Anomalous Thermodynamics in Homogenized Generalized Langevin Systems

Soon Hoe Lim SOON.HOE.LIM@SU.SE
Nordita
Royal Institute of Technology and Stockholm University
Stockholm 106 91, Sweden

Abstract

We study functionals, such as heat and work, along trajectories of a class of multi-dimensional generalized Langevin systems in various limiting situations that correspond to different level of homogenization. These are the situations where one or more of the inertial time scale(s), the memory time scale(s) and the noise correlation time scale(s) of the systems are taken to zero. We find that, unless one restricts to special situations evoking symmetry, it is generally not possible to express the effective evolution of these functionals solely in terms of trajectory of the homogenized process describing the system dynamics via the widely adopted Stratonovich convention. In fact, an anomalous term is often needed for a complete description, implying that convergence of these functionals needs more information than simply the limit of the dynamical process. We trace the origin of such impossibility to area anomaly, thereby linking symmetry breaking and area anomaly. This hold important consequences for many nonequilibrium systems that can be modeled by generalized Langevin equations. Our convergence results hold in a strong pathwise sense.

Keywords: Generalized Langevin Systems, Functionals Along Trajectories, Stochastic Thermodynamics, Homogenization, Area Anomaly, Nonequilibrium Systems

Contents

1. Introduction

We consider a class of non-Markovian Langevin equations, whose coefficients are possibly state-dependent, describing the dynamics of a particle moving in a force field and interacting with the environment. The evolution of the particle's position, $\boldsymbol{x}_t \in \mathbb{R}^d$, $t \geq 0$, is given by the solution to the following stochastic integro-differential equation (SIDE) (Lim et al., 2020):

$$m\ddot{\boldsymbol{x}}_t = \boldsymbol{F}(t, \boldsymbol{x}_t) - \boldsymbol{\gamma}_0(\boldsymbol{x}_t)\dot{\boldsymbol{x}}_t - \boldsymbol{g}(\boldsymbol{x}_t)\int_0^t \boldsymbol{\kappa}(t-s)\boldsymbol{h}(\boldsymbol{x}_s)\dot{\boldsymbol{x}}_s ds + \boldsymbol{\sigma}_0(\boldsymbol{x}_t)\boldsymbol{\eta}_t + \boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{\xi}_t, \tag{1}$$

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

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

In the SIDE (1), overdot denotes derivative with respect to time t, $m > 0$ is the mass of the particle, the matrix-valued functions $\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^{d\times q}$, $\boldsymbol{h} : \mathbb{R}^d \to \mathbb{R}^{q\times d}$, $\boldsymbol{\sigma} : \mathbb{R}^d \to \mathbb{R}^{d\times r}$, $\boldsymbol{\gamma}_0 : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $\boldsymbol{\sigma}_0 : \mathbb{R}^d \to \mathbb{R}^{d\times b}$ are the coefficients of the equation, and $\boldsymbol{F} : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}^d$ is a force field acting on the particle. Here d, q, r and b are, possibly distinct, positive integers. Here and throughout the paper, the superscript T denotes transposition of matrices or vectors and E denotes mathematical expectation. The SIDE (1) can be viewed as a Newton's equation of motion (i.e., $m\ddot{\boldsymbol{x}}_t = \boldsymbol{F}(t, \boldsymbol{x}_t)$) with additional forcing terms to be described in the following.

The second and third term on the right hand side of (1) represent the drag experienced by the particle. This drag is modeled by a sum of two deterministic damping terms of different nature. The second term, proportional to the particle's velocity, models instantaneous damping. On the other hand, the third term, involving an integral over the particle's past velocities with the kernel $\boldsymbol{\kappa}(t-s)$, describes non-instantaneous, distributed delayed, damping due to the back-action effects of the environment up to current time. The matrix-valued function $\boldsymbol{\kappa} : \mathbb{R} \to \mathbb{R}^{q\times q}$ is called a memory function and it decays sufficiently fast at infinities.

The forth and fifth term on the right hand side of (1) represent two stochastic forcings (noises) of different nature imparted to the particle. They are $\boldsymbol{\sigma}_0(\boldsymbol{x}_t)\boldsymbol{\eta}_t$, which is a Gaussian white noise, and $\boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{\xi}_t$, which is a Gaussian colored noise, both of which are possibly multiplicative. Here the process $\boldsymbol{\eta}_t$ represents a b-dimensional white noise, and $\boldsymbol{\xi}_t$ is a r-dimensional mean zero stationary Gaussian process with the covariance function $\boldsymbol{R}(t) = E[\boldsymbol{\xi}_t\boldsymbol{\xi}_0^T]$. The two noise processes are mutually independent. The initial conditions $\boldsymbol{x}$ and $\boldsymbol{v}$ are random variables independent of the noise process $\{(\boldsymbol{\eta}_t, \boldsymbol{\xi}_t) : \ t \geq 0\}$. Precise definition and assumptions, as well as physical motivation, for the memory function and the noise processes will be given in Section 2.

Therefore, (1) is a *generalized Langevin equation* (GLE), containing the Langevin-Kramers equation studied in (Hottovy et al., 2015) (by setting $\boldsymbol{h}$ and $\boldsymbol{\sigma}$ to zero) and the GLE studied in (Lim and Wehr, 2018) (by setting $\boldsymbol{\gamma}_0$ and $\boldsymbol{\sigma}_0$ to zero) as special cases. The most basic form of GLE, which is a special case of (1), was first introduced by Mori in (Mori, 1965) and subsequently used to model many systems in statistical and biological physics (Ermak and McCammon, 1978). The GLE has attracted increasing attentions in recent years, due to its successful application in modeling anomalously diffusing systems, active matter systems and many other nonequilibrium systems (Goychuk, 2012; Lysy et al., 2016; Gottwald et al., 2017; Sevilla, 2018).

We remark that GLEs of the form (1), despite being more general in the above sense, are still not the most general ones. Depending on modeling details (for instance, the form of the coupling among various degrees of freedom), one may need to add other forces such as a Basset force (to account for the effect of hydrodynamic backflow (Fodor et al., 2015)) in the GLEs, or consider GLEs for a set of reaction coordinates/gross variables instead, in which case the resulting GLEs may feature renormalization of bare potential fields, resulting in a potential of mean force (see Section II.B in (Hänggi et al., 1990) or the recent paper (Talkner and Hänggi, 2020) and the references therein). While it is important to keep in mind of these more general models, we will not study them in this paper.

One particular instance, of important relevance in statistical mechanics, that we will revisit often is when the coefficients and/or functions defining the GLE (1) are related in the following way.

Relation 1.1 *Fluctuation-dissipation relations.*

(a) $\boldsymbol{\sigma}_0\boldsymbol{\sigma}_0^T = \boldsymbol{\gamma}_0$ *(i.e. the fluctuation-dissipation relation of the first kind holds);*

(b) $\boldsymbol{\kappa}(t) = \boldsymbol{R}(t)$ *and* $\boldsymbol{g} = \boldsymbol{h}^T = \boldsymbol{\sigma}$ *(i.e. the fluctuation-dissipation relation of the second kind holds).*

It turns out that the GLE (1), with $\boldsymbol{\gamma}_0$ and $\boldsymbol{\sigma}_0$ zero and satisfying (b) in Relation 1.1, can be derived from a microscopic Hamiltonian model (Kac-Zwanzig or Caldeira-Leggett type) for a small system interacting with a heat bath, or via the Mori-Zwanzig projection approach. See, for instance, Appendix A in (Lim and Wehr, 2018) or (Hänggi, 1997; Zwanzig, 1973; Rey-Bellet, 2006; Leimkuhler and Sachs, 2019). In this case, there will be proportionality constants, containing the temperature of the heat bath as a parameter, in the fluctuation-dissipation relations. Since these constants could be absorbed into $\boldsymbol{g}$, $\boldsymbol{h}$ or $\boldsymbol{\sigma}$, we choose not to include them explicitly in Assumption 1.1. Lastly, we remark that the term $-\boldsymbol{\gamma}_0(\boldsymbol{x}_t)\boldsymbol{v}_t$ (when $\boldsymbol{\gamma}_0$ is non-zero) could be used to model forces of different nature acting on the particle, in particular when $\boldsymbol{\gamma}_0$ is not positive definite (and therefore cannot model a damping term) – see Example A.3. Throughout this paper, $\boldsymbol{\gamma}_0$ is either zero or non-zero, in which case it is either positive definite or not positive definite.

There are numerous studies focusing on asymptotic analysis and model reduction of GLEs, aiming to justify the use of low-dimensional phenomenological equations such as the Langevin-Kramers equations and the overdamped Langevin equations for modeling of statistical systems. See, for instance, (Ottobre and Pavliotis, 2011; Lim and Wehr, 2018; Nguyen, 2018). There are also many works studying asymptotics of functionals along trajectory of these phenomenological equations (Celani et al., 2012; Bo and Celani, 2017; Ge and Jin, 2018; Pan et al., 2018; Birrell, 2018). On the other hand, to our best knowledge works performing asymptotic analysis of functionals along trajectory of generalized Langevin systems, in particular for functionals appearing in stochastic thermodynamics of GLEs, are scarce.

In this paper we present a comprehensive multiple time scales analysis (homogenization) of these functionals, as well as of the GLE dynamics, in various limiting situations. The main goal is *to apply the multiscale analysis to investigate the issue of discretization choice for a class of stochastic integrals appearing in stochastic thermodynamics*. This issue concerns with justification (or not) of the widespread use of Stratonovich convention (midpoint discretization) for defining functionals, such as heat and work, along trajectories of these phenomenological models, used in deriving the law of energy balance in the energetics literature (Sekimoto, 2010; Seifert, 2012). From mathematical viewpoint, the Stratonovich choice of discretization guarantees the vector fields involved transform under a change of coordinates (Chetrite and Gawędzki, 2008) and is therefore suitable for formulation of coordinate-free SDEs on manifolds. However, this choice needs to be carefully justified at a more fundamental level, for instance by taking a GLE as starting point for analysis, in which case the functionals (stochastic integrals) along the phase-space trajectories are uniquely defined (i.e. their discretization is free of ambiguities). Performing homogenization on these functionals allows us to find out its limiting expression in the considered limit. This limiting expression is then compared to the functional defined along the trajectory of the limiting dynamics.

In our previous contribution in (Bo et al., 2019), we have shown that for systems in which noise correlation is shorter-lived than inertia (usually the case for microscopic colloids in water at room temperature) the correct discretization for these functionals is Stratonovich – this is the result obtained by performing a Markovian limit first and then the small mass limit. This result holds under the conditions that (i) the processes which generate the colored noise are equilibrium ones, and (ii) in the small mass limit the velocity degrees of freedom reach an equilibrium distribution with the local temperature (this holds when the fluctuation-dissipation relation is obeyed). For systems that violate these conditions, the interpretation of the (limiting) functionals is less immediately clear. The main motivation and contribution of this paper is, in fact, to investigate and identify the limiting behavior of these functionals beyond the aforementioned setting via a systematic multiscale analysis considering different hierarchies of the time scales involved. The results obtained in this paper not only recover our earlier results in (Bo et al., 2019), but also give new results and uncover interesting insights in more general settings. We emphasize that the present paper has a rather applied flavor, i.e., it focuses on applying the general homogenization theorems obtained in our earlier works (Lim et al., 2020; Lim and Wehr, 2018) to shed light on the above issue in the field of stochastic thermodynamics, rather than presenting novel mathematical techniques and proofs for homogenization. Moreover, the notion of stochastic Lévy areas is, for the first time, connected to thermodynamic quantities such as work done on physical systems.

This paper is organized as follows. In Section 2, we define the class of GLE models to be studied in this paper. We give three examples, of relevance in applications to study nonequilibrium systems, of these models in Appendix A. In Section 3, we motivate and introduce a class of functionals along trajectories of the GLE. In Section 4, we study homogenization for a class of SDE systems with state-dependent coefficients and their functionals. The convergence results are obtained in a strong pathwise sense. They follow from applications of the homogenization theorem in Appendix B. We discuss the mathematical implications of these results, in particular we link symmetry breaking and area anomaly. In Section 5, we illustrate and discuss this link in the context of a Brownian particle in a magnetic field to build some intuition on area anomaly before moving on to study the more general situations of GLEs. Section 6 contains the main contributions of the paper. There, building on the results in Section 4, we study homogenization for generalized Langevin dynamics as well as the functionals introduced in Section 3. We then discuss the conditions under which a Stratonovich functional is recovered for various limiting situations, as well as the consequences due to interplay between symmetry breaking and area anomaly. We conclude the paper in Section 7.

2. Generalized Langevin Equations (GLEs)

In this section we define our GLE models, following closely the notation in (Lim and Wehr, 2018). In the GLE (1), the memory function $\boldsymbol{\kappa} : \mathbb{R} \to \mathbb{R}^{q \times q}$ is taken to be *Bohl*, i.e. the matrix elements of $\boldsymbol{\kappa}(t)$ 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 (Trentelman et al., 2002). The noise process $\boldsymbol{\xi}_t$ is a r-dimensional mean zero stationary real-valued Gaussian vector process having a Bohl covariance function, $\boldsymbol{R}(t) := E\boldsymbol{\xi}_t\boldsymbol{\xi}_0^T = \boldsymbol{R}^T(-t)$, and, therefore, its spectral density, $\boldsymbol{S}(\omega) := \int_{-\infty}^{\infty} \boldsymbol{R}(t)e^{-i\omega t}dt$, is a rational function (Willems and Van Schuppen, 1980).

Note that the Gaussian process $\boldsymbol{\xi}_t$ 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 (Kalman, 1960), is called *stochastic realization*. We describe $\boldsymbol{\kappa}(t)$ and $\boldsymbol{\xi}_t$ in detail below.

