_\eta^2 g^2(X_t) + m_0\alpha(\tau_\kappa + \tau_\eta)]}, \tag{86}$$

$$S^{(3)}(X_t) = \left(\frac{\sigma(X_t)}{g^2(X_t)}\right)' \frac{\sigma(X_t)\tau_\eta(\tau_\kappa + \tau_\eta)}{2[\tau_\eta^2 g^2(X_t) + m_0\alpha(\tau_\kappa + \tau_\eta)]}, \tag{87}$$

where the prime $'$ denotes derivative with respect to X_t.

Proof With $x_t^\epsilon := (x_t^\epsilon, z_t^\epsilon, \zeta_t^\epsilon) \in \mathbb{R}^3$ and $v_t^\epsilon := (v_t^\epsilon, y_t^\epsilon, \eta_t^\epsilon) \in \mathbb{R}^3$, SDEs (55), (56) become:

$$dx_t^\epsilon = v_t^\epsilon dt, \tag{88}$$

$$\epsilon dv_t^\epsilon = -\gamma(x_t^\epsilon)v_t^\epsilon dt + F(x_t^\epsilon)dt + \sigma dW_t, \tag{89}$$

where

$$\gamma(x_t^\epsilon) = \begin{bmatrix} 0 & \frac{g(x_t^\epsilon)}{m_0} & -\frac{\sigma(x_t^\epsilon)}{m_0} \\ -\frac{\alpha}{\tau_\kappa}g(x_t^\epsilon) & \frac{\alpha}{\tau_\kappa} & 0 \\ 0 & 0 & \frac{\alpha}{\tau_\eta} \end{bmatrix}, \quad F(x_t^\epsilon) = \begin{bmatrix} \frac{F(x_t^\epsilon)}{m_0} \\ 0 \\ 0 \end{bmatrix}, \quad \sigma = \begin{bmatrix} 0 \\ 0 \\ \frac{\alpha}{\tau_\eta} \end{bmatrix}. \tag{90}$$

It follows from Corollary 3 that the matrix γ is positive stable; one can also calculate its eigenvalues explicitly and see that their real parts are positive. The eigenvalues of γ are

$$\frac{\alpha}{\tau_\eta}, \quad \frac{\alpha}{2\tau_\kappa} \pm \frac{1}{2}\sqrt{\frac{\alpha^2 m_0 - 4\alpha g(x_t^\epsilon)^2 \tau_\kappa}{m_0 \tau_\kappa^2}}, \tag{91}$$

and so their real parts are indeed positive.

On the other hand, the solution, $J \in \mathbb{R}^{3\times 3}$, to the Lyapunov equation,

$$\gamma J + J\gamma^* = \sigma\sigma^*, \tag{92}$$

can be computed (using Mathematica®) to be:

$$J = \begin{bmatrix} \frac{\sigma^2}{2m_0 g^2}\left[\frac{\tau_\kappa^2 g^2 + m_0\alpha(\tau_\kappa + \tau_\eta)}{\tau_\eta^2 g^2 + m_0\alpha(\tau_\kappa + \tau_\eta)}\right] & \frac{\alpha\sigma^2(\tau_\kappa + \tau_\eta)}{2g(\tau_\eta^2 g^2 + m_0\alpha(\tau_\kappa + \tau_\eta))} & \frac{\alpha\sigma(\tau_\kappa + \tau_\eta)}{2(\tau_\eta^2 g^2 + m_0\alpha(\tau_\kappa + \tau_\eta))} \\ \frac{\alpha\sigma^2(\tau_\kappa + \tau_\eta)}{2g(\tau_\eta^2 g^2 + m_0\alpha(\tau_\kappa + \tau_\eta))} & \frac{\alpha\sigma^2(\tau_\kappa + \tau_\eta)}{2(\tau_\eta^2 g^2 + m_0\alpha(\tau_\kappa + \tau_\eta))} & \frac{\tau_\eta\alpha\sigma g}{2(\tau_\eta^2 g^2 + m_0\alpha(\tau_\kappa + \tau_\eta))} \\ \frac{\alpha\sigma(\tau_\kappa + \tau_\eta)}{2(\tau_\eta^2 g^2 + m_0\alpha(\tau_\kappa + \tau_\eta))} & \frac{\tau_\eta\alpha\sigma g}{2(\tau_\eta^2 g^2 + m_0\alpha(\tau_\kappa + \tau_\eta))} & \frac{\alpha}{2\tau_\eta} \end{bmatrix}. \tag{93}$$

The result then follows from Corollary 3. $\qquad\square$

Remark 11 Note that here the matrix γ in (90) is not symmetric and the smallest eigenvalue of its symmetric part can be negative. Moreover, the initial condition v_0^ϵ depends on ϵ through the component η_0^ϵ (which is a zero mean Gaussian random variable with variance $\frac{\alpha}{2\epsilon\tau_\eta}$). Thus, we cannot apply the main results in [8] to obtain the convergence result. This is our main motivation to revisit the Smoluchowski–Kramers limit of SDEs in Sect. 3 under a weakened spectral assumption on the matrix γ (or $\hat{\gamma}$ in the multidimensional case) and a relaxed assumption concerning the ϵ dependence of v_0^ϵ (or $\hat{v}_0^\epsilon$ in the multidimensional case).

Remark 12 In the important case when the fluctuation–dissipation relation (i.e. $\tau_\kappa = \tau_\eta$, $h = g$ and g is proportional to σ) holds for the one-dimensional models of the first sub-class, the correction drift terms $S^{(2)}$ and $S^{(3)}$ cancel each other and the resulting (single) noise-induced drift term coincides with that obtained in the limit as $m \to 0$ of the systems with no memory, driven by a white noise to which Theorem 1 applies directly! However, when the relation fails, we obtain three different drift corrections induced by vanishing of all time scales. Again, the presence of these correction terms may have significant consequences for the dynamics of the systems (see Sect. 6).

5.2 SIDEs Driven by a Non-Markovian Colored Noise

The following corollary provides a homogenized SDE for the particle position in the limit, in which the inertial time scale, the memory time scale and the noise correlation time scale vanish at the same rate in the case when the pre-limit dynamics are driven by the harmonic noise. We emphasize that in this case the original system is driven by a noise which is not a Markov process.

Corollary 5 *Let $d = d_1 = d_2 = q_1 = q_2 = q = r$. We set, in the SDEs (37)–(42): $\tau_\xi = \tau_h$, $W_t^{(q_2)} = W_t^{(d)} = W_t$ and*

$$\boldsymbol{\Gamma}_2 = \begin{bmatrix} \mathbf{0} & -\boldsymbol{I} \\ \boldsymbol{\Omega}^2 & \boldsymbol{\Omega}^2 \end{bmatrix}, \quad \boldsymbol{\Gamma}_1 = \frac{1}{2}\begin{bmatrix} \boldsymbol{\Omega}^2 & 4\boldsymbol{I} - \boldsymbol{\Omega}^2 \\ -\boldsymbol{\Omega}^2 & \boldsymbol{\Omega}^2 \end{bmatrix} =: \boldsymbol{T}\boldsymbol{\Gamma}_2\boldsymbol{T}^{-1}, \tag{94}$$

$$\boldsymbol{M}_2 = \frac{1}{2}\begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}^2 \end{bmatrix}, \quad \boldsymbol{M}_1 = 2\boldsymbol{T}\boldsymbol{M}_2\boldsymbol{T}^*, \tag{95}$$

$$\boldsymbol{C}_2 = [\boldsymbol{I} \ \ \mathbf{0}], \quad \boldsymbol{C}_1 = \boldsymbol{C}_2\boldsymbol{T}^{-1}, \tag{96}$$

to obtain SDEs equivalent to equations (27)–(29) with $\mu = \theta = \nu = 1$. Let $x_t^\epsilon \in \mathbb{R}^d$ be the solution to the SDEs (37)–(42), with the matrices $g(x)$ and $h(x)$ positive definite for every $x \in \mathbb{R}^d$. Moreover, $g(x)$, $h(x)$ and the diagonal matrix $\boldsymbol{\Omega}^2$ are commuting. Suppose that Assumptions 5–7 are satisfied.

Then as $\epsilon \to 0$, the process x_t^ϵ converges to the solution, X_t, of the following Itô SDE

$$dX_t = S(X_t)dt + (gh)^{-1}(X_t)F(X_t)dt + (gh)^{-1}(X_t)\sigma(X_t)dW_t, \tag{97}$$

with $S(X_t) = S^{(1)}(X_t) + S^{(2)}(X_t) + S^{(3)}(X_t)$, where the $S^{(k)}$ are the noise-induced drift terms whose ith components are given by the expressions

$$S_i^{(1)}(X) = m_0\frac{\partial}{\partial X_l}[((gh)^{-1})_{ij}(X)](J_{11})_{jl}(X), \tag{98}$$

$$S_i^{(2)}(X) = -\tau_\kappa\left(\frac{\partial}{\partial X_l}[(h^{-1})_{ij}(X)](J_{21})_{jl}(X) + \frac{\partial}{\partial X_l}[h^{-1}(I - 2\boldsymbol{\Omega}^{-2}))_{ij}(X)](J_{31})_{jl}(X)\right), \tag{99}$$

$$S_i^{(3)}(X) = \tau_h\left(\frac{\partial}{\partial X_l}[((gh)^{-1}\sigma)_{ij}(X)](J_{41})_{jl}(X) + \frac{\partial}{\partial X_l}[((gh)^{-1}\sigma\boldsymbol{\Omega}^{-2})_{ij}(X)](J_{51})_{jl}(X)\right), \tag{100}$$

where $i, j, l = 1, \ldots, d$. In the above,

$$\hat{\boldsymbol{J}} := \begin{bmatrix} \boldsymbol{J}_{11} & \cdots & \boldsymbol{J}_{15} \\ \vdots & \ddots & \vdots \\ \boldsymbol{J}_{51} & \cdots & \boldsymbol{J}_{55} \end{bmatrix} \in \mathbb{R}^{5d \times 5d}, \quad \text{where } \boldsymbol{J}_{kl} \in \mathbb{R}^{d \times d}, \ k, l = 1, \ldots, 5, \tag{101}$$

is the block matrix solving the Lyapunov equation

$$\hat{J}\hat{\gamma}^* + \hat{\gamma}\hat{J} = \hat{\sigma}\hat{\sigma}^*, \tag{102}$$

and

$$\hat{\gamma} = \begin{bmatrix} \mathbf{0} & \frac{g(X)}{m_0} & \frac{g(X)}{m_0} & -\frac{\sigma(X)}{m_0} & \mathbf{0} \\ -\frac{h(X)}{\tau_\kappa} & \frac{\Omega^2}{2\tau_\kappa} & \frac{2I}{\tau_\kappa} - \frac{\Omega^2}{2\tau_\kappa} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\frac{\Omega^2}{2\tau_\kappa} & \frac{\Omega^2}{2\tau_\kappa} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\frac{1}{\tau_h}I \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{\Omega^2}{\tau_h} & \frac{\Omega^2}{\tau_h} \end{bmatrix}, \quad \hat{\sigma} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \frac{\Omega^2}{\tau_h} \end{bmatrix}. \tag{103}$$

In the above, $\hat{\gamma} \in \mathbb{R}^{5d \times 5d}$ is a 5 by 5 block matrix with each block an $\mathbb{R}^{d \times d}$-valued matrix, $\hat{\sigma} \in \mathbb{R}^{5d \times d}$ is a 5 by 1 block matrix with each block a $\mathbb{R}^{d \times d}$-valued matrix, I is a $d \times d$ identity matrix, $\mathbf{0}$ in $\hat{\gamma}$ and $\hat{\sigma}$ is a $d \times d$ zero matrix, and W is a d-dimensional Wiener process. The convergence is obtained in the same sense as in Theorem 2.

Note that the oscillatory nature of covariance function of the harmonic noise in the pre-limit SIDE makes the noise-induced drift in the resulting limiting SDE more complicated (there are more terms) compared to the case of OU process in the first sub-class. Therefore, we write the system of matrix equations that the J_{kl} satisfy in the form of a matrix Lyapunov equation in Corollary 5, without breaking it up into equations for individual blocks. This could of course be done, leading to a (more complicated) analog of (78)–(82). The proof of Corollary 5 is essentially identical to the proof of Corollary 3, so we omit it.

Again, for special one-dimensional systems, we are going to make the result more explicit.

Corollary 6 *In the one-dimensional case, we drop the boldface and write $X_t := X_t \in \mathbb{R}$, $g(x) := g(x)$, with $g : \mathbb{R} \to \mathbb{R}$, etc.. We assume that $h = g$ and $\Omega := \Omega$ is a real constant. The homogenized equation is given by:*

$$dX_t = S(X_t)dt + g^{-2}(X_t)F(X_t)dt + g^{-2}(X_t)\sigma(X_t)dW_t, \tag{104}$$

with $S(X_t) = S^{(1)}(X_t) + S^{(2)}(X_t) + S^{(3)}(X_t)$, where the noise-induced drift terms $S^{(k)}(X_t)$ have the following explicit expressions (computed using Mathematica®) that depend on the parameters m_0, τ_κ and τ_h:

$$S^{(1)}(X) = m_0 \left(\frac{1}{g^2(X)}\right)' J_{11}(X), \tag{105}$$

$$S^{(2)}(X) = -\tau_\kappa \left(\frac{1}{g(X)}\right)' \left(J_{21}(X) + \left(1 - \frac{2}{\Omega^2}\right)J_{31}(X)\right), \tag{106}$$

$$S^{(3)}(X) = \tau_h \left(\frac{\sigma(X)}{g^2(X)}\right)' \left(J_{41}(X) + \frac{1}{\Omega^2}J_{51}(X)\right), \tag{107}$$

where the prime $'$ denotes derivative with respect to X and the $J_{kl}(X)$ are given by:

$$J_{11}(X) = \frac{\sigma^2}{2m_0 g^2 R(X)} \Big(g^4 \tau_\kappa^4 (\tau_\kappa^2 + \tau_\kappa \tau_h \Omega^2 + \tau_h^2 \Omega^2) + m_0^2 \Omega^4 (\tau_\kappa + \tau_h)^2 (\tau_\kappa^2 + \tau_h^2$$
$$+ \tau_\kappa \tau_h (\Omega^2 - 2)) + m_0 \Omega^2 g^2 (\tau_\kappa + \tau_h)[\tau_h^4 + \tau_\kappa^2 \tau_h^2 (\Omega^2 - 2) + \tau_\kappa^4 (\Omega^2 - 1)$$
$$+ \tau_\kappa^3 \tau_h (2 - 3\Omega^2 + \Omega^4)] \Big) \tag{108}$$

$$J_{21}(X) = \frac{\sigma^2(\tau_\kappa + \tau_h)\Omega^2}{4gR(X)}\Big(m_0\Omega^2(\tau_\kappa + \tau_h)(\tau_\kappa^2 + \tau_h^2 + \tau_\kappa\tau_h(\Omega^2 - 2))$$

$$+ g^2(\tau_\kappa^4 + \tau_\kappa^2\tau_h^2 + \tau_h^4 + \tau_\kappa^3\tau_h(\Omega^2 - 1))\Big), \tag{109}$$

$$J_{31}(X) = -\frac{\sigma^2(\tau_\kappa + \tau_h)\Omega^2}{4gR(X)}\Big(-m_0\Omega^2(\tau_\kappa + \tau_h)(\tau_\kappa^2 + \tau_h^2 + \tau_\kappa\tau_h(\Omega^2 - 2))$$

$$+ g^2(\tau_\kappa^4 + \tau_\kappa^2\tau_h^2 - \tau_h^4 + \tau_\kappa^3\tau_h(\Omega^2 - 1))\Big), \tag{110}$$

$$J_{41}(X) = \frac{1}{2}\Big(\sigma\Omega^2(\tau_\kappa + \tau_h)[g^2\tau_h^4 + m_0\Omega^2(\tau_\kappa + \tau_h)(\tau_\kappa^2 + \tau_h^2 + \tau_\kappa\tau_h(\Omega^2 - 2))]\Big), \tag{111}$$

$$J_{51}(X) = -\frac{1}{2}\Big(\sigma\Omega^2(\tau_\kappa + \tau_h)[m_0\Omega^2(\tau_\kappa + \tau_h)(\tau_\kappa^2 + \tau_h^2 + \tau_\kappa\tau_h(\Omega^2 - 2))$$

$$- g^2\tau_\kappa\tau_h^2(\tau_\kappa + \tau_h(\Omega^2 - 1))]\Big), \tag{112}$$

where $g = g(X)$, $\sigma = \sigma(X)$ and

$$R(X) = g^4\tau_h^4(\tau_\kappa^2 + \tau_\kappa\tau_h\Omega^2 + \tau_h^2\Omega^2) + m_0^2\Omega^4(\tau_\kappa + \tau_h)^2(\tau_\kappa^2 + \tau_h^2 + \tau_\kappa\tau_h(\Omega^2 - 2))$$

$$+ g^2m_0\tau_h^2\Omega^2[\tau_h^3\Omega^2 + \tau_\kappa^3(\Omega^2 - 2) + \tau_\kappa^2\tau_h\Omega^2(\Omega^2 - 2) + \tau_\kappa\tau_h^2(2 - 2\Omega^2 + \Omega^4)]. \tag{113}$$

Note that if we send $\Omega \to \infty$ in the expressions for the $S^{(i)}(X)$ ($i = 1, 2, 3$) above, we recover the corresponding expressions given in Corollary 4 (with $\alpha = 1$). This is not surprising, since in this limit the harmonic noise becomes an OU process (with $\alpha = 1$).

Moreover, when $\tau_\kappa = \tau_h = \tau$, the noise-induced drift becomes $S(X) = S^{(1)}(X) + S^{(2)}(X) + S^{(3)}(X)$, where

$$S^{(1)} = \frac{1}{2}\left(\frac{1}{g^2}\right)'\frac{\sigma^2}{g^2}, \tag{114}$$

$$S^{(2)} = -\frac{2\tau\Omega^2\sigma^2}{g}\left(\frac{1}{g}\right)'\left(\frac{g^2\tau + m_0\Omega^2(\Omega^2 - 1)}{4m_0^2\Omega^6 + 2g^2m_0\tau\Omega^4(\Omega^2 - 1) + g^4\tau^2(1 + 2\Omega^2)}\right), \tag{115}$$

$$S^{(3)} = 2\tau\Omega^2\sigma\left(\frac{\sigma}{g^2}\right)'\left(\frac{g^2\tau + m_0\Omega^2(\Omega^2 - 1)}{4m_0^2\Omega^6 + 2g^2m_0\tau\Omega^4(\Omega^2 - 1) + g^4\tau^2(1 + 2\Omega^2)}\right). \tag{116}$$

Again, in the case when the fluctuation–dissipation relation holds we see that the noise-induced drift coincides with that obtained in the limit as $m \to 0$ of the Markovian model in Sect. 3.

6 Application to the Study of Thermophoresis

6.1 Introduction

We revisit the dynamics of a free Brownian particle immersed in a heat bath where a temperature gradient is present. This was previously studied in [42]. It was found there that the particle experiences a drift in response to the temperature gradient, due to the interplay

between the inertial time scale and the noise correlation time scale. Such phenomenon is called *thermophoresis*. We refer to [42,43] and the references therein for further descriptions of this phenomenon, including references to experiments.

Here, we will study the dynamics of the particle in a non-equilibrium heat bath, where a *generalized fluctuation–dissipation relation* holds, in which both the diffusion coefficient and the temperature of the heat bath vary with the position. In contrast to [42], we take into account also the memory time scale (in addition to the inertial time scale and the noise correlation time scale) and model the position of the particle as the solution to a SIDE of the form (11). Unlike the model used in [42], the model can be derived heuristically from microscopic dynamics by an argument very similar to that of Appendix A.

For a spherical particle of radius R immersed in a fluid of viscosity μ, which in general is a function of the temperature $T = T(x)$ (and thus depends on x as well), the friction (or damping) coefficient γ satisfies the Stokes law [13]:

$$\gamma(x) = 6\pi\mu(T)R. \tag{117}$$

On the other hand, the damping coefficient $\gamma(x)$ and the noise coefficient $\sigma(x)$ are expressed in terms of the diffusion coefficient $D(x)$ and the temperature $T(x)$ as follows:

$$\gamma(x) = \frac{k_B T(x)}{D(x)}, \quad \sigma(x) = \frac{k_B T(x)\sqrt{2}}{\sqrt{D(x)}}. \tag{118}$$

In the following, we study two one-dimensional non-Markovian models of thermophoresis. The first model is driven by a Markovian colored noise and the second model by a non-Markovian one.