Let $\boldsymbol{\Gamma}_1 \in \mathbb{R}^{d_1 \times d_1}$, $\boldsymbol{M}_1 \in \mathbb{R}^{d_1 \times d_1}$, $\boldsymbol{C}_1 \in \mathbb{R}^{q \times d_1}$, $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{d_1 \times q_1}$, $\boldsymbol{\Gamma}_2 \in \mathbb{R}^{d_2 \times d_2}$, $\boldsymbol{M}_2 \in \mathbb{R}^{d_2 \times d_2}$, $\boldsymbol{C}_2 \in \mathbb{R}^{r \times d_2}$, $\boldsymbol{\Sigma}_2 \in \mathbb{R}^{d_2 \times q_2}$ 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 $(\boldsymbol{\Gamma}_1, \boldsymbol{M}_1, \boldsymbol{C}_1)$ of matrices as follows:

$$\boldsymbol{\kappa}(t) = \boldsymbol{C}_1 e^{-\boldsymbol{\Gamma}_1|t|}\boldsymbol{M}_1\boldsymbol{C}_1^T. \tag{3}$$

The covariance of the stationary Gaussian noise process $\boldsymbol{\xi}_t$ will be expressed in terms of the triple $(\boldsymbol{\Gamma}_2, \boldsymbol{M}_2, \boldsymbol{C}_2)$. More precisely, we define it as:

$$\boldsymbol{\xi}_t = \boldsymbol{C}_2\boldsymbol{\beta}_t, \tag{4}$$

where $\boldsymbol{\beta}_t$ is the solution to the Itô SDE:

$$d\boldsymbol{\beta}_t = -\boldsymbol{\Gamma}_2\boldsymbol{\beta}_t dt + \boldsymbol{\Sigma}_2 d\boldsymbol{W}_t^{(q_2)}, \tag{5}$$

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 $\boldsymbol{\beta}_0$.

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

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

It follows from positive stability of $\boldsymbol{\Gamma}_i$ that this equation indeed has a unique solution (Bellman, 1997). 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:

$$\boldsymbol{R}(t) = \boldsymbol{C}_2 e^{-\boldsymbol{\Gamma_2}|t|} \boldsymbol{M}_2 \boldsymbol{C}_2^T, \tag{7}$$

and so the triple $(\boldsymbol{\Gamma}_2, \boldsymbol{M}_2, \boldsymbol{C}_2)$ completely specifies the probability distribution of $\boldsymbol{\xi}_t$. For concrete examples of noise process that can be realized using the above formalism, see (Lim and Wehr, 2018).

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 (Lim and Wehr, 2018). The triples that specify the memory function in (3) and the noise process in (4) are unique up to the following transformations:

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

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

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

$$m\ddot{\boldsymbol{x}}_t = \boldsymbol{F}(t, \boldsymbol{x}_t) - \boldsymbol{\gamma}_0(\boldsymbol{x}_t)\dot{\boldsymbol{x}}_t - \boldsymbol{g}(\boldsymbol{x}_t) \int_0^t \boldsymbol{C}_1 e^{-\boldsymbol{\Gamma}_1(t-s)} \boldsymbol{M}_1 \boldsymbol{C}_1^T \boldsymbol{h}(\boldsymbol{x}_s)\dot{\boldsymbol{x}}_s ds + \boldsymbol{\sigma}_0(\boldsymbol{x}_t)\boldsymbol{\eta}_t + \boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{C}_2 \boldsymbol{\beta}_t, \tag{9}$$

where $\boldsymbol{\beta}_t$ is the solution to the SDE (5). Introducing the auxiliary variable

$$\boldsymbol{y}_t = \int_0^t e^{-\boldsymbol{\Gamma}_1(t-s)} \boldsymbol{M}_1 \boldsymbol{C}_1^T \boldsymbol{h}(\boldsymbol{x}_s)\boldsymbol{v}_s ds, \tag{10}$$

and setting $\boldsymbol{\eta}_t dt = d\boldsymbol{B}_t$, where $\boldsymbol{B}_t \in \mathbb{R}^b$ is a Wiener process independent of $\boldsymbol{W}_t^{(q_2)}$, the SIDE can be cast as the following Itô SDE system for the Markov process $\boldsymbol{z}_t = (\boldsymbol{x}_t, \boldsymbol{v}_t, \boldsymbol{y}_t, \boldsymbol{\beta}_t) \in \mathbb{R}^{d \times d \times d_1 \times d_2}$:

$$d\boldsymbol{x}_t = \boldsymbol{v}_t dt, \tag{11}$$

$$m d\boldsymbol{v}_t = \boldsymbol{F}(t, \boldsymbol{x}_t)dt - \boldsymbol{\gamma}_0(\boldsymbol{x}_t)\boldsymbol{v}_t dt - \boldsymbol{g}(\boldsymbol{x}_t)\boldsymbol{C}_1 \boldsymbol{y}_t dt + \boldsymbol{\sigma}_0(\boldsymbol{x}_t) d\boldsymbol{B}_t + \boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{C}_2 \boldsymbol{\beta}_t dt, \tag{12}$$

$$d\boldsymbol{y}_t = -\boldsymbol{\Gamma}_1 \boldsymbol{y}_t dt + \boldsymbol{M}_1 \boldsymbol{C}_1^T \boldsymbol{h}(\boldsymbol{x}_t)\boldsymbol{v}_t dt, \tag{13}$$

$$d\boldsymbol{\beta}_t = -\boldsymbol{\Gamma}_2 \boldsymbol{\beta}_t dt + \boldsymbol{\Sigma}_2 d\boldsymbol{W}_t^{(q_2)}. \tag{14}$$

We refer to Appendix A for three examples of GLE system arising in nonequilibrium statistical mechanics. Several remarks concerning the system (11)-(14) are now in order.

Remark 2.1 *On one hand, $\boldsymbol{z}_t$ is the solution to a hypoelliptic SDE system of the form*

$$d\boldsymbol{z}_t = \boldsymbol{a}(t, \boldsymbol{z}_t)dt + \boldsymbol{B}(t, \boldsymbol{z}_t)d\boldsymbol{U}_t, \tag{15}$$

where $\boldsymbol{U}_t$ is a Wiener process and $\boldsymbol{B}$ is a matrix-valued function that is not full rank, since the noise does not act in all directions of $\boldsymbol{z}$. Therefore, from mathematical point of view our study of the GLE and functionals along its trajectory can be viewed as study of the above hypoelliptic SDE system (Pavliotis, 2014) and the associated functionals. On the other hand, the process $\boldsymbol{r}_t = (\boldsymbol{x}_t, \boldsymbol{v}_t, \boldsymbol{y}_t)$ gives the coordinates of the generalized Langevin system. It is a non-Markov process satisfying an Itô SDE of the form:

$$d\boldsymbol{r}_t = \boldsymbol{b}(t, \boldsymbol{r}_t)dt + \boldsymbol{\Phi}(\boldsymbol{r}_t)d\boldsymbol{B}_t + \boldsymbol{\Phi}_a(\boldsymbol{r}_t)\boldsymbol{\beta}_t dt, \tag{16}$$

where the driving noise consists of a white noise and a Gaussian colored noise. Note that the augmented process $\boldsymbol{z}_t = (\boldsymbol{r}_t, \boldsymbol{\beta}_t)$ is the Markov process solving the SDE (15).

Remark 2.2 *One could have absorbed the constant matrices $\boldsymbol{C}_i$ into the coefficients $\boldsymbol{\sigma}$, $\boldsymbol{g}$, $\boldsymbol{h}$ but we choose to keep them as parameters for our memory function and colored noise models. The one-dimensional case $(d = 1)$ where $\boldsymbol{C}_i = 1$, $\boldsymbol{\Gamma}_i = \alpha_i > 0$, $\boldsymbol{\Sigma}_i = \alpha_i$, $\boldsymbol{M}_i = \alpha_i/2$, for $i = 1, 2$ (we will drop the boldface when denoting the processes and coefficients in the one-dimensional case – for instance, $\boldsymbol{x}_t = x_t$, $\boldsymbol{g} = g$, $\boldsymbol{W}_t = W_t$, etc.), follows as a special case. In this case, the memory function and covariance function of the colored noise process are exponentials, with possibly different decay rates α_i.*

Remark 2.3 *In order to be able to study the GLE as a finite-dimensional Markovian system it is crucial that the memory function and covariance function of the colored noise process be Bohl. In the case where, for instance, these functions decay as a power law, the resulting GLE cannot be studied as a finite-dimensional SDE system and one needs to work in the infinite-dimensional setting (Kupferman, 2004; Glatt-Holtz et al., 2018). However, our formalism allows us to approximate an arbitrary memory function, such as the ones decaying as a power law (long-range memory), on a finite time scale (Siegle et al., 2011). Therefore, our finite-dimensional consideration allows us to cover a sufficiently large class of systems with memory.*

3. Functionals Along Trajectories of GLEs

We are interested in the asymptotic behavior of a class of functionals along the trajectory $(\boldsymbol{r}_t)_{t\geq 0}$, where $\boldsymbol{r}_t = (\boldsymbol{x}_t, \boldsymbol{v}_t, \boldsymbol{y}_t)$[1], of the generalized Langevin systems described by (9) in various limiting situations. These situations are when wide separation of time scales exists in the systems and thereby allowing simplification of the dynamics via elimination of the fast degrees of freedom and description of the system solely in terms of the slow degrees of freedom. These functionals take the form of:

$$\mathcal{F}_t = \int_0^t r(s, \boldsymbol{r}_s)ds + \int_0^t \boldsymbol{p}(s, \boldsymbol{r}_s) \circ^? d\boldsymbol{r}_s \tag{17}$$

1. Since $\boldsymbol{y}_t$ is a functional of $(\boldsymbol{x}_s, \boldsymbol{v}_s)_{0\leq s\leq t}$, it suffices to consider the trajectory $(\boldsymbol{x}_t, \boldsymbol{v}_t)_{t\geq 0}$ instead of $(\boldsymbol{x}_t, \boldsymbol{v}_t, \boldsymbol{y}_t)_{t\geq 0}$

which, in differential form, is:

$$d\mathcal{F}_t = r(t, \boldsymbol{r}_t)dt + \boldsymbol{p}(t, \boldsymbol{r}_t) \circ^? d\boldsymbol{r}_t, \tag{18}$$

where $\circ^?$ denotes the (to be specified) discretization rule defining the stochastic integral in (17). Since different discretization rules lead to different properties of the functional, the discretization rule should be assigned in such a way that the physical behavior of the modeled system is captured correctly (Hottovy et al., 2012; Farago and Grønbech-Jensen, 2014; Sokolov, 2010; Yang et al., 2017). Here and throughout the paper, we are using calligraphic font for denoting a functional. We emphasize that, in contrast to the case of Langevin-Kramers model, the process $\boldsymbol{r}_t$, being a component of the Markov process $(\boldsymbol{r}_t, \boldsymbol{\beta}_t)$, is generally non-Markov.

We are going to introduce and define a special subclass of functionals (17) along the trajectory of the GLE (9) (or equivalently the SDE system (11)-(14)) in the following. These functionals are various thermodynamic functionals of interest arising in stochastic thermodynamics (Seifert, 2012) of the GLE. To begin with, we split the force field as $\boldsymbol{F}(t, \boldsymbol{x}) = -\boldsymbol{\nabla}_{\boldsymbol{x}} U(t, \boldsymbol{x}) + \boldsymbol{f}_{nc}(t, \boldsymbol{x})$, where the scalar-valued function U represents a potential and $\boldsymbol{f}_{nc}$ represents a non-conservative external force, driving the system out of equilibrium.

When considering these functionals, there are two cases of interest. The first case is the case when $\boldsymbol{\sigma}_0 = \boldsymbol{0}$, in which case there is no ambiguity in defining the stochastic integral in (17). The second case is when $\boldsymbol{\sigma}_0$ is non-zero, in which case we need to specify the convention $\circ^?$ for the stochastic integral, usually taken to be Stratonovich. We will consider only the first case here. Therefore, *we set $\boldsymbol{\sigma}_0$ to zero from now on unless specified otherwise*, and replace $\circ^?$ by $\cdot$ to denote dot product. More precisely, when $\boldsymbol{\sigma}_0$ vanishes (and therefore the corresponding $\boldsymbol{\Phi}$ in (16) vanishes), the equation for $\boldsymbol{r}_t$ does not contain a white noise term. In this case, the process $\boldsymbol{r}_t$ is more regular than the one in the case of non-vanishing $\boldsymbol{\sigma}_0$ and the stochastic integral defining $\boldsymbol{r}_t$ is uniquely defined, in particular its properties are independent of the discretization choice.

We define a *heat-like* and *work-like* functional along the stochastic trajectory $(\boldsymbol{r}_t)_{t\geq 0}$ as the functional satisfying the following (controlled) differential equations:

$$d\mathcal{Q}_t = \left(-\boldsymbol{g}(\boldsymbol{x}_t)\int_0^t \boldsymbol{\kappa}(t-s)\boldsymbol{h}(\boldsymbol{x}_s)\boldsymbol{v}_s ds + \boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{\xi}_t - \gamma_0(\boldsymbol{x}_t)\boldsymbol{v}_t\right) \cdot d\boldsymbol{x}_t, \tag{19}$$

$$= \int_0^t \left(m\boldsymbol{v}_s \cdot d\boldsymbol{v}_s - \boldsymbol{F}(s, \boldsymbol{x}_s) \cdot d\boldsymbol{x}_s\right), \tag{20}$$

$$d\mathcal{W}_t = \frac{\partial U}{\partial t}dt + \boldsymbol{f}_{nc}(t, \boldsymbol{x}_t) \cdot d\boldsymbol{x}_t \tag{21}$$

respectively. The above functionals are free of ambiguities in the discretization procedure and are thus uniquely defined.

We emphasize that, as we discussed in (Bo et al., 2019), the functionals above are not, generally and strictly speaking, defining physical heat and work for the generalized Langevin systems. This emphasis leads to our usage of the terminology "heat-like" and "work-like" functional instead of heat and work throughout the paper. These heat-like and work-like functionals are rather defined in a manner that ensures a first law for energy balance is satisfied as follows. Let us define[2] the internal energy of the system as:

$$\mathcal{E}_t = \frac{1}{2} m |\boldsymbol{v}_t|^2 + U(t, \boldsymbol{x}_t). \tag{22}$$

Then, the above definitions for heat-like and work-like functional are consistent with the first law of stochastic thermodynamics in the sense that the energy $\mathcal{E}_t$ is conserved along individual trajectories. Indeed, using $d\mathcal{E} = m\boldsymbol{v} \cdot d\boldsymbol{v} + dU$, one obtains the law:

$$d\mathcal{E} = d\mathcal{W} + d\mathcal{Q}, \tag{23}$$

where $\mathcal{W}$ and $\mathcal{Q}$ are defined in (20) and (21) respectively, and we use the convention that $\mathcal{Q} < 0$ if the heat is transferred or dissipated from the system into the environment.