6.2 A Thermophoresis Model with Ornstein–Uhlenbeck Noise

In this section we model evolution of the position, $x_t \in \mathbb{R}$, of a particle by the following SIDE:

$$m\ddot{x}_t = -\sqrt{\gamma(x_t)}\int_0^t \alpha e^{-\alpha(t-s)}\sqrt{\gamma(x_s)}\dot{x}_s ds + \sigma(x_t)\eta_t, \tag{119}$$

where η_t is a stationary process, satisfying the SDE:

$$d\eta_t = -\alpha\eta_t dt + \alpha dW_t. \tag{120}$$

The above equations are obtained by setting $d = 1$, $\boldsymbol{F} = 0$, $\boldsymbol{h} = \boldsymbol{g} = g := \sqrt{\gamma}$, $\boldsymbol{\sigma} = \sigma$ in (15) and $A = \alpha$ in (13), where γ and σ are given by (118). Note that the noise correlation function is proportional to the memory kernel in the SIDE (119), i.e.

$$E[\eta_t\eta_s] = \frac{\alpha}{2}e^{-\alpha|t-s|} = \frac{1}{2}\kappa_1(t-s), \ s, t \geq 0 \tag{121}$$

as in (21). Together with (118), this implies that (119) satisfies the generalized fluctuation–dissipation relation (see the statement of Corollary 2 and Remark 12). Note also that g is a constant multiple of σ if and only if T is position-independent.

We now consider the effective dynamics of the particle in the limit when all the three characteristic time scales vanish at the same rate. In the following, the prime $'$ denotes derivative with respect to the argument of the function.

Corollary 7 *Let $\epsilon > 0$ be a small parameter and let the particle's position, $x_t^\epsilon \in \mathbb{R}$ $(t \geq 0)$, satisfy the following rescaled version of* (119), (120):

$$dx_t^\epsilon = v_t^\epsilon dt, \tag{122}$$

$$m_0 \epsilon dv_t^\epsilon = \sigma(x_t^\epsilon)\eta_t^\epsilon dt - \sqrt{\gamma(x_t^\epsilon)} \left(\int_0^t \frac{\alpha}{\tau \epsilon} e^{-\frac{\alpha}{\tau \epsilon}(t-s)} \sqrt{\gamma(x_s^\epsilon)} v_s^\epsilon ds \right) dt, \tag{123}$$

$$\tau \epsilon d\eta_t^\epsilon = -\alpha \eta_t^\epsilon dt + \alpha dW_t, \tag{124}$$

where m_0, α, τ are positive constants, and (W_t) is a one-dimensional Wiener process. The initial conditions are random variables $x_0^\epsilon = x$, $v_0^\epsilon = v$, independent of ϵ and (statistically) independent of (W_t), and η_0^ϵ is distributed according to the invariant distribution of the SDE (124).

Assume that the assumptions of Corollary 3 are satisfied (in particular, $\gamma(x) > 0$ for every $x \in \mathbb{R}$). Then, in the limit as $\epsilon \to 0$, x_t^ϵ converges (in the same sense as in Corollary 3) to the process $X_t \in \mathbb{R}$, satisfying the SDE:

$$dX_t = b_1(X_t)dt + \sqrt{2D(X_t)}dW_t, \tag{125}$$

with the noise-induced drift, $b_1(X) = S^{(1)}(X) + S^{(2)}(X) + S^{(3)}(X)$, where

$$S^{(1)}(X) = D'(X) - \frac{D(X)T'(X)}{T(X)}, \tag{126}$$

$$S^{(2)}(X) = \left[-\frac{k_B T(X)D'(X)}{D(X)} + k_B T'(X) \right] \cdot \left[\frac{\tau D(X)}{\tau k_B T(X) + 2m_0 \alpha D(X)} \right], \tag{127}$$

$$S^{(3)}(X) = \left[\frac{k_B T(X)D'(X)}{D(X)} \right] \cdot \left[\frac{\tau D(X)}{\tau k_B T(X) + 2m_0 \alpha D(X)} \right]. \tag{128}$$

Proof The corollary follows from Corollary 4. In particular, the expressions for $S^{(1)}$, $S^{(2)}$ and $S^{(3)}$ follow from applying Corollary 4 to the present system (see (85)–(87)). □

We give some remarks and discussions of the contents of Corollary 7 before we end this subsection.

Remark 13 We see that in this case a part of $S^{(2)}$ cancels $S^{(3)}$ and therefore the noise-induced drift simplifies to:

$$b_1(X) = D'(X) - \frac{2m_0 \alpha D^2(X)}{\tau k_B T(X) + 2m_0 \alpha D(X)} \frac{T'(X)}{T(X)}. \tag{129}$$

Using the Stokes law (117) which gives

$$D(X) = \frac{k_B}{6\pi R} \frac{T(X)}{\mu(T)}, \tag{130}$$

where $\mu(T) = \mu(T(X))$, we have

$$b_1(X) = k_B T'(X) \left(\frac{\tau}{2(\alpha m_0 + 3\pi R\tau \mu(T))} - \frac{\mu'(T)T(X)}{6\pi R\mu^2(T)} \right). \tag{131}$$

Equation (129) gives the thermophoretic drift in the limit when the three characteristic time scales vanish. Since it arises in the absence of an external force acting on the particle, it is a "spurious drift" caused by the presence of the temperature gradient and the state-dependence of the diffusion coefficient. Compared to Eq. (101) in [8], the drift term derived here contains a correction term due to the temperature profile.

Discussion We discuss some physical implications of the thermophoretic drift given in (129). As discussed in [42], the sign of $b_1(X)$ determines the direction in which the particle is expected to travel. The particle will eventually reach some boundaries, which can be either absorbing or reflecting. We are going to consider the reflecting boundaries case.

In the reflecting boundaries case, the position of the particle, satisfying the SDE (125), reaches a steady-state distribution $\rho_\infty(X)$ in the limit $t \to \infty$. Assuming that the particle is confined to the interval (a, b), $a < b$, one can compute the stationary density:

$$\rho_\infty(X) = C \exp\left(-\int_a^X \frac{2\alpha}{r\gamma(y) + 2\alpha} \frac{T'(y)}{T(y)} dy\right), \tag{132}$$

where in terms of the original parameters of the model, $r := \tau/m_0 > 0$, and C is a normalizing constant. In particular, in absence of temperature gradient ($T'(y) = 0$), the particle is equally likely to be found anywhere in (a, b), whereas when a temperature gradient is present, the distribution of the particle's position is not uniform. In the limit $r \to \infty$, the particle's position is again distributed uniformly on (a, b). On the other hand, in the limit $r \to 0$ the stationary density is inversely proportional to the temperature, i.e. $\rho_\infty(X) = \tilde{C}T(X)^{-1}$, where $\tilde{C}$ is a normalizing constant. Thus, the particle is more likely to be found in the colder region. In the special case when $D(X)$ is proportional to $T(X)$, so that γ is independent of X, we have

$$\rho_\infty(X) = \tilde{C}T(X)^{-\frac{2\alpha}{2\alpha+r\gamma}}, \tag{133}$$

where $\tilde{C}$ is a normalizing constant, so the particle is more likely to be found in the colder region, with the likelihood decreasing as r increases.

Next, we are going to study the sign of the thermophoretic drift directly using (129) (this is in contrast to the approach in [42], where $\mu(T)$ is expanded around a fixed temperature). We find that $b_1(X) > 0$ if and only if $r > r_c$ and r_c is the critical ratio of τ/m_0, given by:

$$r_c = \frac{\alpha}{3\pi R\mu(T)} \left(\frac{\mu'(T)T(X)}{\mu(T) - \mu'(T)T(X)}\right), \tag{134}$$

where $\mu(T) = \mu(T(X))$ is obtained from the Stokes law. For $r = r_c$, the stationary density (132) reduces to:

$$\rho_\infty^c(X) = C\frac{\mu(T(X))}{T(X)}, \tag{135}$$

where C is a normalizing constant. Importantly, note that the drift does not change sign if T is independent of X.

We now discuss a special case. When $\mu(T) = \mu_0 > 0$ is a constant (so that $\gamma(X)$ is a constant), the thermophoretic drift is given by:

$$b_1(X) = \frac{k_B T'(X)}{6\pi R\mu_0} \left[1 - \frac{\alpha}{\alpha + 3\pi r R\mu_0}\right]. \tag{136}$$

In agreement with the result in [42], $b_1(X)$ has the same sign as $T'(X)$, leading to a flow towards the hotter region. The steady-state density is

$$\rho_\infty(X) = CT(X)^{-\frac{\alpha}{\alpha+3\pi r R\mu_0}}, \tag{137}$$

where C is a normalizing constant, and the particle is more likely to be found in the colder region for all $r > 0$, even though the thermophoretic drift actually directs the particle towards the hotter regions. This effect is in agreement with experiments, and is explained by the presence of reflecting boundary conditions.

Remark 14 Strictly speaking, reflecting boundary conditions should be first considered for the positive-epsilon version of the process. For instance, at the moment of hitting a boundary point, the velocity can be required to change its sign instantaneously. One has to ask whether such a modification might have an effect on the limiting process. One could, in principle, resolve this problem by adopting the following strategy. Instead of reflecting from the boundaries, the particle may continue going in the same direction, with the coefficients of the equation obtained by reflection in the boundary point (i.e. $\sigma(b + x_t^\epsilon) = \sigma(b - x_t^\epsilon)$, where b is one of the boundary points etc.). The above asymptotic analysis may be conducted for the resulting system. The coefficients of this system are periodic (with the period $2(b - a)$) and continuous, but nondifferentiable at isolated points, so this would involve additional technical work. Also, one would have to prove that putting the limiting system back on a bounded interval indeed gives rise to a process reflecting at the boundaries. We do not pursue the details here.

6.3 A Thermophoresis Model with Non-Markovian (Harmonic) Noise

We repeat the analysis of the previous subsection in the case when the colored noise is a harmonic noise. We set $d = 1$, $\boldsymbol{F} = 0$, $\boldsymbol{h} = \boldsymbol{g} = g := \sqrt{\gamma}$, $\boldsymbol{\sigma} = \sigma$, $\boldsymbol{\Omega} = \Omega$, $\boldsymbol{\Omega}_0 = \Omega_0 := \Omega\sqrt{1 - \Omega^2/4}$, $\boldsymbol{\Omega}_1 = \Omega_1 := \Omega/\sqrt{1 - \Omega^2/4}$ (where $|\Omega| < 2$) in the SIDE (20) and study the effective dynamics of the resulting system as before. The case where $|\Omega| > 2$ can be studied analogously. The following result follows from Corollary 6.

Corollary 8 *Let $\epsilon > 0$ be a small parameter and the particle's position, $x_t^\epsilon \in \mathbb{R}$ ($t \geq 0$), satisfy the following rescaled SDEs:*

$$dx_t^\epsilon = v_t^\epsilon dt, \tag{138}$$

$$m_0\epsilon dv_t^\epsilon = -\frac{\sqrt{\gamma(x_t^\epsilon)}}{\tau\epsilon}\left(\int_0^t e^{-\frac{\Omega^2}{2\tau\epsilon}(t-s)}\left[\cos\left(\frac{\Omega_0}{\tau\epsilon}(t-s)\right)\right.\right.$$
$$\left.\left. +\frac{\Omega_1}{2}\sin\left(\frac{\Omega_0}{\tau\epsilon}(t-s)\right)\right]\sqrt{\gamma(x_s^\epsilon)}v_s^\epsilon ds\right)dt + \sigma(x_t^\epsilon)h_t^\epsilon dt, \tag{139}$$

$$\tau\epsilon dh_t^\epsilon = u_t^\epsilon dt, \tag{140}$$

$$\tau\epsilon du_t^\epsilon = -\Omega^2 u_t^\epsilon dt - \Omega^2 h_t^\epsilon dt + \Omega^2 dW_t, \tag{141}$$

where m_0 and τ are positive constants, Ω, Ω_0 and Ω_1 are constants defined as before, and (W_t) is a one-dimensional Wiener process. The initial conditions are given by the random variables $x_0^\epsilon = x$, $v_0^\epsilon = v$, independent of ϵ, and $(h_0^\epsilon, u_0^\epsilon)$ are distributed according to the invariant measure of the SDEs (140), (141).

Assume that the assumptions in Corollary 5 are satisfied. Then, in the limit as $\epsilon \to 0$, the process x_t^ϵ converges (in the same sense as Corollary 5) to the process $X_t \in \mathbb{R}$, satisfying the SDE:

$$dX_t = b_2(X_t)dt + \sqrt{2D(X_t)}dW_t, \tag{142}$$

where the noise-induced drift term is given by:

$$b_2(X) = D'(X) \tag{143}$$

$$-\frac{(4m_0^2\Omega^6 D^2(X) + \tau^2(k_BT(X))^2)D(X)}{4m_0^2\Omega^6 D^2(X) + 2k_BT(X)m_0\tau\Omega^4(\Omega^2 - 1)D(X) + \tau^2(1 + 2\Omega^2)(k_BT(X))^2}\frac{T'(X)}{T(X)}. \tag{144}$$

We next discuss the contents of Corollary 8.

Remark 15 Note that $b_2(X)$ differs from $b_1(X)$ obtained previously and $b_2(X) \to b_1(X)$, with $\alpha = 1$ in the expression for $b_1(X)$, in the limit $\Omega \to \infty$.

Discussion In the reflecting boundaries case, the stationary distribution of the particle's position is

$$\rho_\infty(X)$$
$$= C \exp\left(-\int_a^X \frac{D(y)(4\Omega^6 D^2(y) + r^2(k_B T(y))^2)}{4\Omega^6 D^2(y) + 2r\Omega^4(\Omega^2 - 1)D(y)k_B T(y) + r^2(1 + 2\Omega^2)(k_B T(y))^2} \frac{T'(y)}{T(y)} dy\right),$$
(145)

where $r := \tau/m_0 > 0$ and C is a normalizing constant. Similarly to the previous model, in the absence of temperature gradient (i.e. when T is a constant), the particle is equally likely to be found anywhere in (a, b). When a temperature gradient is present, distribution of the particle's position is not uniform. However, in contrast to the previous model, in the limit $r \to \infty$ the particle is not distributed uniformly on (a, b) and in the limit $r \to 0$ the stationary density is no longer inversely proportional to the temperature. Both distributions depend on the diffusion coefficient $D(X)$ as well as on the temperature profile $T(X)$.

We can also study the sign of the thermophoretic drift. In this case there can be up to two critical ratios, r_c, at which $b_2(X)$ changes sign, as the equation $b_2(X) = 0$ is a quadratic equation in r. In the special case when $\mu(T) = \mu_0 > 0$ is a constant (and thus so is $\gamma(X)$), the thermophoretic drift is given by:

$$b_2(X) = \frac{k_B T'(X)}{6\pi R \mu_0}\left[1 - \frac{\Omega^6 + 9\pi^2 R^2 r^2 \mu_0^2}{\Omega^6 + 3\pi R r \Omega^4(\Omega^2 - 1)\mu_0 + 9\pi^2 R^2 r^2(1 + 2\Omega^2)\mu_0^2}\right]. \quad (146)$$

In contrast to the result in previous model, $b_2(X)$ has the same sign as $T'(X)$ provided that

$$r > \frac{\Omega^2(1 - \Omega^2)}{6\pi R \mu_0}. \quad (147)$$

Thus, $b_2(X)$ and $T'(X)$ do not share the same sign for all $r > 0$, unless $|\Omega| \geq 1$. According to this model, presence of a temperature gradient allows us to tune the parameters (m_0, τ, Ω) to control the direction which the particle travels. The steady-state density in this case is

$$\rho_\infty(X) = CT(X)^{-\frac{\Omega^6 + 9\pi^2 R^2 r^2 \mu_0^2}{\Omega^6 + 3\pi R r \Omega^4(\Omega^2 - 1)\mu_0 + 9\pi^2 R^2 r^2(1 + 2\Omega^2)\mu_0^2}}, \quad (148)$$

where C is a normalizing constant. The particle will be more likely found in the colder region for all $r > 0$ if $|\Omega| \geq 1$, whereas this might not be true for all $r > 0$ if $|\Omega| < 1$.

7 Conclusions and Final Remarks

We have studied homogenization of a class of GLEs in the limit when three characteristic time scales, i.e. the inertial time, the characteristic memory time in the damping term, and the correlation time of colored noise driving the equations, vanish at the same rate. We have derived effective equations, which are simpler in three respects:

1. The velocity variables have been homogenized. As a result, the number of degrees of freedom is reduced and there are no fast variables left.

2. The equations become regular SDEs, since the memory time has been taken to zero.
3. The system is driven by a white noise.

Importantly, *noise-induced drifts* are present in the limiting equations, resulting from the dependence of the coefficients of the original model on the state of the system. We have applied the general results to a study of thermophoretic drift, correcting the formulae obtained in an earlier work [42]. In systems, satisfying a fluctuation–dissipation relation, the noise-induced drifts in the limiting SDEs for the particle's position reduce to a single term, and for special cases the limiting SDEs coincide with that of [8]. However, in the more general case, new terms appear, absent in the case without memory. To prove the main theorem, we have employed the main result of [8], proven here in a different version under a relaxed assumption on the damping matrix and the initial conditions.

Homogenization of other specific non-Markovian models can also be studied using the methods of this paper. An example is a system with exponentially decaying memory kernel, driven by white noise in the limit as the inertial and memory time scales vanish at the same rate. In this case the noise-induced drift in the limiting equation will consist of two terms, not three, as in the case studied here. Moreover, one could also study the case when the time scales of the system do not vanish at the same rate, along the lines of [3].

The colored noises considered in this paper have correlations decaying exponentially (short-range memory). It would be interesting to study cases where the GLE is driven by other colored noises such as fractional Gaussian noises, with covariances decaying as a power, relevant for modeling anomalous diffusion phenomena in fields ranging from biology to finance [44]. As mentioned in Sect. 2, we will explore homogenization for GLEs with vanishing effective damping and diffusion constant in a future work.

Appendices

Appendix A: Derivation of SIDEs From a Hamiltonian Model

For completeness, we provide a derivation for a special case of SIDEs (11), (15) and (20) from a Hamiltonian model of a small system (Brownian particle) in contact with a heat bath in thermal equilibrium. The particle is moving in a potential U. The heat bath is modeled as a system of non-interacting harmonic oscillators whose initial energy is distributed according to the Gibbs distribution at temperature T. The Brownian particle is coupled to each oscillator in the bath. This model is used widely to study many systems in statistical physics [11, 45]. Our goal is to derive, heuristically, a stochastic integro-differential equation (SIDE) for the position and momentum variables of the particle from the Hamiltonian dynamics. This derivation serves to motivate the class of SIDEs that we are studying in this paper. We emphasize that our derivation here is certainly not original and follows closely that in [46] (see also an abstract approach in [47]).

One approach to derive the equations is to assume first that there are finitely many harmonic oscillators in the bath (Kac–Zwanzig model [32,48]). We then take the thermodynamic limit by sending the number of oscillators to infinity in the resulting equations (replacing finite

sum over oscillator frequencies by an integral), arguing that the set of frequencies must be dense to allow dissipation of energy from the system to the bath and to eliminate Poincaré recurrence. Another approach, which is more technically involved, is to replace the finite system of oscillator equations by a system modeled by a wave equation [27,49]. We will derive the SIDEs by adopting the former approach in the multi-dimensional case.

We consider the situation where the coupling is nonlinear in the particle's position and linear in the bath variables. Let $\hat{\boldsymbol{x}} = (\boldsymbol{x}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \in \mathbb{R}^{d+Np}$ and $\hat{\boldsymbol{p}} = (\boldsymbol{p}, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N) \in \mathbb{R}^{d+Np}$ (here, $\boldsymbol{x}, \boldsymbol{p} \in \mathbb{R}^d$ and $\boldsymbol{x}_i, \boldsymbol{p}_i \in \mathbb{R}^p$ for $i = 1, \ldots, N$). Hereafter, the superscript $*$ denotes transposition and $|\boldsymbol{b}|^2 := \boldsymbol{b}^* \boldsymbol{b} = \sum_{k=1}^n b_k^2$ denotes square of the norm of vector $\boldsymbol{b} := (b_1, \ldots, b_n) \in \mathbb{R}^n$.

The Hamiltonian of the system plus bath is:

$$H(\hat{\boldsymbol{x}}, \hat{\boldsymbol{p}}) = \frac{|\boldsymbol{p}|^2}{2m} + U(\boldsymbol{x}) + \sum_{k=1}^N \left(\frac{|\boldsymbol{p}_k|^2}{2} + \frac{1}{2}\omega_k^2 \left| \boldsymbol{x}_k - \frac{\boldsymbol{c}_k^*}{\omega_k^2} \boldsymbol{f}(\boldsymbol{x}) \right|^2 \right), \qquad (149)$$

where m is the particle's mass, $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{p} \in \mathbb{R}^d$ are respectively the particle's position and momentum, $\boldsymbol{x}_k \in \mathbb{R}^p$, $\boldsymbol{p}_k \in \mathbb{R}^p$ and $\omega_k \in \mathbb{R}^+$ $(k = 1, \ldots, N)$ are respectively the position, momentum and frequency of the kth bath oscillator (with unit mass), $\boldsymbol{f}(\boldsymbol{x}) := (f_1(\boldsymbol{x}), \ldots, f_r(\boldsymbol{x})) \in \mathbb{R}^r$ is a vector function of $\boldsymbol{x} := (x^{(1)}, \ldots, x^{(d)})$ and the $\boldsymbol{c}_k \in \mathbb{R}^{r \times p}$ (so $\boldsymbol{c}_k^* \in \mathbb{R}^{p \times r}$) are coupling matrices that specify the coupling strength between the system and the kth bath oscillator.