Next, we specialize the above definition to the setting where the heat-like and work-like functional become physical heat and work. This is the case where $\boldsymbol{\gamma}_0 = \boldsymbol{0}$, the fluctuation-dissipation relation of the second kind holds, and the colored noise models a heat bath which is in equilibrium at temperature T. In this case, the resulting GLE can be derived from a microscopic Hamiltonian model (see an earlier remark in Section 2) for a Brownian particle (weakly) interacting with an equilibrium heat bath at temperature T. The *thermodynamic entropy produced in the environment*, from an initial state $(\boldsymbol{x}_0, \boldsymbol{v}_0)$ at the initial time to a final state $(\boldsymbol{x}_t, \boldsymbol{v}_t)$ at time t, is defined as:

$$\mathcal{S}_t = -\beta \mathcal{Q}_t = \beta \int_0^t \left(\boldsymbol{F}(s, \boldsymbol{x}_s) \cdot d\boldsymbol{x}_s - m\boldsymbol{v}_s \cdot d\boldsymbol{v}_s \right). \tag{24}$$

where $\beta = 1/k_B T$. It is a measure of irreversibility of the generalized Langevin dynamics. The heat can be interpreted as the change of bath energy over the system trajectory and it is a functional of the system history alone (Aurell, 2018). In the more general case beyond the above setting, the above definition does not generally define a thermodynamic entropy, and so we are going to simply refer to it as an entropy-like functional. Finally, we emphasize that the integrals defining the dynamical process $\boldsymbol{r}_t$ and functionals $\mathcal{Q}_t$, $\mathcal{R}_t$ here are *uniquely defined and will be taken to be the starting point for multiple time scale analysis (homogenization)*, for which (the interpretation of) their limiting expression will be of interest.

4. Homogenization of Slow-Fast SDE Systems and Their Functionals

Asymptotic analysis of functionals along trajectories of approximating stochastic processes has long histories and is an important tool for stochastic modeling of noisy systems. An important early example comes from the classic work of Wong and Zakai (Wong and Zakai, 1965), who considered the limiting behavior of the family of real-valued stochastic integrals $y_n(t) = \int_0^t u(B_n(s)) dB_n(s)$, where u is some sufficiently nice function and $B_n(t)$ is a sequence of sufficiently smooth functions approximating a Wiener process. They found that $y_n(t)$ converges to the Stratonovich integral, $y(t) = \int_0^t u(B(s)) \circ dB(s)$, where $\circ$ denotes Stratonovich

2. Note that a fluctuating internal energy is by no means unique, but can assume many different forms which all would give the same "mean value", but different higher moments (Hänggi, 2019). As a consequence, (22) is nothing more than a definition.

product and $B(t)$ is a Wiener process, in the limit as $n \to \infty$. The result holds in one dimension and may fail in higher dimensions, in which case one has additional (anomalous) drift terms due to Lévy area correction (Lévy et al., 1951; Ikeda and Watanabe, 2014; Sussmann, 1991) (see Section 11.7.7 in (Pavliotis and Stuart, 2008) for an explicit example).

Each $y_n(t)$ is a functional along trajectories of the approximating functions $B_n(t)$. In the special case where the fast process $B_n(t)$ satisfies an Itô SDE, driven by a white noise, the key technique is to embed the functional into a higher dimensional Markov process. The goal is then to determine the limiting behavior of the slow process $y_n(t)$, as components of the Markov process, as $n \to \infty$. In the context of the above example, one has $dz_n(t) = dB_n(t)$, $dy_n(t) = u(z_n(t))dz_n(t)$, and $B_n(t)$ is a process embedded in a SDE system. If, for instance, $B_n(t)$ is an integrated Ornstein-Uhlenbeck process, then we have $dB_n(t) = C_n(t)dt$, $dC_n(t) = -\lambda_n C_n(t)dt + \sigma_n dW_t$, where W_t is a Wiener process and λ_n, σ_n are some suitable increasing sequences in n.

We are going to study a generalization of the above example problem to a class of multi-dimensional diffusion processes. Our setting is sufficiently general to cover all the asymptotic problems (homogenization) for GLEs and their functionals in this paper. In this section we focus on homogenization in the general setting. Examples and applications in the context of stochastic energetics will be studied and discussed in detail in Section 5 and Section 6.

Consider the following family of Itô SDE systems for $\boldsymbol{Z}_t^\epsilon = (\boldsymbol{X}_t^\epsilon, \boldsymbol{Y}_t^\epsilon, \boldsymbol{\mathcal{A}}_t^\epsilon, \mathcal{B}_t^\epsilon) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R}$:

$$d\boldsymbol{X}_t^\epsilon = \boldsymbol{U}_1(t, \boldsymbol{X}_t^\epsilon)\boldsymbol{Y}_t^\epsilon dt + \boldsymbol{u}_1(t, \boldsymbol{X}_t^\epsilon)dt + \tilde{\boldsymbol{\sigma}}(t, \boldsymbol{X}_t^\epsilon)d\tilde{\boldsymbol{W}}_t, \tag{25}$$

$$\epsilon d\boldsymbol{Y}_t^\epsilon = -\boldsymbol{U}_2(t, \boldsymbol{X}_t^\epsilon)\boldsymbol{Y}_t^\epsilon dt + \boldsymbol{u}_2(t, \boldsymbol{X}_t^\epsilon)dt + \boldsymbol{\sigma}(t, \boldsymbol{X}_t^\epsilon)d\boldsymbol{W}_t, \tag{26}$$

$$d\boldsymbol{\mathcal{A}}_t^\epsilon = \boldsymbol{r}(t, \boldsymbol{X}_t^\epsilon)dt + \boldsymbol{P}(t, \boldsymbol{X}_t^\epsilon)d\boldsymbol{X}_t^\epsilon, \tag{27}$$

$$d\mathcal{B}_t^\epsilon = \epsilon \boldsymbol{Y}_t^\epsilon \cdot d\boldsymbol{Y}_t^\epsilon, \tag{28}$$

where $\boldsymbol{U}_1 : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$, $\boldsymbol{U}_2 : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{m\times m}$, $\boldsymbol{u}_1 : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{u}_2 : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^m$, $\tilde{\boldsymbol{\sigma}} : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{n\times d_s}$, $\boldsymbol{\sigma} : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{m\times d_f}$, $\tilde{\boldsymbol{W}}_t \in \mathbb{R}^{d_s}$ and $\boldsymbol{W}_t \in \mathbb{R}^{d_f}$ are independent Wiener processes on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ such that the usual conditions (Karatzas and Shreve, 2014) hold, $\boldsymbol{r} : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^l$, $\boldsymbol{P} : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^{l\times n}$, $\epsilon > 0$ is a small parameter, and $\cdot$ denotes dot product. The variables $\boldsymbol{Z}_t^\epsilon$ model physical processes or states of a system with dimensionless variables. Let $\mathbb{E}$ denote expectation with respect to $\mathbb{P}$.

We take $\mathcal{B}_0^\epsilon = \epsilon|\boldsymbol{Y}_0^\epsilon|^2/2$, so that

$$\mathcal{B}_t^\epsilon = \mathcal{B}_0^\epsilon + \epsilon \int_0^t \boldsymbol{Y}_s^\epsilon \cdot d\boldsymbol{Y}_s^\epsilon = \mathcal{B}_0^\epsilon + \frac{\epsilon}{2}\int_0^t d\left(|\boldsymbol{Y}_s^\epsilon|^2\right) = \frac{\epsilon}{2}|\boldsymbol{Y}_t^\epsilon|^2. \tag{29}$$

The above systems are variants of the one considered in (Bo and Celani, 2013) (see also (Bo and Celani, 2017, 2014)). All the equations contain fast dynamics but the dynamics in $\boldsymbol{Y}^\epsilon$ is one order of magnitude faster than in $\boldsymbol{X}^\epsilon$, $\boldsymbol{\mathcal{A}}^\epsilon$ and $\mathcal{B}^\epsilon$. Our goal is to eliminate the variable $\boldsymbol{Y}^\epsilon$ in (25)-(28) and derive an effective description for the slow process $\boldsymbol{Q}_t^\epsilon = (\boldsymbol{X}_t^\epsilon, \boldsymbol{\mathcal{A}}_t^\epsilon, \mathcal{B}_t^\epsilon)$ in the limit $\epsilon \to 0$.

We now introduce our notation and provide some reminders on transformation of stochastic integrals.

Notation. Consider the diffusion process $\boldsymbol{Z}_t \in \mathbb{R}^N$, $t \geq 0$, satisfying the Itô SDE:

$$d\boldsymbol{Z}_t = \boldsymbol{b}(t, \boldsymbol{Z}_t)dt + \boldsymbol{\sigma}(t, \boldsymbol{Z}_t)d\boldsymbol{W}_t, \tag{30}$$

where $\boldsymbol{b} : \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}^N$, $\boldsymbol{\sigma} : \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}^{N\times M}$ (differentiable in $\boldsymbol{Z}$), and $\boldsymbol{W}_t \in \mathbb{R}^M$ is a Wiener process. Equivalently, it can be cast as the following *Stratonovich SDE*:

$$d\boldsymbol{Z}_t = \boldsymbol{u}(t, \boldsymbol{Z}_t)dt + \boldsymbol{\sigma}(t, \boldsymbol{Z}_t) \circ d\boldsymbol{W}_t, \tag{31}$$

where $\boldsymbol{u}(t, \boldsymbol{Z}_t) = \boldsymbol{b}(t, \boldsymbol{Z}_t) - \boldsymbol{c}(t, \boldsymbol{Z}_t)$, the symbol $\circ$ denotes Stratonovich convention (without the symbol $\circ$, Itô convention is taken), and, in index-free notation,

$$\boldsymbol{c} = \frac{1}{2}[\boldsymbol{\nabla} \cdot (\boldsymbol{\sigma}\boldsymbol{\sigma}^T) - \boldsymbol{\sigma}\boldsymbol{\nabla} \cdot (\boldsymbol{\sigma}^T)]. \tag{32}$$

In the above, $\boldsymbol{\nabla}\cdot$ denotes divergence operator which contracts a matrix-valued function to a vector-valued function: for the matrix-valued function $\boldsymbol{A}(\boldsymbol{Z})$, the ith component of its divergence is given by

$$(\boldsymbol{\nabla} \cdot \boldsymbol{A})^i = \sum_{j=1}^N \frac{\partial A^{ij}}{\partial Z^j}. \tag{33}$$

Equivalently, in components,

$$c^i = \frac{1}{2}\frac{\partial \sigma^{ij}}{\partial X^k}\sigma^{kj}, \tag{34}$$

where σ^{ij} denotes the (i, j)-entry of the matrix $\boldsymbol{\sigma}$, Z^k the kth component of the vector $\boldsymbol{Z}$, and we have used Einstein's summation convention for repeated indices.

We make the following assumptions on the SDE systems (25)-(28):

Assumption 4.1 *The global solutions, defined on $[0, T]$, to the pre-limit SDEs* (25)-(28) *and to the limiting SDEs* (35)-(37) *a.s. exist and are unique for all $\epsilon > 0$ (i.e. there are no explosions).*

Assumption 4.2 *The matrix-valued functions $\{\boldsymbol{U}_2(t, \boldsymbol{X}); t \in [0, T], \boldsymbol{X} \in \mathbb{R}^n\}$ are uniformly positive stable, i.e. all real parts of the eigenvalues of $\boldsymbol{U}_2(t, \boldsymbol{X})$ are bounded from below, uniformly in t and $\boldsymbol{X}$, by a positive constant.*

Assumption 4.3 *For $t \in [0, T]$, $\boldsymbol{X} \in \mathbb{R}^n$, and $i = 1, 2$, the functions $\boldsymbol{u}_i(t, \boldsymbol{X})$, $\tilde{\boldsymbol{\sigma}}(t, \boldsymbol{X})$, $\boldsymbol{\sigma}(t, \boldsymbol{X})$, $\boldsymbol{r}(t, \boldsymbol{X})$ are continuous and bounded in t and $\boldsymbol{X}$, and Lipschitz in $\boldsymbol{X}$, whereas the functions $\boldsymbol{U}_i(t, \boldsymbol{X})$, $\boldsymbol{P}(t, \boldsymbol{X})$, $(\boldsymbol{U}_i)_{\boldsymbol{X}}(t, \boldsymbol{X})$, $\boldsymbol{P}_{\boldsymbol{X}}(t, \boldsymbol{X})$ are continuous in t, continuously differentiable in $\boldsymbol{X}$, bounded in t and $\boldsymbol{X}$, and Lipschitz in $\boldsymbol{X}$. Moreover, the functions $(\boldsymbol{U}_i)_{\boldsymbol{X}\boldsymbol{X}}(t, \boldsymbol{X})$ $(i = 1, 2)$ and $\boldsymbol{P}_{\boldsymbol{X}\boldsymbol{X}}(t, \boldsymbol{X})$ are bounded for every $t \in [0, T]$ and $\boldsymbol{X} \in \mathbb{R}^n$.*

Assumption 4.4 *The initial condition $\boldsymbol{X}_0^\epsilon = \boldsymbol{X}^\epsilon \in \mathbb{R}^n$ is an $\mathcal{F}_0$-measurable random variable that may depend on ϵ, and we assume that $\mathbb{E}[|\boldsymbol{X}^\epsilon|^p] = O(1)$ as $\epsilon \to 0$ for all $p > 0$. Also, $\boldsymbol{X}^\epsilon$ converges, in the limit as $\epsilon \to 0$, to a random variable $\boldsymbol{X}$ as follows: $\mathbb{E}[|\boldsymbol{X}^\epsilon - \boldsymbol{X}|^p] = O(\epsilon^{pr_0})$, where $r_0 > 1/2$ is a constant, as $\epsilon \to 0$. The same conditions are assumed for $\boldsymbol{A}_0^\epsilon$. The initial condition $\boldsymbol{Y}_0^\epsilon = \boldsymbol{Y}^\epsilon \in \mathbb{R}^m$ is an $\mathcal{F}_0$-measurable random variable that may depend on ϵ, and we assume that for every $p > 0$, $\mathbb{E}[|\epsilon\boldsymbol{Y}^\epsilon|^p] = O(\epsilon^\alpha)$ as $\epsilon \to 0$, for some $\alpha \geq p/2$.*

The following theorem follows from a straightforward application of Theorem B.1. The last statement in the theorem follows from the proof of Theorem B.1 (see (Lim et al., 2020) for details).