To derive an equation for the particle's position and momentum, we start by substituting the expression for $H(\hat{\boldsymbol{x}}, \hat{\boldsymbol{p}})$ into the Hamilton's equations to obtain:

$$\dot{\boldsymbol{x}} = \frac{\boldsymbol{p}}{m}, \qquad (150)$$

$$\dot{\boldsymbol{p}} = -\nabla_{\boldsymbol{x}} U(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x}) \sum_k \boldsymbol{c}_k \left(\boldsymbol{x}_k - \frac{\boldsymbol{c}_k^*}{\omega_k^2} \boldsymbol{f}(\boldsymbol{x}) \right), \qquad (151)$$

$$\dot{\boldsymbol{x}}_k = \dot{\boldsymbol{p}}_k, \quad k = 1, \ldots, N, \qquad (152)$$

$$\dot{\boldsymbol{p}}_k = -\omega_k^2 \boldsymbol{x}_k + \boldsymbol{c}_k^* \boldsymbol{f}(\boldsymbol{x}), \quad k = 1, \ldots, N, \qquad (153)$$

where $\boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}^{d \times r}$ denotes the Jacobian matrix $\left(\frac{\partial f_i}{\partial x^{(j)}} \right)_{ij}$.

Next, we eliminate the bath variables $\boldsymbol{x}_k, \boldsymbol{p}_k, k = 1, \ldots, N$, from the system's dynamics. Solving for $\boldsymbol{x}_k(t)$ in terms of $\boldsymbol{x}(t)$:

$$\boldsymbol{x}_k(t) = \boldsymbol{x}_k(0) \cos(\omega_k t) + \frac{\boldsymbol{p}_k(0)}{\omega_k} \sin(\omega_k t) + \frac{\boldsymbol{c}_k^*}{\omega_k} \int_0^t \sin(\omega_k(t - s)) \boldsymbol{f}(\boldsymbol{x}(s)) ds. \qquad (154)$$

Substituting this into (151), we obtain:

$$\dot{\boldsymbol{p}}(t) = -\nabla_{\boldsymbol{x}} U(\boldsymbol{x}(t)) + \boldsymbol{g}(\boldsymbol{x}(t)) \sum_k \frac{\boldsymbol{c}_k \boldsymbol{c}_k^*}{\omega_k^2} \left(\int_0^t \omega_k \sin(\omega_k(t - s)) \boldsymbol{f}(\boldsymbol{x}(s)) ds - \boldsymbol{f}(\boldsymbol{x}(t)) \right)$$

$$+ \boldsymbol{g}(\boldsymbol{x}(t)) \boldsymbol{F}(t), \qquad (155)$$

where

$$\boldsymbol{F}(t) = \sum_k \boldsymbol{c}_k \left(\boldsymbol{x}_k(0) \cos(\boldsymbol{w}_k t) + \frac{\boldsymbol{p}_k(0)}{\omega_k} \sin(\omega_k t) \right). \qquad (156)$$

In the integral term above, we integrate by parts to obtain:

$$\int_0^t \omega_k \sin(\omega_k(t-s)) f(x(s)) ds = f(x(t)) - \cos(\omega_k t) f(x(0))$$

$$- \int_0^t \cos(\omega_k(t-s)) g^*(x(s)) \dot{x}(s) ds. \quad (157)$$

Using this, the equation for $p(t)$ becomes the generalized Langevin equation (GLE):

$$\dot{p}(t) = -\nabla_x U(x(t)) - g(x(t)) \int_0^t \kappa(t-s) g^*(x(s)) \dot{x}(s) ds + g(x(t)) \xi(t), \quad (158)$$

where

$$\kappa(t) = \sum_k \frac{c_k c_k^*}{\omega_k^2} \cos(\omega_k t) \in \mathbb{R}^{r \times r} \quad (159)$$

and

$$\xi(t) = F(t) - \kappa(t) f(x(0)) = \sum_k c_k \left(\left(x_k(0) - \frac{c_k^*}{\omega_k^2} f(x(0)) \right) \cos(\omega_k t) + \frac{p_k(0)}{\omega_k} \sin(\omega_k t) \right). \quad (160)$$

Note that $\xi(t) \in \mathbb{R}^r$ is expressed in terms of the initial values of the variables $x_k'(0) := x_k(0) - \frac{c_k^*}{\omega_k^2} f(x(0)) \in \mathbb{R}^p$ and $p_k(0) \in \mathbb{R}^p$. If all these initial values are known, then $\xi(t)$ is a deterministic force. However, one rarely has a complete information about these initial values and this is where the introduction of randomness can help to simplify the model. In view of this, we assume that the variables $x_k'(0)$ and $p_k(0)$ are random and are distributed according to a Gibbs measure, with the density:

$$\rho((x_k, p_k) \mid x(0) = x) = Z^{-1} \exp\left(-\beta \left(\sum_{k=1}^N \frac{|p_k|^2}{2} + \frac{1}{2} \omega_k^2 \left| x_k - \frac{c_k^*}{\omega_k^2} f(x) \right|^2 \right) \right), \quad (161)$$

where $\beta = 1/(k_B T)$ and Z is the partition function. Taking the averages of the bath variables with respect to the above density:

$$E_\rho \left[x_k'(0) \mid x(0) = x \right] = 0, \quad E_\rho[p_k(0) \mid x(0) = x] = 0, \quad (162)$$

$$E_\rho[x_k'(0)(x_k'(0))^* \mid x(0) = x] = \frac{k_B T}{\omega_k^2} I, \quad E_\rho[(p_k(0)(p_k(0))^* \mid x(0) = x] = k_B T I, \quad (163)$$

where E_ρ denotes mathematical expectation with respect to ρ and $I \in \mathbb{R}^{p \times p}$ is identity matrix.

Note that $\xi(t)$ is a stationary Gaussian process, if it is conditionally averaged with respect to ρ [32]. It follows from this distribution of the bath variables that we have the *fluctuation–dissipation relation*:

$$E_\rho[\xi(t)] = 0, \quad E_\rho[\xi(t)\xi(s)^*] = k_B T \kappa(t-s), \quad (164)$$

where $\kappa(t-s)$ is the memory kernel whose formula is given in (159). Later, we will generalize the resulting covariance of the process $\xi(t)$ to an integral expression. We remark that the memory function $\kappa(t)$ and the "color" of the noise $\xi(t)$ are determined by the bath spectrum and the system-bath coupling.

Now we pass to the continuum limit by replacing the sum over k in $\boldsymbol{\kappa}(t)$ by an integral $\int_{\mathbb{R}^+} d\omega n(\omega)$, where $n(\omega)$ is a density of states. Then, if the $\boldsymbol{c}_k$ are replaced by $\boldsymbol{c}(\omega) \in \mathbb{R}^{r \times p}$, the memory function $\boldsymbol{\kappa}(t)$ becomes the function:

$$\boldsymbol{\kappa}(t) = \int_{\mathbb{R}^+} d\omega n(\omega) \frac{\boldsymbol{c}(\omega)\boldsymbol{c}(\omega)^*}{\omega^2} \cos(\omega t), \tag{165}$$

where $\hat{\boldsymbol{\kappa}}_c(\omega) := n(\omega)\boldsymbol{c}(\omega)\boldsymbol{c}(\omega)^*/\omega^2 \in L^1(\mathbb{R}^+)$. Repeating the same procedure for the noise process and also replacing the $\boldsymbol{x}_k'(0)$ and $\boldsymbol{p}_k(0)$ by $\boldsymbol{x}'(\omega)$ and $\boldsymbol{p}(\omega)$ respectively, $\boldsymbol{\xi}(t)$ becomes:

$$\boldsymbol{\xi}(t) = \int_{\mathbb{R}^+} d\omega n(\omega)\boldsymbol{c}(\omega) \left(\boldsymbol{x}'(\omega)\cos(\omega t) + \frac{\boldsymbol{p}(\omega)}{\omega}\sin(\omega t) \right). \tag{166}$$

The choice of the $n(\omega)$ and $\boldsymbol{c}(\omega)$ specifies the memory function and therefore (by the fluctuation–dissipation relation) the statistical properties of the noise process. We write $\boldsymbol{\kappa}(t)$ as an inverse Fourier transform of a measure:

$$\boldsymbol{\kappa}(t) = \frac{1}{2\pi} \int_{\mathbb{R}} S(\omega)e^{i\omega t}d\omega, \tag{167}$$

where the measure is absolutely continuous with respect to the Lebesgue measure, with the density $S(\omega) = \pi\hat{\boldsymbol{\kappa}}_c(\omega) \geq 0$. The density $S(\omega)$ is known as the *spectral density* of the bath.

In the following examples, we take $n(\omega) = 2\omega^2/\pi$ (Debye-type spectrum for phonon bath).

Example 1 If we choose $\boldsymbol{c}(\omega) \in \mathbb{R}^{d \times d}$ to be the identity matrix $\boldsymbol{I}$ multiplied by a scalar constant that is independent of ω, then $\boldsymbol{\kappa}(t)$ is proportional to $\delta(t)\boldsymbol{I}$. This leads to a Langevin equation driven by white noise, in which the damping term is instantaneous. In this case, we have the SDE system for $(\boldsymbol{x}_t, \boldsymbol{v}_t) \in \mathbb{R}^{d \times d}$:

$$d\boldsymbol{x}_t = \boldsymbol{v}_t dt, \tag{168}$$

$$m d\boldsymbol{v}_t = -\nabla_{\boldsymbol{x}} U(\boldsymbol{x}_t)dt - \boldsymbol{g}(\boldsymbol{x}_t)\boldsymbol{g}^*(\boldsymbol{x}_t)\boldsymbol{v}_t dt + \boldsymbol{g}(\boldsymbol{x}_t)\boldsymbol{\xi}_t dt, \tag{169}$$

where $\boldsymbol{\xi}_t$ is a white noise.

Example 2 If we choose $\boldsymbol{c}(\omega) \in \mathbb{R}^{d \times d}$ to be the diagonal matrix with the kth entry

$$\frac{\alpha_k}{\sqrt{\alpha_k^2 + \omega^2}}, \tag{170}$$

where the $\alpha_k > 0$, then we have:

$$\boldsymbol{\kappa}(t) = \boldsymbol{A}e^{-\boldsymbol{A}|t|}, \tag{171}$$

where $\boldsymbol{A}$ is the constant diagonal matrix with the kth entry equal α_k. This gives SIDE (15). On the other hand, choosing $\boldsymbol{c}(\omega)$ to be the diagonal matrix with the kth entry

$$\left(\frac{\omega_{kk}}{\tau_{kk}} \right)^2 \frac{1}{\sqrt{\omega^2(\omega_{kk}^2/\tau_{kk})^2 + (\omega^2 - (\omega_{kk}/\tau_{kk})^2)^2}} \tag{172}$$

allows us to obtain the covariance function of a harmonic noise process, where the ω_{kk} and τ_{kk} are the diagonal entries of the matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\tau}$ respectively. In the general case where $\boldsymbol{\kappa}(t)$ is of the form (3), one may take $\boldsymbol{M}_1 = \boldsymbol{I}, \boldsymbol{\Gamma}_1$ to be positive definite (so that the Lyapunov equation gives $\boldsymbol{\Gamma}_1 = \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_1^*/2$) and choose

$$\boldsymbol{c}(\omega) = \frac{1}{\sqrt{2}} \boldsymbol{C}_1(\boldsymbol{\Gamma}_1^2 + \omega^2 \boldsymbol{I})^{-1/2} \boldsymbol{\Sigma}_1. \tag{173}$$

Conclusion The goal of this paper is to study effective dynamics of systems which can be modeled by equations of the form (11). As argued above, Hamiltonian systems describing particles interacting with heat baths can be modeled by equations of the form (11).

Appendix B: Outline of the Proof of Theorem 1

We provide minimal outline of the proof of Theorem 1 in the following.

Appendix B.1: Derivation of Limiting SDE

Denoting $\boldsymbol{W}_t^{(k)} = \boldsymbol{W}_t$, we start by rewriting SDE (31) as

$$\boldsymbol{v}_t^m dt = -m\boldsymbol{\gamma}^{-1}(\boldsymbol{x}_t^m)d\boldsymbol{v}_t^m + \boldsymbol{\gamma}^{-1}(\boldsymbol{x}_t^m)\boldsymbol{F}(\boldsymbol{x}_t^m)dt + \boldsymbol{\gamma}^{-1}(\boldsymbol{x}_t^m)\boldsymbol{\sigma}(\boldsymbol{x}_t^m)d\boldsymbol{W}_t. \tag{174}$$

The integral form of the above is given by

$$\boldsymbol{x}_t^m = \boldsymbol{x} - m\int_0^t \boldsymbol{\gamma}^{-1}(\boldsymbol{x}_s^m)d\boldsymbol{v}_s^m + \int_0^t \boldsymbol{\gamma}^{-1}(\boldsymbol{x}_s^m)\boldsymbol{F}(\boldsymbol{x}_s^m)ds + \int_0^t \boldsymbol{\gamma}^{-1}(\boldsymbol{x}_s^m)\boldsymbol{\sigma}(\boldsymbol{x}_s^m)d\boldsymbol{W}_s. \tag{175}$$

We are interested in the limit as $m \to 0$ of the process $\boldsymbol{x}_t^m$. As $m \to 0$, we expect the sum of the second and third integral terms in the right hand side above to converge to $\int_0^t \boldsymbol{\gamma}^{-1}(\boldsymbol{X}_s)\boldsymbol{F}(\boldsymbol{X}_s)ds + \int_0^t \boldsymbol{\gamma}^{-1}(\boldsymbol{X}_s)\boldsymbol{\sigma}(\boldsymbol{X}_s)d\boldsymbol{W}_s$.

To examine the asymptotics of the first integral term when m becomes small, we integrate by parts to write its ith component as:

$$\int_0^t (\gamma^{-1})_{ij}(\boldsymbol{x}_s^m)d(mv_s^m)_j = (\gamma^{-1})_{ij}(\boldsymbol{x}_t^m)m(v_t^m)_j - (\gamma^{-1})_{ij}(\boldsymbol{x})m(v^m)_j$$
$$- \int_0^t \frac{\partial}{\partial x_l^m}[(\gamma^{-1})_{ij}(\boldsymbol{x}_s^m)]m(v_s^m)_j(v_s^m)_l ds. \tag{176}$$

Note that the product $m(v_s^m)_j(v_s^m)_l$ is the (j, l)-entry of the matrix $mv_s^m(v_s^m)^*$.

We now examine the asymptotic behavior of the above expression in the limit as $m \to 0$. Following [8], we express the matrix $mv_s^m(v_s^m)^*$, $s \in [0, t]$, as a solution to an equation by applying Itô's formula to the matrix $(mv_s^m)(m(v_s^m)^*)$. This leads to:

$$d[(mv_s^m)(m(v_s^m)^*)] = -[\boldsymbol{\gamma}(\boldsymbol{x}_s^m)(mv_s^m(v_s^m)^*ds) + (mv_s^m(v_s^m)^*ds)\boldsymbol{\gamma}^*(\boldsymbol{x}_s^m)]$$
$$+ \boldsymbol{\sigma}(\boldsymbol{x}_s^m)\boldsymbol{\sigma}^*(\boldsymbol{x}_s^m)ds + d\boldsymbol{U}_s^m + d(\boldsymbol{U}_s^m)^*, \tag{177}$$

where

$$d\boldsymbol{U}_s^m = (\boldsymbol{F}(\boldsymbol{x}_s^m)ds + \boldsymbol{\sigma}(\boldsymbol{x}_s^m)d\boldsymbol{W}_s)m(v_s^m)^*, \tag{178}$$
$$d(\boldsymbol{U}_s^m)^* = mv_s^m(\boldsymbol{F}^*(\boldsymbol{x}_s^m)ds + d\boldsymbol{W}_s^*\boldsymbol{\sigma}^*(\boldsymbol{x}_s^m)). \tag{179}$$

Denoting $mv_s^m(v_s^m)^*ds$ by $\boldsymbol{V}$, $-\boldsymbol{\gamma}(\boldsymbol{x}_s^m)$ by $\boldsymbol{Q}$ and letting

$$\boldsymbol{C} := d[(mv_s^m)(m(v_s^m)^*)] - \boldsymbol{\sigma}(\boldsymbol{x}_s^m)\boldsymbol{\sigma}^*(\boldsymbol{x}_s^m)ds - d\boldsymbol{U}_s^m - d(\boldsymbol{U}_s^m)^*, \tag{180}$$

we can write the above equation as the following Lyapunov equation

$$\boldsymbol{Q}\boldsymbol{V} + \boldsymbol{V}\boldsymbol{Q}^* = \boldsymbol{C}. \tag{181}$$

By Assumption 2, all real parts of the eigenvalues of Q are negative, thus the Lyapunov equation has a unique solution given by:

$$V = -\int_0^\infty e^{Qy} C e^{Q^* y} dy. \tag{182}$$

Writing this out explicitly, we obtain:

$$m v_s^m (v_s^m)^* ds = -\int_0^\infty e^{-\gamma(x_s^m)y} d[(m v_s^m)(m(v_s^m)^*)] e^{-\gamma^*(x_s^m)y} dy$$

$$+ \int_0^\infty e^{-\gamma(x_s^m)y} (\sigma(x_s^m)\sigma^*(x_s^m)ds) e^{-\gamma^*(x_s^m)y} dy$$

$$+ \int_0^\infty e^{-\gamma(x_s^m)y} (dU_s^m + d(U_s^m)^*) e^{-\gamma^*(x_s^m)y} dy. \tag{183}$$

Based on the prior result in [8], we expect that only the second term on the right hand side has a nonzero limit as $m \to 0$. The other terms are expected to vanish as $m \to 0$. Thus, in the limit $m \to 0$, we expect that $m v_t^m (v_t^m)^*$ converges to the solution, J, of the Lyapunov equation:

$$J \gamma^* + \gamma J = \sigma \sigma^*, \tag{184}$$

given in the statement of Theorem 1 (see Eq. (35)).

Appendix B.2: Moment Estimates

To justify the above convergence arguments, we provide estimate on the pth moment of the momentum process $p_t^m := m v_t^m$, in the limit $m \to 0$.

Proposition 1 *Suppose that Assumption 1–4 hold. For all $p \geq 1$, $T > 0$, there exists a positive random variable m_1 such that:*

$$\mathbb{E}\left[\sup_{t \in [0,T]} |p_t^m|^p; m \leq m_1\right] \to 0 \tag{185}$$

as $m \to 0$.

Proof For $t \in [0, T]$, $m > 0$, the process p_t^m satisfies the SDE:

$$dp_t^m = -\frac{\gamma(x_t^m)}{m} p_t^m dt + F(x_t^m)dt + \sigma(x_t^m)dW_t. \tag{186}$$

Let $\tau \in [0, t]$ and rewrite the above equation as:

$$dp_t^m = -\frac{\gamma(x_\tau^m)}{m} p_t^m dt + \frac{1}{m} \left(\gamma(x_\tau^m) - \gamma(x_t^m)\right) p_t^m dt + F(x_t^m)dt + \sigma(x_t^m)dW_t, \tag{187}$$

which admits the following solution representation:

$$p_t^m = e^{-\frac{\gamma(x_\tau^m)}{m}t} p_0^m + \int_0^t e^{-\frac{\gamma(x_\tau^m)}{m}(t-s)} F(x_s^m)ds + \int_0^t e^{-\frac{\gamma(x_\tau^m)}{m}(t-s)} \sigma(x_s^m)dW_s$$

$$+ \frac{1}{m} \int_0^t e^{-\frac{\gamma(x_\tau^m)}{m}(t-s)} (\gamma(x_\tau^m) - \gamma(x_s^m)) p_s^m ds. \tag{188}$$

We set $\tau = t$ in the above representation to obtain:

$$
\boldsymbol{p}_t^m = e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}t} \boldsymbol{p}_0^m + \int_0^t e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}(t-s)} \boldsymbol{F}(\boldsymbol{x}_s^m) ds + \int_0^t e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}(t-s)} \boldsymbol{\sigma}(\boldsymbol{x}_s^m) d\boldsymbol{W}_s
$$
$$
+ \frac{1}{m} \int_0^t e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}(t-s)} (\boldsymbol{\gamma}(\boldsymbol{x}_t^m) - \boldsymbol{\gamma}(\boldsymbol{x}_s^m)) \boldsymbol{p}_s^m ds. \tag{189}
$$