Theorem 4.1 *Under the Assumption 4.1-4.4, in the limit $\epsilon \to 0$, the family of processes $(\boldsymbol{X}_t^\epsilon, \boldsymbol{\mathcal{A}}_t^\epsilon)$, $t \in [0,T]$, converges to $(\boldsymbol{X}_t, \boldsymbol{\mathcal{A}}_t)$ solving the Itô SDE:*

$$\begin{aligned} d\boldsymbol{X}_t &= [\boldsymbol{u}_1(t,\boldsymbol{X}_t) + \boldsymbol{U}_1(t,\boldsymbol{X}_t)\boldsymbol{U}_2^{-1}(t,\boldsymbol{X}_t)\boldsymbol{u}_2(t,\boldsymbol{X}_t)]dt + \boldsymbol{S}_{Ito}(t,\boldsymbol{X}_t)dt + \tilde{\boldsymbol{\sigma}}(t,\boldsymbol{X}_t)d\tilde{\boldsymbol{W}}_t \\ &\quad + \boldsymbol{U}_1(t,\boldsymbol{X}_t)\boldsymbol{U}_2^{-1}(t,\boldsymbol{X}_t)\boldsymbol{\sigma}(t,\boldsymbol{X}_t)d\boldsymbol{W}_t, \qquad (35) \\ d\boldsymbol{\mathcal{A}}_t &= \boldsymbol{r}(t,\boldsymbol{X}_t)dt + \boldsymbol{P}(t,\boldsymbol{X}_t)d\boldsymbol{X}_t + d\boldsymbol{\mathcal{A}}_t', \qquad (36) \\ d\boldsymbol{\mathcal{A}}_t' &= [\boldsymbol{\nabla}\cdot(\boldsymbol{P}(t,\boldsymbol{X}_t)\boldsymbol{U}_1(t,\boldsymbol{X}_t)\boldsymbol{\mu}(t,\boldsymbol{X}_t)\boldsymbol{U}_1^T(t,\boldsymbol{X}_t)), \\ &\quad - \boldsymbol{P}(t,\boldsymbol{X}_t)\boldsymbol{\nabla}\cdot(\boldsymbol{U}_1(t,\boldsymbol{X}_t)\boldsymbol{\mu}(t,\boldsymbol{X}_t)\boldsymbol{U}_1^T(t,\boldsymbol{X}_t))]dt, \quad \textit{or, in component:} \qquad (37) \\ d(\mathcal{A}_t')^k &= U_1^{ia}U_1^{jb}(U_2^{-1}J)^{ab}\frac{\partial P^{ki}}{\partial X^j}dt. \qquad (38) \end{aligned}$$

In the above, $\boldsymbol{S}_{Ito}$ is the noise-induced drift:

$$\boldsymbol{S}_{Ito} = \boldsymbol{\nabla}\cdot(\boldsymbol{U}_1\boldsymbol{U}_2^{-1}\boldsymbol{J}\boldsymbol{U}_1^T) - \boldsymbol{U}_1\boldsymbol{U}_2^{-1}\boldsymbol{\nabla}\cdot(\boldsymbol{J}\boldsymbol{U}_1^T), \qquad (39)$$

with $\boldsymbol{J}$ solving the Lyapunov equation

$$\boldsymbol{U}_2\boldsymbol{J} + \boldsymbol{J}\boldsymbol{U}_2^T = \boldsymbol{\sigma}\boldsymbol{\sigma}^T, \qquad (40)$$

and $\boldsymbol{\mu} = \boldsymbol{U}_2^{-1}\boldsymbol{J}$. The convergence is in the following sense: for all finite $T > 0$,

$$\sup_{t\in[0,T]} |\boldsymbol{X}_t^\epsilon - \boldsymbol{X}_t| \to 0, \quad \sup_{t\in[0,T]} |\boldsymbol{\mathcal{A}}_t^\epsilon - \boldsymbol{\mathcal{A}}_t| \to 0, \qquad (41)$$

in probability, in the limit as $\epsilon \to 0$. The family of functionals $\mathcal{B}_t^\epsilon = \frac{\epsilon}{2}|\boldsymbol{Y}_t^\epsilon|^2$ converges to $Tr(\boldsymbol{J}(t,\boldsymbol{X}_t))$ as $\epsilon \to 0$ in the following sense: for all finite $T > 0$,

$$\sup_{t\in[0,T]} \int_0^t |\mathcal{B}_s^\epsilon - Tr(\boldsymbol{J}(s,\boldsymbol{X}_s)|ds \to 0 \qquad (42)$$

in probability as $\epsilon \to 0$.

The following remarks describe the link between *symmetry breaking* (violation of a detailed balance condition) and *area anomaly* (concerning the appearance of the anomalous contributions, $\boldsymbol{S}_{Ito}dt$ and $d\boldsymbol{\mathcal{A}}_t'$, in the homogenized equations).

Remark 4.1 *We recall some connections to relevant concepts from nonequilibrium statistical mechanics (Pavliotis, 2014). Define the matrix $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, by*

$$\mu^{ab} := \int_0^\infty \mathbb{E}Y_\tau^a Y_0^b d\tau, \qquad (43)$$

$$2\mu_S^{ab} := \mu^{ab} + \mu^{ba} =: \nu^{ac}\nu^{bc}. \qquad (44)$$

Let $\boldsymbol{L}_0$ be the infinitesimal generator corresponding to the fast dynamics in $\boldsymbol{Y}$, i.e. $\boldsymbol{L}_0 = -\boldsymbol{U}_2(t,\boldsymbol{X})\boldsymbol{Y}\cdot\boldsymbol{\nabla}_{\boldsymbol{Y}} + \frac{1}{2}(\boldsymbol{\sigma}(t,\boldsymbol{X})\boldsymbol{\sigma}^T(t,\boldsymbol{X})) : \boldsymbol{\nabla}_{\boldsymbol{Y}}\boldsymbol{\nabla}_{\boldsymbol{Y}}$, where $\boldsymbol{A} : \boldsymbol{\nabla}_{\boldsymbol{Y}}\boldsymbol{\nabla}_{\boldsymbol{Y}} := \sum_{i,j} A^{ij}\frac{\partial^2}{\partial Y^i \partial Y^j}$. Using the time integral representation formula for $(-\boldsymbol{L}_0)^{-1}$, one finds $\mu^{ab} = \overline{Y^b(-\boldsymbol{L}_0^{-1}Y^a)}$, where overbar denotes averaging with respect to the invariant density of a mean zero Gaussian process with the covariance matrix $\boldsymbol{J}$. This is an example of the Green-Kubo formula, which is important for the calculation of transport coefficients (Pavliotis, 2010). It is straightforward to compute that $\boldsymbol{\mu} = \boldsymbol{U}_2^{-1}\boldsymbol{J}$ and $\boldsymbol{\nu} = \boldsymbol{U}_2^{-1}\boldsymbol{\sigma}$. Recall that $\boldsymbol{J}$ solves the Lyapunov equation (40), which can be rewritten as $\boldsymbol{L} + \boldsymbol{L}^T = \boldsymbol{D}$, where $\boldsymbol{L} := \boldsymbol{U}_2\boldsymbol{J}$ is the Onsager matrix of kinetic coefficient (associated to the fast dynamics) and $\boldsymbol{D} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$ is the diffusion matrix (Godrèche and Luck, 2018).

It is well known that the detailed balance condition (the condition for the fast process to be reversible, or equivalently, for its infinitesimal generator to be symmetric), for a given t and $\boldsymbol{X}$, holds if and only if $\boldsymbol{U}_2\boldsymbol{D}$ is symmetric, i.e. $\boldsymbol{U}_2\boldsymbol{D} = \boldsymbol{D}\boldsymbol{U}_2^T$ (Gardiner, 2009). In this case, the stationary covariance matrix is $\boldsymbol{U}_2^{-1}\boldsymbol{D}/2$ and the corresponding stationary state is an equilibrium one. In particular, this symmetry condition implies that $\boldsymbol{\mu}$ is symmetric and $\boldsymbol{\mu} = \boldsymbol{\mu}_S$. The converse is not true unless $\boldsymbol{U}_2^2\boldsymbol{J}$ is symmetric. When the symmetry condition is broken, the fast process is irreversible and has a nonequilibrium stationary state. One can quantify the irreversibility of the process as follows. We write $\boldsymbol{L} = \boldsymbol{D}/2 + \boldsymbol{Q}$ and $\boldsymbol{L}^T = \boldsymbol{D}/2 - \boldsymbol{Q}$ so that we can use $\boldsymbol{Q} = (\boldsymbol{L} - \boldsymbol{L}^T)/2$, the antisymmetric part of the Onsager matrix, to measure the irreversibility of the fast process. If the fast process is reversible, then the Onsager matrix $\boldsymbol{L} = \boldsymbol{D}/2$ is symmetric and $\boldsymbol{Q} = \boldsymbol{0}$. We refer to (Godrèche and Luck, 2018; Macieszczak et al., 2018) and the references therein for a list of works on quantification of the asymmetry of the Onsager matrix.

Remark 4.2 *In the case when $\tilde{\boldsymbol{\sigma}} = \tilde{\boldsymbol{\sigma}}(t)$ and $\boldsymbol{\sigma} = \boldsymbol{\sigma}(t)$ are independent of the state, we have:*

$$\begin{aligned} d\boldsymbol{X}_t = (\boldsymbol{u}_1(t,\boldsymbol{X}_t) + \boldsymbol{U}_1(t,\boldsymbol{X}_t)\boldsymbol{U}_2^{-1}(t,\boldsymbol{X}_t)\boldsymbol{u}_2(t,\boldsymbol{X}_t) + d\boldsymbol{X}_t'' \\ + \tilde{\boldsymbol{\sigma}}(t)d\tilde{\boldsymbol{W}}_t + \boldsymbol{U}_1(t,\boldsymbol{X}_t)\boldsymbol{U}_2^{-1}(t,\boldsymbol{X}_t)\boldsymbol{\sigma}(t)\circ d\boldsymbol{W}_t, \end{aligned} \tag{45}$$

with $d\boldsymbol{X}_t'' = \boldsymbol{H}_{Str}(t,\boldsymbol{X}_t))dt$, where $\boldsymbol{H}_{Str}$ is the additional drift term which can be written in two equivalent ways. The first one is in terms of $\boldsymbol{Q}$, $\boldsymbol{L}$ and $\boldsymbol{\nu}$ introduced earlier and $\boldsymbol{H}_{Str}$ is written compactly as a sum of three contributions:

$$\boldsymbol{H}_{Str} = \boldsymbol{\nabla}\cdot(\boldsymbol{U}_1\boldsymbol{U}_2^{-1}(\boldsymbol{U}_1\boldsymbol{U}_2^{-1}\boldsymbol{Q})^T) - \boldsymbol{U}_1\boldsymbol{U}_2^{-1}\boldsymbol{\nabla}\cdot((\boldsymbol{U}_1\boldsymbol{U}_2^{-1}\boldsymbol{L})^T) + \frac{1}{2}(\boldsymbol{U}_1\boldsymbol{U}_2^{-1})\boldsymbol{\sigma}\boldsymbol{\nabla}\cdot((\boldsymbol{U}_1\boldsymbol{\nu})^T). \tag{46}$$

The second way is in terms of $\boldsymbol{Q}$, Lie brackets of vector fields and $\boldsymbol{\nu}$:

$$H^i_{Str} = \frac{\partial(U_1U_2^{-1})^{ip}}{\partial X^k}(U_1U_2^{-1})^{kl}Q^{lp} = \frac{1}{2}Q^{lp}[\boldsymbol{G}_l,\boldsymbol{G}_p]^i, \tag{47}$$

where the vector fields $\boldsymbol{G}_l$ are associated to the lth column of the matrix $\boldsymbol{U}_1\boldsymbol{U}_2^{-1}$ and $[\cdot,\cdot]$ denotes the Lie bracket of two vector fields. The antisymmetric matrix $\boldsymbol{Q}$ (which, as discussed earlier, measures the irreversibility of the fast process) encodes the stochastic area of the limiting dynamical process, and $\boldsymbol{H}_{Str}$ would vanish in the one-dimensional case (c.f.

(Ikeda and Watanabe, 2014), or Section 2 in (Lejay and Lyons, 2003) for the point of view of interpolation problem for trajectories).

The irreversibility of the fast process generates macroscopic current in the stationary state and induces some loops in the trajectories. It turns out that the area generated by these loops is of $O(1)$ as $\epsilon \to 0$. As a result, zooming in the small scale $\boldsymbol{X}_t$ "spins" around a modified mean trajectory (Lejay et al., 2002; Lejay and Lyons, 2003). We refer the reader to Section 5 for an illustration of such phenomenon in a simple example. The phenomena of area anomaly has been discovered and studied recently in different problem settings (Chevyrev et al., 2016; Lopusanschi and Simon, 2017, 2018) (see also the references therein). One rigorous framework for understanding these phenomena is based on the theory of rough paths (Lyons, 1998; Friz and Victoir, 2010; Friz and Hairer, 2020).

Remark 4.3 *The evolution of the effective functional is described by:*

$$d\boldsymbol{\mathcal{A}}_t = \boldsymbol{r}dt + \boldsymbol{P} \circ d\boldsymbol{X}_t + d\boldsymbol{\mathcal{A}}_t'', \tag{48}$$

$$d\boldsymbol{\mathcal{A}}_t'' = \left[\boldsymbol{\nabla} \cdot \left(\boldsymbol{P}\left(\boldsymbol{U}_1 \boldsymbol{\mu}_A^T \boldsymbol{U}_1^T - \frac{1}{2}\tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^T\right)\right) - \boldsymbol{P}\boldsymbol{\nabla} \cdot \left(\boldsymbol{U}_1 \boldsymbol{\mu}_A^T \boldsymbol{U}_1^T - \frac{1}{2}\tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^T\right)\right] dt, \tag{49}$$

where $\boldsymbol{\mu}_A$ is the antisymmetric part of $\boldsymbol{\mu}$. In component form, we have:

$$d(\mathcal{A}_t'')^i = \frac{1}{2} U_1^{kb} U_1^{ja} \mu_A^{ab} \left(\frac{\partial P^{ij}}{\partial X^k} - \frac{\partial P^{ik}}{\partial X^j}\right) dt - \frac{\partial P^{ij}}{\partial X^k} (\tilde{\sigma}\tilde{\sigma}^T)^{kj} dt. \tag{50}$$

Therefore, whenever $\boldsymbol{\mu}_A = \boldsymbol{0}$ (a sufficient condition for this is when $\boldsymbol{Q} = \boldsymbol{0}$ and $\tilde{\boldsymbol{\sigma}} = \boldsymbol{0}$), $d\boldsymbol{\mathcal{A}}_t'' = \boldsymbol{0}$ and the effective SDE for the functional $\boldsymbol{\mathcal{A}}_t$ can be expressed entirely in terms of the trajectory of the slow process in the Stratonovich prescription. Otherwise, the loops induced by irreversibility of the fast dynamics in the $\boldsymbol{X}$-trajectory generally cause $\boldsymbol{\mathcal{A}}_t$, a functional of the $\boldsymbol{X}$-trajectory, to "spin" around a modified mean trajectory in the limit. Similar results, albeit in a different and more abstract context, were also shown and discussed in (Lejay and Lyons, 2003). In the very special case when $\boldsymbol{r} = \boldsymbol{0}$, $\boldsymbol{P}$ is an identity matrix and $\boldsymbol{\mathcal{A}}_0^\epsilon = \boldsymbol{X}_0^\epsilon = \boldsymbol{0}$, we have $\boldsymbol{\mathcal{A}}_t^\epsilon = \boldsymbol{X}_t^\epsilon$ and therefore the effective description for both dynamical variable and functional coincides – see Remark 4.1 for expression of the anomalous contribution in this case.

Finally, we remark that even in the general case when $\boldsymbol{\mu}_A$ is non-zero, the effective SDE for the functional $\boldsymbol{A}_t$ can be expressed entirely in terms of the trajectory of the slow process (albeit generally not in the Stratonovich prescription), and therefore the area anomaly due to $\boldsymbol{\mathcal{A}}_t$ here is different from the entropy anomaly studied in (Bo and Celani, 2017), where new independent noise terms need to be introduced in the effective equation for the entropy production.

5. An Example: Stochastic Area as Work Functional and Its Homogenization

In this section, we are going to discuss the area anomaly phenomenon and its consequences in the context of simple Langevin systems describing the motion of a Brownian particle in magnetic field. In the next section, we will study how such phenomena manifests itself for functionals along trajectories of a wide class of multi-dimensional generalized Langevin systems approximating, in various time scale separation scenarios, that of an effective Langevin system.

Let $(\boldsymbol{q}_s)_{s\geq 0}$, $\boldsymbol{q}_s = (q_s^1, q_s^2) \in \mathbb{R}^2$, be a stochastic process. Let $\boldsymbol{q}_0 = \boldsymbol{0}$. The *stochastic area of* $(\boldsymbol{q}_s)_{s\in[0,t]}$ *on the interval* $[0,t]$ is defined as the random variable:

$$S_t = \frac{1}{2}\int_0^t (q_s^1 dq_s^2 - q_s^2 dq_s^1). \tag{51}$$

Viewing t as a continuous-time parameter, this gives rise to the area process $(S_t)_{t\geq 0}$. The above formula, with $\boldsymbol{q}_s = \boldsymbol{W}_s$ (i.e. a 2D Wiener process), is an object first introduced and studied by Lévy in (Lévy et al., 1951). His formula formally defines the area (which is random) included by the curve $\mathcal{C}_t = \{Q^1 = q_s^1, Q^2 = q_s^2,\ s\in[0,t]\}$ and its chord. Extension of this definition to the case when $\boldsymbol{q}_s \in \mathbb{R}^d$, $d > 2$, is straightforward (Malham and Wiese, 2010).

Let $(\boldsymbol{\eta}_s^\epsilon)_{s\in[0,T]}$, $\epsilon > 0$, be a family of sufficiently smooth approximations of the Wiener process $(\boldsymbol{W}_s)_{s\in[0,T]}$, where $\boldsymbol{\eta}_s^\epsilon$ converges to $\boldsymbol{W}_s$ as $\epsilon \to 0$ in a pathwise sense. A natural question is then whether or not the stochastic area of $(\boldsymbol{\eta}_s)_{s\in[0,t]}$ converges to Lévy's stochastic area as $\epsilon \to 0$. We will show that this is generally not true and discuss the consequences in the context of a physical system. We would expect similar conclusion to hold had we replaced S_t with other functionals.

Consider the motion of a charged (non-relativistic) particle undergoing Brownian motion in the presence of a magnetic field. Such motion is of interest in astrophysics, as motion from interacting charged particles produces observed light curves with interesting peculiarities (Harko and Mocanu, 2016). For simplicity, here we consider the case where the magnetic field, $\boldsymbol{B}$, points in the z-direction with a constant magnitude B and study the motion of the particle in the 2D plane perpendicular to the magnetic field[3]. In the absence of external forces and noise where the magnetic field is the dominant factor determining the motion, the particle revolves in a circular orbit with a frequency Ω, producing current loops. In this case, the magnetic force is $\boldsymbol{F}_B = \Omega \boldsymbol{V} \times \boldsymbol{e}_3 = \Omega(v^2, -v^1, 0)$, where $\boldsymbol{V} = (v^1, v^2, v^3) \in \mathbb{R}^3$ is the velocity of the charged particle, $\times$ denotes cross product and $\boldsymbol{e}_3 = (0,0,1)$. It does no work on the particle, even though the direction of motion of the particle is changed.

Taking into account this magnetic force, as well as a drag and noise term to model collisions of the charged particle with surrounding particles, the evolution of position $\boldsymbol{q}_t = (q_t^1, q_t^2)$ and velocity $\boldsymbol{v}_t = (v_t^1, v_t^2)$ of the particle on the 2D plane (assuming $\boldsymbol{q}_0 = \boldsymbol{0}$) can be described by the SDE:

$$d\boldsymbol{q}_t = \boldsymbol{v}_t dt, \tag{52}$$

$$m d\boldsymbol{v}_t = -\Omega \boldsymbol{J}\boldsymbol{v}_t dt - \boldsymbol{v}_t dt + \boldsymbol{A} d\boldsymbol{W}_t, \tag{53}$$

where $m > 0$ is the mass of the particle, $\Omega = \frac{qB}{c}$ (with q the charge of the particle, c the speed of light and B the magnitude of the constant magnetic force) is the Lamor

3. The analysis beyond this case is straightforward but involves richer physics. For instance, the charged particle may spiral in a non-trivial configuration-dependent manner when the magnetic field is position-dependent and points in arbitrary direction.

frequency (up to a multiplicative factor of $1/m$), $\boldsymbol{A} = \boldsymbol{I} + \Omega \boldsymbol{J}$ $\Bigg($with $\boldsymbol{I}$ identity matrix and $\boldsymbol{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\Bigg)$, and $\boldsymbol{W}_t$ is a Wiener process. Note that $\boldsymbol{A}$ is positive stable (but not symmetric unless $\Omega = 0$) and $\boldsymbol{A}\boldsymbol{W}_t$ is a Brownian motion with the covariance matrix $(1+\Omega^2)\boldsymbol{I}$.

Let us now suppose that the charged particle is additionally subject to an external, non-conservative force field, $\boldsymbol{f}_{nc}(t, \boldsymbol{q})$, so that the equations of motion become:

$$d\boldsymbol{q}_t = \boldsymbol{v}_t dt, \tag{54}$$
$$m d\boldsymbol{v}_t = -\Omega \boldsymbol{J} \boldsymbol{v}_t dt - \boldsymbol{v}_t dt + \boldsymbol{f}_{nc}(t, \boldsymbol{q}_t) dt + \boldsymbol{A} d\boldsymbol{W}_t. \tag{55}$$

In this case, following the approach in stochastic energetics (Sekimoto, 2010), we write the kinetic energy of the charged particle as $\mathcal{E}_t := \frac{1}{2} m v_t^2 = \mathcal{Q}_t + \mathcal{W}_t$, where the heat $\mathcal{Q}_t$ and work $\mathcal{W}_t$ satisfies:

$$d\mathcal{Q}_t = m\boldsymbol{v}_t \circ d\boldsymbol{v}_t - \boldsymbol{f}_{nc}(t, \boldsymbol{q}_t) \cdot d\boldsymbol{q}_t, \tag{56}$$
$$d\mathcal{W}_t = \boldsymbol{f}_{nc}(t, \boldsymbol{q}_t) \cdot d\boldsymbol{q}_t, \tag{57}$$

where $\circ$ denotes Stratonovich integration and $\cdot$ denotes inner product. *In the special case where $\boldsymbol{f}_{nc}(t, \boldsymbol{q}) = \frac{1}{2}(-q^2, q^1)$, the resulting work is exactly the stochastic area of the position process, i.e. $d\mathcal{W}_t = \frac{1}{2}(\boldsymbol{J}\boldsymbol{q}_t)^T d\boldsymbol{q}_t = dS_t$.* We will work with this special case in the following.

Setting $m = \epsilon$, we now consider the following rescaled family of the system (52)-(53), together with the SDEs defining the stochastic areas, S_t^ϵ, of the (rescaled) position process of the charged particle:

$$d\boldsymbol{q}_t^\epsilon = \boldsymbol{v}_t^\epsilon dt, \tag{58}$$
$$\epsilon d\boldsymbol{v}_t^\epsilon = -\boldsymbol{A}\boldsymbol{v}_t^\epsilon dt + \boldsymbol{A} d\boldsymbol{W}_t, \tag{59}$$
$$dS_t^\epsilon = \frac{1}{2}(\boldsymbol{J}\boldsymbol{q}_t^\epsilon)^T d\boldsymbol{q}_t^\epsilon. \tag{60}$$

A straightforward application of Theorem 4.1 allows us to find out whether the family of stochastic areas of $(\boldsymbol{q}_s^\epsilon)_{s \in [0,t]}$ converges to Lévy's stochastic area as $\epsilon \to 0$.

Corollary 5.1 *In the limit $\epsilon \to 0$, the family of processes $(\boldsymbol{q}_t^\epsilon, S_t^\epsilon)$ converges to $(\boldsymbol{W}_t, \bar{S}_t)$, where*

$$\bar{S}_t = S_t^{Levy} - \frac{\Omega}{2} t, \tag{61}$$

with S_t^{Levy} Lévy's stochastic area. More precisely, for all finite $T > 0$, $\sup_{t\in[0,T]} |\boldsymbol{q}_t^\epsilon - \boldsymbol{W}_t|$, $\sup_{t \in [0,T]} |S_t^\epsilon - \bar{S}_t| \to 0$ in probability[4], as $\epsilon \to 0$.

Therefore, unless $\Omega = 0$ the stochastic area (which here carries the meaning of work) of the pre-limit process does not converge to Lévy's area in the small mass limit, even though the pre-limit process converges to a Wiener process. The correct limiting area (work) includes an additional term (which we refer to as area anomaly) that depends on

4. In fact, a stronger L^p convergence result (for $p > 0$) can be obtained in this case.

the frequency at which the charged particle circles around the magnetic field, retaining in the limit the information on how the charged particle is moving under presence of the magnetic force.

The frequency Ω can be interpreted as a symmetry breaking parameter. Indeed, when $\Omega > 0$, $\boldsymbol{A}$ is not symmetric, and so the irreversibility (breaking of detailed balance) of the fast velocity process generates macroscopic current in the stationary state and induces loops in the position space whose areas are of $O(1)$ as $\epsilon \to 0$. This irreversibility can be quantified using the antisymmetric part of the Onsager matrix (Godrèche and Luck, 2018), which in this case can be computed to be $\boldsymbol{Q} = \frac{1+\Omega^2}{2}\left(\frac{\boldsymbol{A}-\boldsymbol{A}^T}{2}\right) = \Omega \boldsymbol{J}$, whose off-diagonal entries encode the area anomaly.

From a physical point of view, such phenomenon may be experimentally realized, along the line of (Argun et al., 2017), in a microscopic heat engine generating a torque via circular motion, from which work may possibly be extracted. On the other hand, rich mathematical insights on the phenomenon can be obtained using the theory of rough paths (Friz et al., 2015; Bruned et al., 2016).

6. Homogenization of GLEs and Their Functionals

In this section we explore five homogenization procedures for the GLEs and the associated functionals of interest:

(5.1) a Markovian limit;

(5.2) a limit where the small mass limit is taken after the Markovian limit in (5.1);

(5.3) the small mass limit;

(5.4) a limit where a Markovian limit is taken after the small mass limit in (5.3); and

(5.5) a joint Markovian and small mass limit.

For each procedure, we first state the problem, motivation as well as the assumptions, and then present the results. These results are obtained by applying Theorem 4.1, upon verifying the assumptions. Since the verification is straightforward we omit the proof for these results. We then discuss the commutativity of these procedures and the consequences of imposing/breaking various symmetry conditions (including fluctuation-dissipation relations).

For all these homogenization procedures, we are studying the case where the colored noise comes from two independent sources evolving on different time scales. The noise is modeled by $\boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{\xi}_t = \boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{C}_2\boldsymbol{\beta}_t = \boldsymbol{\sigma}_s(\boldsymbol{x}_t)\boldsymbol{\xi}_t^{(s)} + \boldsymbol{\sigma}_f(\boldsymbol{x}_t)\boldsymbol{\xi}_t^{(f)}$, where $\boldsymbol{\xi}_t^{(s)} = \boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)}$ and $\boldsymbol{\xi}_t^{(f)} = \boldsymbol{C}_f\boldsymbol{\beta}_t^{(f)}$, with $\boldsymbol{\beta}_t^{(s)}$ and $\boldsymbol{\beta}_t^{(f)}$ satisfying SDEs of the form (14) with different damping and diffusion constants, i.e.:

$$d\boldsymbol{\beta}_t^{(s)} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}, \tag{62}$$

$$d\boldsymbol{\beta}_t^{(f)} = -\boldsymbol{\Gamma}_f\boldsymbol{\beta}_t^{(f)}dt + \boldsymbol{\Sigma}_f d\boldsymbol{W}_t^{(d_f)}. \tag{63}$$

Here $\boldsymbol{\sigma}_f$ is a non-zero matrix, $\boldsymbol{\sigma}_s$ is a possibly zero matrix, $\boldsymbol{W}_t^{(d_s)} \in \mathbb{R}^{d_s}$ and $\boldsymbol{W}_t^{(d_f)} \in \mathbb{R}^{d_f}$ are independent Wiener processes on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ satisfying the usual conditions, and $\boldsymbol{\xi}_t^{(f)}$ denotes the part of the noise whose correlation times are much smaller than those of $\boldsymbol{\xi}_t^{(s)}$ (the superscript (s) and (f) indicate "slow" and "fast" respectively). The matrices $\boldsymbol{\Gamma}_i$ $(i = s, f)$ are positive stable and $\boldsymbol{M}_i = \boldsymbol{M}_i^T > 0$ satisfies the Lyapunov equation $\boldsymbol{\Gamma}_i \boldsymbol{M}_i + \boldsymbol{M}_i \boldsymbol{\Gamma}_i^T = \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_i^T$. We denote the covariance of $\boldsymbol{\xi}_t^{(s)}$ and $\boldsymbol{\xi}_t^{(f)}$ as $\boldsymbol{R}_s(t)$ and $\boldsymbol{R}_f(t)$ respectively.

In the case when $\boldsymbol{\sigma}_s$ is zero, $\boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{\xi}_t = \boldsymbol{\sigma}_f(\boldsymbol{x}_t)\boldsymbol{\xi}_t^{(f)}$, in which case all the noise correlation time scales are small, so taking these time scales to zero performs the full white noise limit for the GLE. Otherwise, not all noise correlation time scales are small and therefore not all of these time scales will be taken to zero, performing only a partial white noise limit for the GLE – this retains the influence of the colored noise on the system in the limit.