Therefore,

$$
|\boldsymbol{p}_t^m| \leq \left| e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}t} \boldsymbol{p}_0^m \right| + \int_0^t \left\| e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}(t-s)} \right\| \cdot |\boldsymbol{F}(\boldsymbol{x}_s^m)| ds + \left| \int_0^t e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}(t-s)} \boldsymbol{\sigma}(\boldsymbol{x}_s^m) d\boldsymbol{W}_s \right|
$$
$$
+ \sup_{s\in[0,T]} |\boldsymbol{p}_s^m| \cdot \frac{1}{m} \int_0^t \left\| e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}(t-s)} \right\| \cdot \|(\boldsymbol{\gamma}(\boldsymbol{x}_t^m) - \boldsymbol{\gamma}(\boldsymbol{x}_s^m))\| ds. \tag{190}
$$

By assumption on the boundedness and spectrum of $\boldsymbol{\gamma}(\boldsymbol{x}_\tau) \in \mathbb{R}^{n\times n}$ (see Assumption 1-2), there exist positive constants $\kappa > 0$ and $C > 0$ such that

$$
\left\| e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_\tau^m)}{m}s} \right\| \leq Ce^{-\frac{\kappa}{m}s}, \tag{191}
$$

for all $s, \tau \in [0, T]$, $m > 0$. Indeed, applying the formula (A.2.4) in [50], one has:

$$
\left\| e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_\tau^m)}{m}s} \right\| \leq e^{-\frac{\Lambda}{m}s} \left(1 + 2\|\boldsymbol{\gamma}\| \sum_{k=1}^{n-1} \frac{1}{k!} \left(\frac{2s\|\boldsymbol{\gamma}\|}{m} \right)^k \right), \tag{192}
$$

where $\Lambda := \min_k Re(\lambda_k) > 0$ and the λ_k are the eigenvalues of $\boldsymbol{\gamma}$. Therefore, there exists a constant $C > 0$ such that $\left\| e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_\tau^m)}{m}s} \right\| \leq Ce^{-\frac{\Lambda}{2m}s}$.

Using this, we obtain the following $\mathbb{P}$-a.s. estimate:

$$
\sup_{t\in[0,T]} |\boldsymbol{p}_t^m| \leq C \left| \boldsymbol{p}_0^m \right| + C \sup_{t\in[0,T]} \int_0^t e^{-\frac{\kappa}{m}(t-s)} |\boldsymbol{F}(\boldsymbol{x}_s^m)| ds
$$
$$
+ \sup_{t\in[0,T]} \left| \int_0^t e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}(t-s)} \boldsymbol{\sigma}(\boldsymbol{x}_s^m) d\boldsymbol{W}_s \right|
$$
$$
+ \sup_{t\in[0,T]} |\boldsymbol{p}_t^m| \left(\sup_{t\in[0,T]} \frac{\tilde{C}}{m} \int_0^t e^{-\frac{\kappa}{m}(t-s)} \|\boldsymbol{\gamma}(\boldsymbol{x}_t^m) - \boldsymbol{\gamma}(\boldsymbol{x}_s^m)\| ds \right), \tag{193}
$$

where $\tilde{C} > 0$ is a constant.

Next, the key observation on (193) is that the term in parentheses in (193) can be made small by choosing a sufficiently small m. More precisely, using an adapted version of Lemma A.2.4 in [50], there exists a (generally random) $m_1 > 0$ such that for $m \leq m_1$,

$$
\sup_{t\in[0,T]} |\boldsymbol{p}_t^m| \leq C \left| \boldsymbol{p}_0^m \right| + C \sup_{t\in[0,T]} \int_0^t e^{-\frac{\kappa}{m}(t-s)} |\boldsymbol{F}(\boldsymbol{x}_s^m)| ds
$$
$$
+ \sup_{t\in[0,T]} \left| \int_0^t e^{-\frac{\boldsymbol{\gamma}(\boldsymbol{x}_t^m)}{m}(t-s)} \boldsymbol{\sigma}(\boldsymbol{x}_s^m) d\boldsymbol{W}_s \right| + \frac{1}{2} \sup_{t\in[0,T]} |\boldsymbol{p}_t^m|, \tag{194}
$$

and so:

$$
\sup_{t\in[0,T]} |\boldsymbol{p}_t^m| \leq 2C \left| \boldsymbol{p}_0^m \right| + 2C \sup_{t\in[0,T]} \int_0^t e^{-\frac{\kappa}{m}(t-s)} |\boldsymbol{F}(\boldsymbol{x}_s^m)| ds
$$

$$+ 2 \sup_{t \in [0,T]} \left| \int_0^t e^{-\frac{\gamma(x_t^m)}{m}(t-s)} \sigma(x_s^m) dW_s \right| \tag{195}$$

$$\leq 2Cm \left| v_0^m \right| + 2C \frac{m}{\kappa} \sup_{u \in \mathbb{R}^n} |F(u)| + 2 \sup_{t \in [0,T]} \left| \int_0^t e^{-\frac{\gamma(x_t^m)}{m}(t-s)} \sigma(x_s^m) dW_s \right|. \tag{196}$$

Therefore, for $p \geq 1$,

$$\mathbb{E} \left[\sup_{t \in [0,T]} |p_t^m|^p; m \leq m_1 \right] \leq C_1(p) \mathbb{E} \left[\left| m v_0^m \right|^p; m \leq m_1 \right] + C_2(p) \mathbb{E} \left[m^p; m \leq m_1 \right]$$

$$+ C_3(p) \mathbb{E} \left[\sup_{t \in [0,T]} \left| \int_0^t e^{-\frac{\gamma(x_t^m)}{m}(t-s)} \sigma(x_s^m) dW_s \right|^p; m \leq m_1 \right], \tag{197}$$

where $C_1(p)$, $C_2(p)$ and $C_3(p)$ are some positive constants.

We estimate the last term in the above, using in particular the Burkholder-Davis-Gundy inequality [33]:

$$\mathbb{E} \left[\sup_{t \in [0,T]} \left| \int_0^t e^{-\frac{\gamma(x_t^m)}{m}(t-s)} \sigma(x_s^m) dW_s \right|^p; m \leq m_1 \right]$$

$$\leq \mathbb{E} \left[\sup_{t \in [0,T]} \left| \int_0^t e^{-\frac{\gamma(x_t^m)}{m}(t-s)} \sigma(x_s^m) dW_s \right|^p \right] \tag{198}$$

$$\leq C(p,n) \mathbb{E} \left(\int_0^T \| e^{-\frac{\gamma(x_t^m)}{m}(t-s)} \sigma(x_s^m) \|_F^2 ds \right)^{p/2} \tag{199}$$

$$\leq \tilde{C}(p,n) \mathbb{E} \left(\int_0^T e^{-2\kappa(t-s)/m} ds \right)^{p/2} \tag{200}$$

$$\leq \frac{\tilde{C}(p,n)}{\kappa} m^{p/2}, \tag{201}$$

where $C(p,n)$, $\tilde{C}(p,n)$ are positive constants dependent on p and n, and $\| \cdot \|_F$ denotes Frobenius norm. Using this estimate, (197) and Assumption 3, we see that $\mathbb{E} \left[\sup_{t \in [0,T]} |p_t^m|^p; m \leq m_1 \right] \to 0$ as $m \to 0$.   □

We also need the following estimate on a class of integrals with respect to products of the components of the momentum process $p_t^m = m v_t^m$. The estimate is a straightforward modification of the one given in Proposition 2.3 in [37].

Proposition 2 *Suppose that Assumption 1-4 hold. Let $h : \mathbb{R}^n \to \mathbb{R}$ be a C_b^1 function (i.e. continuously differentiable and bounded function) on $[0, T]$, with bounded first derivative $\nabla_x h(x)$ for every $x \in \mathbb{R}^n$. Then for any $p \geq 1$, $T > 0$, $i, j = 1, \ldots, n$,*

$$\mathbb{E} \left[\sup_{t \in [0,T]} \left| \int_0^t h(x_s^m) d((p_s^m)_i (p_s^m)_j) \right|^p; m \leq m_1 \right] \to 0 \tag{202}$$

as $m \to 0$, where the m_1 is from Proposition 1. Here i, j denote the components of the momentum process p_t^m in the standard basis for $\mathbb{R}^n$.

Using the above moment estimates and the proof techniques (the main tools are well known ordinary and stochastic integral inequalities as well as a Gronwall type argument) in [36,37], one obtain the convergence of x_t^m to X_t in the limit as $m \to 0$ in the following sense: for all finite $T > 0$, $p \geq 1$,

$$\mathbb{E}\left[\sup_{t \in [0,T]} |x_t^m - X_t|^p; m \leq m_1\right] \to 0, \tag{203}$$

as $m \to 0$, where the m_1 is from Proposition 1. This implies that for all finite $T > 0$, $\sup_{t \in [0,T]} |x_t^m - X_t| \to 0$ in probability, in the limit as $m \to 0$ (see Lemma 1 below).

Lemma 1 *Let $Y_m \geq 0$ ($m \geq 0$) be a family of random variables. Suppose that there exists a strictly positive random variable $m^*(\omega)$ such that*

$$\lim_{m \to 0} \mathbb{E}\left[Y_m^p; m^* \geq m\right] \to 0$$

for some $p \geq 1$. Then $Y_m \to 0$ in probability as $m \to 0$.

Proof Let g be a bounded continuous function on $\mathbb{R}_+$. We have

$$\mathbb{E}\left[g(Y_m)\right] = \mathbb{E}\left[g(Y_m); m^* \leq m\right] + \mathbb{E}\left[g(Y_m); m^* \geq m\right]. \tag{204}$$

The first term is bounded from above by $(\sup |g|)\mathbb{P}\left[m^* \leq m\right]$ which goes to zero as $m \to 0$, since $\cap_{m>0}\{m^* \leq m\} = \emptyset$. To estimate the second term, let $\epsilon > 0$. Choose $\delta > 0$ such that $|g(y)| < \epsilon$ whenever $y < \delta$. We have

$$\mathbb{E}\left[g(Y_m); m^* \geq m\right] \leq \mathbb{E}\left[g(Y_m); m^* \geq m, Y_m < \delta\right] + \mathbb{E}\left[g(Y_m); m^* \geq m, Y_m \geq \delta\right]. \tag{205}$$

The first term is bounded from above by ϵ. The function $\frac{g(y)}{y}$ on $\{y : y \geq \delta\}$ is bounded by some constant M, so the second term on the right-hand side of the above equation is bounded from above by $M\mathbb{E}\left[Y_m; m^* \geq m\right]$ which goes to zero as $m \to 0$ by assumption. The lemma is proven. $\square$

We end this appendix with a remark on the mode of convergence stated in Theorem 1.