We assume, throughout the rest of the paper, that:

Assumption 6.1 *The matrices*

$$\boldsymbol{K}_i = \boldsymbol{C}_i \boldsymbol{\Gamma}_i^{-1} \boldsymbol{M}_i \boldsymbol{C}_i^T \quad (i = 1, f) \tag{64}$$

are non-zero and invertible (but not necessarily positive definite).

This assumption is necessary for a meaningful Markovian limit and implies that the GLE models normal diffusion (see (Lim et al., 2020) for cases where the assumption is violated). The matrix $\boldsymbol{K}_1$ is the effective damping constant and $\boldsymbol{K}_f$ the effective diffusion constant (for the fast noise process $\boldsymbol{\xi}_t^{(f)}$) in the GLE (Lim and Wehr, 2018).

In all cases, we are assuming that there are no explosions, i.e. almost surely, for every $\epsilon > 0$ there exists global unique solution to the pre-limit SDE system and also to the limiting SDE system on the time interval $[0, T]$. Other assumptions needed concern the initial conditions as well as the regularity and boundedness of the coefficients in the GLE. Note that we have chosen to work with a rather strong assumptions here – they can be relaxed in various directions at an increased cost of technicality but we choose not to pursue this here.

Assumption 6.2 *Regularity and boundedness. For $t \in \mathbb{R}^+$, $\boldsymbol{y} \in \mathbb{R}^d$, the functions $\boldsymbol{F}(t, \boldsymbol{y})$, $\boldsymbol{\sigma}_0(\boldsymbol{y})$, $\boldsymbol{\sigma}_s(\boldsymbol{y})$ and $\boldsymbol{\sigma}_f(\boldsymbol{y})$ are continuous and bounded (in t and $\boldsymbol{y}$) as well as Lipschitz in $\boldsymbol{y}$, whereas the functions $\boldsymbol{\gamma}_0(\boldsymbol{y})$, $\boldsymbol{g}(\boldsymbol{y})$, $\boldsymbol{h}(\boldsymbol{y})$, $(\boldsymbol{\gamma}_0)_{\boldsymbol{y}}(\boldsymbol{y})$, $(\boldsymbol{g})_{\boldsymbol{y}}(\boldsymbol{y})$ and $(\boldsymbol{h})_{\boldsymbol{y}}(\boldsymbol{y})$ are continuously differentiable and Lipschitz in $\boldsymbol{y}$ as well as bounded (in $\boldsymbol{y}$). Moreover, the functions $(\boldsymbol{\gamma}_0)_{\boldsymbol{yy}}(\boldsymbol{y})$, $(\boldsymbol{g})_{\boldsymbol{yy}}(\boldsymbol{y})$ and $(\boldsymbol{h})_{\boldsymbol{yy}}(\boldsymbol{y})$ are bounded for every $\boldsymbol{y} \in \mathbb{R}^d$.*

Assumption 6.3 *Initial conditions. The initial data $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^d$ are $\mathcal{F}_0$-measurable random variables independent of the σ-algebra generated by the Wiener processes $\boldsymbol{W}^{(d_s)}$ and $\boldsymbol{W}^{(d_f)}$. They are independent of ϵ and have finite moments of all orders.*

We may also need one of the following stability assumptions when studying certain procedures:

Assumption 6.4 *The matrix-valued functions $\{\boldsymbol{\gamma}_0(\boldsymbol{x}); \boldsymbol{x} \in \mathbb{R}^d\}$ are uniformly positive stable.*

Assumption 6.5 *The matrix-valued functions $\{\boldsymbol{\Gamma}(\boldsymbol{x}) = \boldsymbol{\gamma}_0(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})\boldsymbol{K}_1 \boldsymbol{h}(\boldsymbol{x}); \boldsymbol{x} \in \mathbb{R}^d\}$ are uniformly positive stable.*

6.1 A Markovian Limit

We introduce the scaling $\boldsymbol{\kappa}(t) \mapsto \frac{1}{\epsilon}\boldsymbol{\kappa}\left(\frac{t}{\epsilon}\right)$ and $\boldsymbol{R}_f(t) \mapsto \frac{1}{\epsilon}\boldsymbol{R}_f\left(\frac{t}{\epsilon}\right)$, where $\epsilon > 0$ is a small parameter, in the GLE (9) with the colored noise term $\boldsymbol{\sigma}(\boldsymbol{x}_t)\boldsymbol{\xi}_t = \boldsymbol{\sigma}_s(\boldsymbol{x}_t)\boldsymbol{\xi}_t^{(s)} + \boldsymbol{\sigma}_f(\boldsymbol{x}_t)\boldsymbol{\xi}_t^{(f)}$ as defined above. This is the limit where all the memory time scales, associated with the history-dependent damping term, and the relevant noise correlation time scales tend to zero at the same rate, and is therefore a partial Markovian limit. Our goal is to study the limit $\epsilon \to 0$ of the resulting generalized Langevin dynamics as well as of the work-like and heat-like functional of the system.

Implementing this scaling, and introducing the auxiliary process

$$\boldsymbol{y}_t^\epsilon = \int_0^t e^{-\boldsymbol{\Gamma}_1(t-s)} \boldsymbol{M}_1 \boldsymbol{C}_1^T \boldsymbol{h}(\boldsymbol{x}_s^\epsilon)\boldsymbol{v}_s^\epsilon ds, \tag{65}$$

the process $(\boldsymbol{x}_t^\epsilon, \boldsymbol{v}_t^\epsilon, \boldsymbol{y}_t^\epsilon, \boldsymbol{\beta}_t^{(f)\epsilon}, \boldsymbol{\beta}_t^{(s)\epsilon})$ satisfies the SDE system:

$$d\boldsymbol{x}_t^\epsilon = \boldsymbol{v}_t^\epsilon dt, \tag{66}$$

$$m d\boldsymbol{v}_t^\epsilon = \boldsymbol{F}(t, \boldsymbol{x}_t^\epsilon)dt - \boldsymbol{\gamma}_0(\boldsymbol{x}_t^\epsilon)\boldsymbol{v}_t^\epsilon dt - \boldsymbol{g}(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_1\boldsymbol{y}_t^\epsilon dt + \boldsymbol{\sigma}_f(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_f\boldsymbol{\beta}_t^{(f)\epsilon}dt + \boldsymbol{\sigma}_s(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)\epsilon}dt, \tag{67}$$

$$\epsilon d\boldsymbol{y}_t^\epsilon = -\boldsymbol{\Gamma}_1\boldsymbol{y}_t^\epsilon dt + \boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}(\boldsymbol{x}_t^\epsilon)\boldsymbol{v}_t^\epsilon dt, \tag{68}$$

$$\epsilon d\boldsymbol{\beta}_t^{(f)\epsilon} = -\boldsymbol{\Gamma}_f\boldsymbol{\beta}_t^{(f)\epsilon}dt + \boldsymbol{\Sigma}_f d\boldsymbol{W}_t^{(d_f)}, \tag{69}$$

$$d\boldsymbol{\beta}_t^{(s)\epsilon} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)\epsilon}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}. \tag{70}$$

The heat-like functional $\mathcal{Q}_t$ and work-like functional $\mathcal{W}_t$ satisfy the following SDEs:

$$d\mathcal{Q}_t^\epsilon = m\boldsymbol{v}_t^\epsilon \cdot d\boldsymbol{v}_t^\epsilon - \boldsymbol{F}(t, \boldsymbol{x}_t^\epsilon) \cdot d\boldsymbol{x}_t^\epsilon, \tag{71}$$

$$d\mathcal{W}_t^\epsilon = \frac{\partial U}{\partial t}dt + \boldsymbol{f}_{nc}(t, \boldsymbol{x}_t^\epsilon) \cdot d\boldsymbol{x}_t^\epsilon, \tag{72}$$

where $(\boldsymbol{x}_t^\epsilon, \boldsymbol{v}_t^\epsilon)$ solves the SDE system (66)-(70). Note that in the special case of $d = 2$ with $U := 0$, $\boldsymbol{f}_{nc}(t, \boldsymbol{x}) := \frac{1}{2}(-x^2, x^1)$, the work-like functional is simply stochastic area of the position process and the heat-like functional is the difference between the kinetic energy and this area.

The dynamics in $\boldsymbol{y}^\epsilon$ and $\boldsymbol{\beta}^{(f)\epsilon}$ are an order of magnitude faster than those in $\boldsymbol{x}^\epsilon$, $\boldsymbol{v}^\epsilon$, $\boldsymbol{\beta}^{(s)\epsilon}$, $\mathcal{Q}^\epsilon$ and $\mathcal{W}^\epsilon$, and one has the following results.

Corollary 6.1 *Under appropriate assumptions on the initial conditions and the coefficients (i.e. Assumption 6.1-6.3) of the pre-limit SDEs* (66)-(70)*, the family of processes* $(\boldsymbol{x}_t^\epsilon, \boldsymbol{v}_t^\epsilon, \boldsymbol{\beta}_t^{(s)\epsilon})$*, satisfying the SDEs* (66)-(70)*, converges, as* $\epsilon \to 0$*, to the solution* $(\boldsymbol{x}_t, \boldsymbol{v}_t, \boldsymbol{\beta}_t^{(s)})$ *of the Itô SDE system:*

$$d\boldsymbol{x}_t = \boldsymbol{v}_t dt, \tag{73}$$

$$m d\boldsymbol{v}_t = \boldsymbol{F}(t, \boldsymbol{x}_t)dt - \boldsymbol{\Gamma}(\boldsymbol{x}_t)\boldsymbol{v}_t dt + \boldsymbol{\Sigma}(\boldsymbol{x}_t)d\boldsymbol{W}_t^{(d_f)} + \boldsymbol{\sigma}_s(\boldsymbol{x}_t)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)}dt, \tag{74}$$

$$d\boldsymbol{\beta}_t^{(s)} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}, \tag{75}$$

where $\boldsymbol{\Gamma} = \boldsymbol{\gamma}_0 + \boldsymbol{g}\boldsymbol{K}_1\boldsymbol{h}$ *and* $\boldsymbol{\Sigma} = \boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}\boldsymbol{\Sigma}_f$*. The convergence is in the strong pathwise sense as before.*

Note that $\boldsymbol{\Sigma}(\boldsymbol{x}_t)\boldsymbol{W}_t^{(d_f)} = \boldsymbol{\sigma}_f(\boldsymbol{x}_t)\boldsymbol{B}_t$, where $\boldsymbol{B}_t$ is a Brownian motion with covariance $\boldsymbol{K}_f + \boldsymbol{K}_f^T$.

Corollary 6.2 *Let $\boldsymbol{\Theta}_A$ denote the antisymmetric part of the matrix $\boldsymbol{\Theta} = \boldsymbol{\sigma}_f \boldsymbol{K}_f^T \boldsymbol{\sigma}_f^T$, with $\boldsymbol{K}_f = \boldsymbol{C}_f \boldsymbol{\Gamma}_f^{-1} \boldsymbol{M}_f \boldsymbol{C}_f^T$, where $\boldsymbol{M}_f$ solves the Lyapunov equation $\boldsymbol{\Gamma}_f \boldsymbol{M}_f + \boldsymbol{M}_f \boldsymbol{\Gamma}_f^T = \boldsymbol{\Sigma}_f \boldsymbol{\Sigma}_f^T$. Under the same assumptions as in Corollary 6.1, the family of processes $(\mathcal{W}_t^\epsilon, \mathcal{Q}_t^\epsilon)$, converges, as $\epsilon \to 0$, to the solution $(\mathcal{Q}_t, \mathcal{W}_t)$ of the SDEs:*

$$d\mathcal{Q}_t = m\boldsymbol{v}_t \circ d\boldsymbol{v}_t - \boldsymbol{F}(t, \boldsymbol{x}_t) \cdot d\boldsymbol{x}_t + d\mathcal{Q}_t^{anom}, \tag{76}$$

$$d\mathcal{W}_t = \frac{\partial U}{\partial t} dt + \boldsymbol{f}_{nc}(t, \boldsymbol{x}_t) \cdot d\boldsymbol{x}_t, \tag{77}$$

where

$$d\mathcal{Q}_t^{anom} = \frac{1}{m} \boldsymbol{\nabla}_{\boldsymbol{v}} \cdot (\boldsymbol{v}_t^T \boldsymbol{\Theta}_A(\boldsymbol{x}_t)) dt, \tag{78}$$

and $(\boldsymbol{x}_t, \boldsymbol{v}_t)$ solves the SDE system (73)-(75). *The convergence is in the strong pathwise sense as before.*

Corollary 6.3 *$d\mathcal{Q}_t^{anom} = 0$ if and only if $\boldsymbol{\mu}_f = \boldsymbol{\Gamma}_f^{-1} \boldsymbol{M}_f$ (or equivalently, $\boldsymbol{K}_f$) is symmetric. In particular, a sufficient condition for $d\mathcal{Q}_t^{anom} = 0$ is when the fast process $\boldsymbol{\beta}_t^{(f)}$ satisfies the detailed balance condition.*

Note that $\boldsymbol{\Theta} = \boldsymbol{\sigma}_f \boldsymbol{C}_f \boldsymbol{M}_f \boldsymbol{\Gamma}_f^{-T} (\boldsymbol{\sigma}_f \boldsymbol{C}_f)^T = \boldsymbol{\sigma}_f \boldsymbol{K}_f^T \boldsymbol{\sigma}_f^T$, which can be related to the Onsager matrix associated to the fast dynamics. It can be shown that the matrix $\boldsymbol{\Theta}$ is, at least in the case when $\boldsymbol{\sigma}_f$ is a non-zero constant, the time integral of the correlation function of the stationary colored noise process $\tilde{\boldsymbol{\xi}}_t := \boldsymbol{\sigma}_f \boldsymbol{C}_f \boldsymbol{\beta}_t^{(f)}$, i.e. $\Theta^{ab} = \int_0^\infty E[\tilde{\xi}_t^a \tilde{\xi}_0^b] dt$, which is in general not symmetric. From Corollary 6.2, we see that, unless $\boldsymbol{\Theta}_A$ vanishes (i.e. when we are in the one-dimensional setting, or in the multi-dimensional setting with all the matrix-valued coefficients diagonal, or when the fast colored noise process admits an equilibrium stationary state), the effective evolution of the functional $\mathcal{Q}_t$ cannot be expressed solely as a Stratonovich integral over the effective trajectory. Interestingly, in the one-dimensional setting, the Stratonovich discretization is justified even if the fluctuation-dissipation relation of the second kind is violated. In the general case, whether $\boldsymbol{\Theta}_A$ vanishes or not is entirely due to the symmetry associated with the fast driving colored process, and, in particular, is independent of the details of the memory function and the slower driving noise process.

6.2 The Markovian Limit Followed by the Small Mass Limit

We rescale $m \mapsto m_0\epsilon$, where $m_0 > 0$ is a proportionality constant, in (73)-(77). The resulting SDE system then becomes:

$$d\boldsymbol{x}_t^\epsilon = \boldsymbol{v}_t^\epsilon dt, \tag{79}$$

$$\epsilon d\boldsymbol{v}_t^\epsilon = \boldsymbol{F}(t,\boldsymbol{x}_t^\epsilon)dt - \boldsymbol{\Gamma}(\boldsymbol{x}_t^\epsilon)\boldsymbol{v}_t^\epsilon dt + \boldsymbol{\Sigma}(\boldsymbol{x}_t^\epsilon)d\boldsymbol{W}_t^{(f)} + \boldsymbol{\sigma}_s(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)\epsilon}dt, \tag{80}$$

$$d\boldsymbol{\beta}_t^{(s)\epsilon} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)\epsilon}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}, \tag{81}$$

$$d\mathcal{Q}_t^\epsilon = m_0\epsilon\boldsymbol{v}_t^\epsilon \circ d\boldsymbol{v}_t^\epsilon - \boldsymbol{F}(t,\boldsymbol{x}_t^\epsilon)\cdot d\boldsymbol{x}_t^\epsilon + \frac{1}{m_0\epsilon}\boldsymbol{\nabla}_{\boldsymbol{v}^\epsilon}\cdot((\boldsymbol{v}_t^\epsilon)^T\boldsymbol{\Theta}_A(\boldsymbol{x}_t^\epsilon))dt, \tag{82}$$

$$d\mathcal{W}_t^\epsilon = \frac{\partial U}{\partial t}dt + \boldsymbol{f}_{nc}(t,\boldsymbol{x}_t^\epsilon)\cdot d\boldsymbol{x}_t^\epsilon. \tag{83}$$

We are going to study the limit $\epsilon \to 0$ of the above system. This corresponds to taking the small mass limit after the Markovian limit is taken on the GLE (9). We assume that Assumption 6.5 holds, which is crucial to ensure that the small mass limit of the system described by (73)-(75) is well defined (Lim and Wehr, 2018).

Corollary 6.4 *Under appropriate assumptions on the initial conditions and the coefficients (i.e. Assumption 6.1-6.3) of the pre-limit SDEs* (79)-(80) *and Assumption 6.5, the family of processes $\boldsymbol{x}_t^\epsilon$, satisfying the SDEs* (79)-(81)*, converges, as $\epsilon \to 0$, to the solution of the following Itô SDE:*

$$d\boldsymbol{x}_t = \boldsymbol{\Gamma}^{-1}(\boldsymbol{x}_t)(\boldsymbol{F}(t,\boldsymbol{x}_t)dt + \boldsymbol{\Sigma}(\boldsymbol{x}_t)d\boldsymbol{W}_t^{(d_f)} + \boldsymbol{\sigma}_s(\boldsymbol{x}_t)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)}dt) + \boldsymbol{H}(\boldsymbol{x}_t)dt, \tag{84}$$

$$d\boldsymbol{\beta}_t^{(s)} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_f)}, \tag{85}$$

where $\boldsymbol{\Gamma} = \gamma_0 + \boldsymbol{g}\boldsymbol{K}_1\boldsymbol{h}$, $\boldsymbol{\Sigma} = \boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}\boldsymbol{\Sigma}_f$, and $\boldsymbol{H}$ is the noise-induced drift whose expression is given by:

$$\boldsymbol{H} = \boldsymbol{\nabla}\cdot(\boldsymbol{\Gamma}^{-1}\boldsymbol{J}) - \boldsymbol{\Gamma}^{-1}\boldsymbol{\nabla}\cdot\boldsymbol{J}, \tag{86}$$

where $\boldsymbol{J}$ solves the Lyapunov equation $\boldsymbol{\Gamma}\boldsymbol{J} + \boldsymbol{J}\boldsymbol{\Gamma}^T = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T = \boldsymbol{\Theta} + \boldsymbol{\Theta}^T$, with $\boldsymbol{\Theta} = \boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{M}_f\boldsymbol{\Gamma}_f^{-T}(\boldsymbol{\sigma}_f\boldsymbol{C}_f)^T$, which was first introduced in Collorary 6.2. The convergence is in the strong pathwise sense as before.

If $\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T$ is symmetric (detailed balance), then $\boldsymbol{J} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\sigma}_f\boldsymbol{K}_f^T\boldsymbol{\sigma}_f^T$ and $\boldsymbol{H}$ simplifies to:

$$\boldsymbol{H} = \boldsymbol{\nabla}\cdot(\boldsymbol{\Gamma}^{-2}\boldsymbol{\sigma}_f\boldsymbol{K}_f^T\boldsymbol{\sigma}_f^T) - \boldsymbol{\Gamma}^{-1}\boldsymbol{\nabla}\cdot(\boldsymbol{\Gamma}^{-1}\boldsymbol{\sigma}_f\boldsymbol{K}_f^T\boldsymbol{\sigma}_f^T). \tag{87}$$

Corollary 6.5 *Assume that $\boldsymbol{\Theta}_A = \boldsymbol{0}$ (i.e. $\boldsymbol{\mu}_f = \boldsymbol{\Gamma}_f^{-1}\boldsymbol{M}_f$ is symmetric). Let $\boldsymbol{K}_A$ denote the antisymmetric part of the matrix $\boldsymbol{K} = \boldsymbol{\Gamma}^{-2}\boldsymbol{\sigma}_f\boldsymbol{K}_f^T\boldsymbol{\sigma}_f^T = \boldsymbol{\Gamma}^{-2}\boldsymbol{\Theta}$. Then, under the same assumptions as in Corollary 6.4, as $\epsilon \to 0$, the family of processes $(\mathcal{W}_t^\epsilon, \mathcal{R}_t^\epsilon)$, satisfying the SDEs* (82)-(83)*, converges to the solution of the following SDEs:*

$$d\mathcal{W}_t = \frac{\partial U}{\partial t}dt + \boldsymbol{f}_{nc}(t,\boldsymbol{x}_t)\circ d\boldsymbol{x}_t + d\mathcal{W}_t', \tag{88}$$

$$d\mathcal{R}_t = \boldsymbol{F}(t,\boldsymbol{x}_t)\circ d\boldsymbol{x}_t + d\mathcal{R}_t', \tag{89}$$

where

$$d\mathcal{W}_t' = [\boldsymbol{\nabla} \cdot (\boldsymbol{f}_{nc}^T(t,\boldsymbol{x}_t)\boldsymbol{K}_A^T(\boldsymbol{x}_t)) - \boldsymbol{f}_{nc}^T(t,\boldsymbol{x}_t)\boldsymbol{\nabla} \cdot \boldsymbol{K}_A^T(\boldsymbol{x}_t)]dt, \tag{90}$$

$$d\mathcal{R}_t' = [\boldsymbol{\nabla} \cdot (\boldsymbol{F}^T(t,\boldsymbol{x}_t)\boldsymbol{K}_A^T(\boldsymbol{x}_t)) - \boldsymbol{F}^T(t,\boldsymbol{x}_t)\boldsymbol{\nabla} \cdot \boldsymbol{K}_A^T(\boldsymbol{x}_t)]dt, \tag{91}$$

and $\boldsymbol{x}_t$ solves the SDE (84). The convergence is in the strong pathwise sense as before.

Corollary 6.6 *Suppose that the assumptions in Corollary 6.5 holds. Then $d\mathcal{W}_t' = d\mathcal{R}_t = 0$ when $\boldsymbol{\gamma}_0 = \boldsymbol{0}$, $\boldsymbol{g} \propto \boldsymbol{h}^T = \boldsymbol{\sigma}_f$ and $\boldsymbol{K}_f = \boldsymbol{K}_1$.*

One can write $\boldsymbol{K}$, using the solution $\boldsymbol{J}$ of the Lyapunov equation, explicitly as:

$$\boldsymbol{K} = (\boldsymbol{\gamma}_0 + \boldsymbol{g}\boldsymbol{K}_1\boldsymbol{h})^{-1} \int_0^\infty e^{-(\boldsymbol{\gamma}_0 + \boldsymbol{g}\boldsymbol{K}_1\boldsymbol{h})y}(\boldsymbol{\Theta} + \boldsymbol{\Theta}^T)e^{-(\boldsymbol{\gamma}_0 + \boldsymbol{g}\boldsymbol{K}_1\boldsymbol{h})^T y} dy, \tag{92}$$

where $\boldsymbol{\Theta}$ is, as we have remarked earlier, the time integral of the correlation function of the stationary colored noise process $\tilde{\boldsymbol{\xi}}_t = \boldsymbol{\sigma}_f \boldsymbol{C}_f \boldsymbol{\beta}_t^{(f)}$ with $\boldsymbol{\sigma}_f$ a constant.

We remark that if $\boldsymbol{\Theta}_A$ is non-zero, then the heat-like functional $\mathcal{Q}_t^\epsilon$ diverges in the considered limit (since $\mathcal{Q}_t^{anom} = O(1/\epsilon^2)$ as $\epsilon \to 0$). In the one-dimensional setting (where $\gamma_0 = 0$, $gh > 0$), the limit of all functionals considered is well-defined and can be expressed solely in terms of trajectory of the slow process via Stratonovich procedure. In the multi-dimensional setting, this is generally not true and, in fact, the functional might even diverge in the considered limit in the absence of symmetry of $\boldsymbol{K}$. In the case $\boldsymbol{\gamma}_0 = \boldsymbol{0}$, two sufficient condition for $d\mathcal{W}_t' = d\mathcal{R}_t = 0$ when $\boldsymbol{\gamma}_0 = \boldsymbol{0}$ are:

(1) when the fluctuation-dissipation relation holds and the driving colored noise process is an equilibrium one (in which case $\boldsymbol{K}_i = \boldsymbol{C}_i \boldsymbol{\Gamma}_i^{-1} \boldsymbol{M}_i \boldsymbol{C}_i^T$, $i = 1, f$, is symmetric) – this is the condition in Corollary 6.6;

(2) when $\boldsymbol{K}_1$ and $\boldsymbol{K}_f$ are proportional to identity (but not necessarily the same), $\boldsymbol{gh}$ is positive definite and commutes with $\boldsymbol{\sigma}_f \boldsymbol{\sigma}_f^T$.

6.3 The Small Mass Limit

We introduce the scaling $m \mapsto m_0\epsilon$ in the GLE and take the limit $\epsilon \to 0$ of the resulting equivalent rescaled SDE system:

$$d\boldsymbol{x}_t^\epsilon = \boldsymbol{v}_t^\epsilon dt, \tag{93}$$

$$m_0 \epsilon d\boldsymbol{v}_t^\epsilon = \boldsymbol{F}(t, \boldsymbol{x}_t^\epsilon)dt - \boldsymbol{\gamma}_0(\boldsymbol{x}_t^\epsilon)\boldsymbol{v}_t^\epsilon dt - \boldsymbol{g}(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_1 \boldsymbol{y}_t^\epsilon dt + \boldsymbol{\sigma}_f(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_f \boldsymbol{\beta}_t^{(f)\epsilon} dt + \boldsymbol{\sigma}_s(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_s \boldsymbol{\beta}_t^{(s)\epsilon} dt, \tag{94}$$

$$d\boldsymbol{y}_t^\epsilon = -\boldsymbol{\Gamma}_1 \boldsymbol{y}_t^\epsilon dt + \boldsymbol{M}_1 \boldsymbol{C}_1^T \boldsymbol{h}(\boldsymbol{x}_t^\epsilon)\boldsymbol{v}_t^\epsilon dt, \tag{95}$$

$$d\boldsymbol{\beta}_t^{(f)\epsilon} = -\boldsymbol{\Gamma}_f \boldsymbol{\beta}_t^{(f)\epsilon} dt + \boldsymbol{\Sigma}_f d\boldsymbol{W}_t^{(d_f)}, \tag{96}$$

$$d\boldsymbol{\beta}_t^{(s)\epsilon} = -\boldsymbol{\Gamma}_s \boldsymbol{\beta}_t^{(s)\epsilon} dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}. \tag{97}$$

The heat-like functional $\mathcal{Q}_t$ and work-like functional $\mathcal{W}_t$ satisfy the following SDEs:

$$d\mathcal{Q}_t^\epsilon = m_0 \epsilon \boldsymbol{v}_t^\epsilon \cdot d\boldsymbol{v}_t^\epsilon - \boldsymbol{F}(t, \boldsymbol{x}_t^\epsilon) \cdot d\boldsymbol{x}_t^\epsilon, \tag{98}$$

$$d\mathcal{W}_t^\epsilon = \frac{\partial U}{\partial t} dt + \boldsymbol{f}_{nc}(t, \boldsymbol{x}_t^\epsilon) \cdot d\boldsymbol{x}_t^\epsilon, \tag{99}$$

where $(\boldsymbol{x}_t^\epsilon, \boldsymbol{v}_t^\epsilon)$ solves the SDE system (93)-(97).

The dynamics in $\boldsymbol{v}^\epsilon$ are an order of magnitude faster than those in the other variables. Under a crucial assumption on the damping matrix $\boldsymbol{\gamma}_0$, the limit is well-defined and we have the following results.

Corollary 6.7 *Under appropriate assumptions on the initial conditions and the coefficients (i.e. Assumption 6.1-6.3) of the pre-limit SDEs* (93)-(97) *and Assumption 6.4, the family of processes $\boldsymbol{x}_t^\epsilon$, satisfying the SDEs* (93)-(97)*, converges, as $\epsilon \to 0$, to the solution of the following Itô SDE:*

$$d\boldsymbol{x}_t = \boldsymbol{\gamma}_0^{-1}(\boldsymbol{x}_t)[\boldsymbol{F}(t,\boldsymbol{x}_t) - \boldsymbol{g}(\boldsymbol{x}_t)\boldsymbol{C}_1\boldsymbol{y}_t + \boldsymbol{\sigma}_f(\boldsymbol{x}_t)\boldsymbol{C}_f\boldsymbol{\beta}_t^{(f)} + \boldsymbol{\sigma}_s(\boldsymbol{x}_t)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)}]dt, \tag{100}$$

$$\begin{aligned} d\boldsymbol{y}_t &= -\boldsymbol{\Gamma}_1\boldsymbol{y}_t dt \\ &\quad + \boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}(\boldsymbol{x}_t)\boldsymbol{\gamma}_0^{-1}(\boldsymbol{x}_t)[\boldsymbol{F}(t,\boldsymbol{x}_t) - \boldsymbol{g}(\boldsymbol{x}_t)\boldsymbol{C}_1\boldsymbol{y}_t + \boldsymbol{\sigma}_f(\boldsymbol{x}_t)\boldsymbol{C}_f\boldsymbol{\beta}_t^{(f)} + \boldsymbol{\sigma}_s(\boldsymbol{x}_t)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)}]dt, \end{aligned} \tag{101}$$

$$d\boldsymbol{\beta}_t^{(f)} = -\boldsymbol{\Gamma}_f\boldsymbol{\beta}_t^{(f)}dt + \boldsymbol{\Sigma}_f d\boldsymbol{W}_t^{(d_f)}, \tag{102}$$

$$d\boldsymbol{\beta}_t^{(s)} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}. \tag{103}$$

The convergence is in the strong pathwise sense as before.