Remark 16 Provided that for $p \geq 1$, there exists a constant $C > 0$ such that $\mathbb{E}\sup_{t \in [0,T]} |p_t^m|^p < C$ for all $m > 0$ (so that the family of random variables $(\sup_{t \in [0,T]} |p_t^m|^p)_{m>0}$ is uniformly integrable), one could, using Proposition 1, Lemma 1 and Theorem 13.7 in [51], obtain L^p-convergence of $\sup_{t \in [0,T]} |p_t^m|$ to zero and hence strengthen the convergence result stated in Theorem 1 to L^p-convergence. The uniform integrability condition is satisfied if $\mathbb{E}\sup_{t \in [0,T]} \|\Phi^m(t)\|^p$ is bounded uniformly in m, where $\Phi^m(t)$ is the fundamental matrix that solves the random initial value problem:

$$\frac{\partial}{\partial t}\Phi^m(t) = -\frac{\gamma(x_t^m)}{m}\Phi^m(t), \quad \Phi^m(0) = I, \quad t \in [0, T]. \tag{206}$$

However, it is not obvious how one could verify the latter condition from our assumptions on γ. Roughly speaking, one does not have a good control of $\sup_{t \in [0,T]} |p_t^m|^p$ outside of the set $\{m \leq m_1\}$. If γ was, in addition, symmetric (and so all the (real) eigenvalues of γ are bounded from below by a positive constant—c.f. [37]), the condition can be easily verified.

References

1. Majda, A.J., Timofeyev, I., Vanden Eijnden, E.: A mathematical framework for stochastic climate models. Commun. Pure Appl. Math. **54**(8), 891–974 (2001)
2. Givon, D., Kupferman, R., Stuart, A.: Extracting macroscopic dynamics: model problems and algorithms. Nonlinearity **17**(6), R55 (2004)
3. Pavliotis, G.A., Stuart, A.M.: Analysis of white noise limits for stochastic systems with two fast relaxation times. Multiscale Model. Simul. **4**(1), 1–35 (2005)
4. Pavliotis, G., Stuart, A.: Multiscale Methods, Texts in Applied Mathematics, vol. 53. Springer, New York (2008)
5. Nelson, E.: Dynamical Theories of Brownian Motion, vol. 2. Princeton University Press, Princeton (1967)
6. Franosch, T., Grimm, M., Belushkin, M., Mor, F.M., Foffi, G., Forró, L., Jeney, S.: Resonances arising from hydrodynamic memory in Brownian Motion. Nature **478**(7367), 85–88 (2011)
7. Gröblacher, S., Trubarov, A., Prigge, N., Cole, G., Aspelmeyer, M., Eisert, J.: Observation of non-Markovian micromechanical Brownian motion. Nat. Commun. **6**, 7606 (2015). https://doi.org/10.1038/ncomms8606
8. Hottovy, S., McDaniel, A., Volpe, G., Wehr, J.: The Smoluchowski–Kramers limit of stochastic differential equations with arbitrary state-dependent friction. Commun. Math. Phys. **336**(3), 1259–1283 (2015)
9. Trentelman, H.L., Stoorvogel, A.A., Hautus, M.: Control Theory for Linear Systems. Communications and Control Engineering Series. Springer, Berlin (2002)
10. Willems, J., Van Schuppen, J.: Stochastic systems and the problem of state space realization. In: Geometrical Methods for the Theory of Linear Systems: Proceedings of a NATO Advanced Study Institute and AMS Summer Seminar in Applied Mathematics held at Harvard University, Cambridge, June 18–29, 1979, vol. 62, p. 283. Springer, Berlin (1980)
11. Mori, H.: Transport, collective motion, and Brownian motion. Prog. Theor. Phys. **33**(3), 423–455 (1965)
12. Kubo, R.: The fluctuation–dissipation theorem. Rep. Prog. Phys. **29**(1), 255 (1966)
13. Toda, M., Kubo, R., Saito, N., Hashitsume, N.: Statistical Physics II: Nonequilibrium Statistical Mechanics. Springer Series in Solid-State Sciences. Springer, Berlin (2012)
14. Goychuk, I.: Viscoelastic subdiffusion: Generalized Langevin equation approach. Adv. Chem. Phys. **150**, 187 (2012)
15. Van Kampen, N.: Remarks on non-Markov processes. Braz. J. Phys. **28**(2), 90–96 (1998)
16. Łuczka, J.: Non-Markovian stochastic processes: colored noise. Chaos: an interdisciplinary. J. Nonlinear Sci. **15**(2), 026107 (2005)
17. Samorodnitsky, G., Taqqu, M.: Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Stochastic Modeling Series. Taylor & Francis, Boca Raton (1994)
18. Stella, L., Lorenz, C.D., Kantorovich, L.: Generalized Langevin equation: an efficient approach to nonequilibrium molecular dynamics of open systems. Phys. Rev. B **89**, 134303 (2014). https://doi.org/10.1103/PhysRevB.89.134303
19. McKinley, S.A., Yao, L., Forest, M.G.: Transient anomalous diffusion of tracer particles in soft matter. J. Rheol. (1978-present) **53**(6), 1487–1506 (2009)
20. Adelman, S., Doll, J.: Generalized Langevin equation approach for atom/solid-surface scattering: general formulation for classical scattering off harmonic solids. J. Chem. Phys. **64**(6), 2375–2388 (1976)
21. Ottobre, M., Pavliotis, G.A.: Asymptotic analysis for the generalized Langevin equation. Nonlinearity **24**, 1629–1653 (2011). https://doi.org/10.1088/0951-7715/24/5/013
22. Kalman, R.E.: A new approach to linear filtering and prediction problems. J. Basic Eng. **82**(1), 35–45 (1960)
23. Bellman, R.: Introduction to Matrix Analysis, vol. 19. SIAM, Philadelphia (1997)
24. Lindquist, A., Picci, G.: Realization theory for multivariate stationary Gaussian processes. SIAM J. Control Optim. **23**(6), 809–857 (1985)
25. Lindquist, A., Picci, G.: Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification. Series in Contemporary Mathematics. Springer, Berlin (2015)
26. Bao, J.D., Hänggi, P., Zhuo, Y.Z.: Non-Markovian Brownian dynamics and nonergodicity. Phys. Rev. E **72**(6), 061,107 (2005)
27. Pavliotis, G.: Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations. Texts in Applied Mathematics. Springer, New York (2014)
28. Hottovy, S., McDaniel, A., Wehr, J.: A small delay and correlation time limit of stochastic differential delay equations with state-dependent colored noise. Markov Proc. Relat. Fields **22**(3), 595–628 (2016)
29. Schimansky-Geier, L., Zülicke, C.: Harmonic noise: effect on bistable systems. Zeitschrift für Physik B Condensed Matter **79**(3), 451–460 (1990)

30. McDaniel, A., Duman, O., Volpe, G., Wehr, J.: An SDE approximation for stochastic differential delay equations with colored state-dependent noise (2014). arXiv:1406.7287

31. Hänggi, P., Jung, P., Zerbe, C., Moss, F.: Can colored noise improve stochastic resonance? J. Stat. Phys. **70**(1), 25–47 (1993)

32. Zwanzig, R.: Nonequilibrium Statistical Mechanics. Oxford University Press, Oxford (2001)

33. Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus, vol. 113. Springer, Berlin (2012)

34. Hottovy, S., Volpe, G., Wehr, J.: Noise-induced drift in stochastic differential equations with arbitrary friction and diffusion in the Smoluchowski–Kramers limit. J. Stat. Phys. **146**(4), 762–773 (2012). https://doi.org/10.1007/s10955-012-0418-9

35. Herzog, D.P., Hottovy, S., Volpe, G.: The small mass limit for Langevin dynamics with unbounded coefficients and positive friction. J. Stat. Phys. **163**(3), 659–673 (2016). https://doi.org/10.1007/s10955-016-1498-8

36. Birrell, J., Hottovy, S., Volpe, G., Wehr, J.: Small mass limit of a Langevin equation on a manifold. Annales Henri Poincaré **18**, 707–755 (2017)

37. Birrell, J., Wehr, J.: Homogenization of Dissipative, Noisy, Hamiltonian Dynamics. Stochastic Processes and Their Applications. Elsevier, Amsterdam (2017)

38. Volpe, G., Wehr, J.: Effective drifts in dynamical systems with multiplicative noise: a review of recent progress. Rep. Prog. Phys. **79**(5), 053901 (2016)

39. Horn, R., Johnson, C.: Topics in Matrix Analysis. Cambridge University Press, Cambridge (1994)

40. Winkelbauer, A.: Moments and absolute moments of the normal distribution (2012). arXiv:1209.4340

41. Eckmann, J.P., Pillet, C.A., Rey-Bellet, L.: Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. Commun. Math. Phys. **201**(3), 657–697 (1999)

42. Hottovy, S., Volpe, G., Wehr, J.: Thermophoresis of Brownian particles driven by colored noise. EPL **99**, 60002 (2012)

43. Piazza, R., Parola, A.: Thermophoresis in colloidal suspensions. J. Phys. **20**(15), 153102 (2008)

44. Kou, S.C.: Stochastic modeling in nanoscale biophysics: subdiffusion within proteins. Ann. Appl. Stat. **2**(2), 501–535 (2008)

45. Ford, G., Kac, M., Mazur, P.: Statistical mechanics of assemblies of coupled oscillators. J. Math. Phys. **6**(4), 504–515 (1965)

46. Hänggi, P.: Generalized Langevin equations: A useful tool for the perplexed modeller of nonequilibrium fluctuations? Stochastic Dynamics, pp. 15–22. Springer, Berlin (1997)

47. Zwanzig, R.: Nonlinear generalized Langevin equations. J. Stat. Phys. **9**(3), 215–220 (1973). https://doi.org/10.1007/BF01008729

48. Ariel, G., Vanden-Eijnden, E.: A strong limit theorem in the Kac–Zwanzig model. Nonlinearity **22**(1), 145 (2008)

49. Rey-Bellet, L.: Open classical systems. Open Quantum Systems II, pp. 41–78. Springer, Berlin (2006)

50. Kabanov, Y., Pergamenshchikov, S.: Two-Scale Stochastic Systems: Asymptotic Analysis and Control. Stochastic Modelling and Applied Probability. Springer, Berlin (2013)

51. Williams, D.: Probability with Martingales. Cambridge University Press, Cambridge (1991)