Corollary 6.8 *Under the same assumptions as in Corollary 6.7, as $\epsilon \to 0$, the family of processes $(\mathcal{W}_t^\epsilon, \mathcal{R}_t^\epsilon)$, satisfying the SDEs* (98)-(99)*, converges to the solution of the following SDEs:*

$$d\mathcal{W}_t = \frac{\partial U}{\partial t}dt + \boldsymbol{f}_{nc}(t,\boldsymbol{x}_t)d\boldsymbol{x}_t, \tag{104}$$

$$d\mathcal{R}_t = \boldsymbol{F}(t,\boldsymbol{x}_t)d\boldsymbol{x}_t, \tag{105}$$

where $\boldsymbol{x}_t$ solves the SDE (100)*. The convergence is in the strong pathwise sense as before.*

Note the above functionals are uniquely defined.

6.4 The Small Mass Limit Followed by a Markovian Limit

We introduce the scaling $\boldsymbol{\kappa}(t) \mapsto \frac{1}{\epsilon}\boldsymbol{\kappa}\left(\frac{t}{\epsilon}\right)$ and $\boldsymbol{R}_f(t) \mapsto \frac{1}{\epsilon}\boldsymbol{R}_f\left(\frac{t}{\epsilon}\right)$ in the SDEs (100)-(103). This is the limit where a Markovian limit is taken after the small mass limit is performed on the GLE.

The resulting rescaled SDEs for the dynamics and functionals become:

$$d\boldsymbol{x}_t^\epsilon = \boldsymbol{\gamma}_0^{-1}(\boldsymbol{x}_t^\epsilon)[\boldsymbol{F}(t,\boldsymbol{x}_t^\epsilon) - \boldsymbol{g}(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_1\boldsymbol{y}_t^\epsilon + \boldsymbol{\sigma}_f(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_f\boldsymbol{\beta}_t^{(f)\epsilon} + \boldsymbol{\sigma}_s(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)\epsilon}]dt, \tag{106}$$

$$\begin{aligned} \epsilon d\boldsymbol{y}_t^\epsilon &= -\boldsymbol{\gamma}_1(\boldsymbol{x}_t^\epsilon)\boldsymbol{y}_t^\epsilon dt \\ &\quad + \boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}(\boldsymbol{x}_t^\epsilon)\boldsymbol{\gamma}_0^{-1}(\boldsymbol{x}_t^\epsilon)[\boldsymbol{F}(t,\boldsymbol{x}_t^\epsilon) + \boldsymbol{\sigma}_f(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_f\boldsymbol{\beta}_t^{(f)\epsilon} + \boldsymbol{\sigma}_s(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)\epsilon}]dt, \end{aligned} \tag{107}$$

$$\epsilon d\boldsymbol{\beta}_t^{(f)\epsilon} = -\boldsymbol{\Gamma}_f\boldsymbol{\beta}_t^{(f)\epsilon}dt + \boldsymbol{\Sigma}_f d\boldsymbol{W}_t^{(d_f)}, \tag{108}$$

$$d\boldsymbol{\beta}_t^{(s)\epsilon} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)\epsilon}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}, \tag{109}$$

$$d\mathcal{W}_t^\epsilon = \frac{\partial U}{\partial t}dt + \boldsymbol{f}_{nc}(t,\boldsymbol{x}_t^\epsilon)d\boldsymbol{x}_t^\epsilon, \tag{110}$$

$$d\mathcal{R}_t^\epsilon = \boldsymbol{F}(t,\boldsymbol{x}_t^\epsilon)d\boldsymbol{x}_t^\epsilon, \tag{111}$$

where $\boldsymbol{\gamma}_1 = \boldsymbol{\Gamma}_1 + \boldsymbol{M}_1 \boldsymbol{C}_1^T \boldsymbol{h} \boldsymbol{\gamma}_0^{-1} \boldsymbol{g} \boldsymbol{C}_1$.

Corollary 6.9 *Under appropriate assumptions on the initial conditions and the coefficients (i.e. Assumption 6.1-6.3) of the pre-limit SDEs* (106)-(109) *and Assumption 6.4, the family of processes $\boldsymbol{x}_t^\epsilon$, satisfying the SDEs* (106)-(109)*, converges, as $\epsilon \to 0$, to the solution of the following Itô SDE:*

$$d\boldsymbol{x}_t = \boldsymbol{\gamma}_2^{-1}(\boldsymbol{x}_t)[\boldsymbol{F}(t,\boldsymbol{x}_t) + \boldsymbol{\sigma}_s(\boldsymbol{x}_t)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)}]dt + \boldsymbol{\gamma}_2^{-1}(\boldsymbol{x}_t)\boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}\boldsymbol{\Sigma}_f d\boldsymbol{W}_t^{(d_f)} + \boldsymbol{S}(\boldsymbol{x}_t)dt, \tag{112}$$

$$d\boldsymbol{\beta}_t^{(s)} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}, \tag{113}$$

where $\boldsymbol{\gamma}_2^{-1} = \boldsymbol{\gamma}_0^{-1}(\boldsymbol{I} - \boldsymbol{g}\boldsymbol{C}_1\boldsymbol{\gamma}_1^{-1}\boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}\boldsymbol{\gamma}_0^{-1})$, $\boldsymbol{\gamma}_1 = \boldsymbol{\Gamma}_1 + \boldsymbol{M}_1 \boldsymbol{C}_1^T \boldsymbol{h} \boldsymbol{\gamma}_0^{-1} \boldsymbol{g} \boldsymbol{C}_1$, and

$$S^i = \frac{\partial R^{ij}}{\partial x^l} T^{jl}. \tag{114}$$

In the above

$$\boldsymbol{R} = -\boldsymbol{\gamma}_0^{-1}[\boldsymbol{g}\boldsymbol{C}_1\boldsymbol{\gamma}_1^{-1} \quad \boldsymbol{g}\boldsymbol{C}_1\boldsymbol{\gamma}_1^{-1}(\boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}\boldsymbol{\gamma}_0^{-1}\boldsymbol{\sigma}_f\boldsymbol{C}_f)\boldsymbol{\Gamma}_f^{-1} - \boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}], \tag{115}$$

$$\boldsymbol{T} = (-\boldsymbol{J}_{11}\boldsymbol{C}_1^T\boldsymbol{g}^T\boldsymbol{\gamma}_0^{-T} + \boldsymbol{J}_{12}\boldsymbol{C}_f^T\boldsymbol{\sigma}_f^T\boldsymbol{\gamma}_0^{-T}, -\boldsymbol{J}_{12}^T\boldsymbol{C}_1^T\boldsymbol{g}^T\boldsymbol{\gamma}_0^{-T} + \boldsymbol{M}_f\boldsymbol{C}_f^T\boldsymbol{\sigma}_f^T\boldsymbol{\gamma}_0^{-T}), \tag{116}$$

where $\boldsymbol{J}_{11}$ and $\boldsymbol{J}_{12}$ solve the matrix equations:

$$\boldsymbol{\gamma}_1\boldsymbol{J}_{12} + \boldsymbol{J}_{12}\boldsymbol{\Gamma}_f^T = \boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}\boldsymbol{\gamma}_0^{-1}\boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{M}_f, \tag{117}$$

$$\boldsymbol{\gamma}_1\boldsymbol{J}_{11} + \boldsymbol{J}_{11}\boldsymbol{\gamma}_1^T = \boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}\boldsymbol{\gamma}_0^{-1}\boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{J}_{12}^T + \boldsymbol{J}_{12}(\boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}\boldsymbol{\gamma}_0^{-1}\boldsymbol{\sigma}_f\boldsymbol{C}_f)^T. \tag{118}$$

The convergence is in the strong pathwise sense as before.

Corollary 6.10 *Under the same assumptions as in Corollary 6.9, as $\epsilon \to 0$, the family of processes $(\mathcal{W}_t^\epsilon, \mathcal{R}_t^\epsilon)$, satisfying the SDEs* (111)-(110)*, converges to the solution of the following SDEs:*

$$d\mathcal{W}_t = \frac{\partial U}{\partial t}dt + \boldsymbol{f}_{nc}(t,\boldsymbol{x}_t) \circ d\boldsymbol{x}_t + d\mathcal{W}_t', \tag{119}$$

$$d\mathcal{R}_t = \boldsymbol{F}(t,\boldsymbol{x}_t) \circ d\boldsymbol{x}_t + d\mathcal{R}_t', \tag{120}$$

$$d\mathcal{W}_t' = [\boldsymbol{\nabla} \cdot (\boldsymbol{f}_{nc}^T(t,\boldsymbol{x}_t)\boldsymbol{\Phi}(\boldsymbol{x}_t)\boldsymbol{\mu}_A^T(\boldsymbol{x}_t)\boldsymbol{\Phi}^T(\boldsymbol{x}_t)) - \boldsymbol{f}_{nc}^T(t,\boldsymbol{x}_t)\boldsymbol{\nabla} \cdot (\boldsymbol{\Phi}(\boldsymbol{x}_t)\boldsymbol{\mu}_A^T(\boldsymbol{x}_t)\boldsymbol{\Phi}^T(\boldsymbol{x}_t))]dt, \tag{121}$$

$$d\mathcal{R}_t' = [\boldsymbol{\nabla} \cdot (\boldsymbol{F}^T(t,\boldsymbol{x}_t)\boldsymbol{\Phi}(\boldsymbol{x}_t)\boldsymbol{\mu}_A^T(\boldsymbol{x}_t)\boldsymbol{\Phi}^T(\boldsymbol{x}_t)) - \boldsymbol{F}^T(t,\boldsymbol{x}_t)\boldsymbol{\nabla} \cdot (\boldsymbol{\Phi}(\boldsymbol{x}_t)\boldsymbol{\mu}_A^T(\boldsymbol{x}_t)\boldsymbol{\Phi}^T(\boldsymbol{x}_t))]dt, \tag{122}$$

where $\boldsymbol{\Phi} = \boldsymbol{\gamma}_0^{-1}[-\boldsymbol{g}\boldsymbol{C}_1 \quad \boldsymbol{\sigma}_f\boldsymbol{C}_f]$, $\boldsymbol{\mu}_A$ is the antisymmetric part of the matrix

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\gamma}_1^{-1}(\boldsymbol{J}_{11} + \boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}\boldsymbol{\gamma}_0^{-1}\boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}\boldsymbol{J}_{12}^T) & \boldsymbol{\gamma}_1^{-1}(\boldsymbol{J}_{12} + \boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}\boldsymbol{\gamma}_0^{-1}\boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}\boldsymbol{M}_f) \\ \boldsymbol{\Gamma}_f^{-1}\boldsymbol{J}_{12}^T & \boldsymbol{\Gamma}_f^{-1}\boldsymbol{M}_f \end{bmatrix}, \tag{123}$$

with $\boldsymbol{J}_{11}$ and $\boldsymbol{J}_{12}$ satisfying (117)-(118)*, and $\boldsymbol{x}_t$ solves the SDE* (112)*. The convergence is in the strong pathwise sense as before.*

6.5 A Joint Markovian and Small Mass Limit

We introduce the scaling $\boldsymbol{\kappa}(t) \mapsto \frac{1}{\epsilon}\boldsymbol{\kappa}\left(\frac{t}{\epsilon}\right)$ and $\boldsymbol{R}_f(t) \mapsto \frac{1}{\epsilon}\boldsymbol{R}_f\left(\frac{t}{\epsilon}\right)$, $m \mapsto m_0\epsilon$ in the GLE (9). This is the limit where the inertial time scale, the memory time scale and some noise correlation time scales of the system tend to zero at the same rate. This will provide a further coarse-grained model compared to the Markovian limit and therefore more information will be lost in the limit. We remark that the small mass limit of our GLE is generally not well-defined (unless $\boldsymbol{\gamma}_0 > \boldsymbol{0}$) and leads to the interesting phenomenon of anomalous gap of the particle's mean-squared displacement (McKinley et al., 2009; Indei et al., 2012; Córdoba et al., 2012).

Introducing the auxiliary variable $\boldsymbol{y}_t$ as before, the resulting rescaled GLE can then be studied as the following SDE system for the Markov process $(\boldsymbol{x}_t^\epsilon, \boldsymbol{v}_t^\epsilon, \boldsymbol{y}_t^\epsilon, \boldsymbol{\beta}_t^{(f)\epsilon}, \boldsymbol{\beta}_t^{(s)\epsilon})$:

$$d\boldsymbol{x}_t^\epsilon = \boldsymbol{v}_t^\epsilon dt, \tag{124}$$

$$m_0\epsilon d\boldsymbol{v}_t^\epsilon = \boldsymbol{F}(t,\boldsymbol{x}_t^\epsilon)dt - \boldsymbol{\gamma}_0(\boldsymbol{x}_t^\epsilon)\boldsymbol{v}_t^\epsilon dt - \boldsymbol{g}(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_1\boldsymbol{y}_t^\epsilon dt + \boldsymbol{\sigma}_f(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_f\boldsymbol{\beta}_t^{(f)\epsilon}dt + \boldsymbol{\sigma}_s(\boldsymbol{x}_t^\epsilon)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)\epsilon}dt, \tag{125}$$

$$\epsilon d\boldsymbol{y}_t^\epsilon = -\boldsymbol{\Gamma}_1\boldsymbol{y}_t^\epsilon dt + \boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}(\boldsymbol{x}_t^\epsilon)\boldsymbol{v}_t^\epsilon dt, \tag{126}$$

$$\epsilon d\boldsymbol{\beta}_t^{(f)\epsilon} = -\boldsymbol{\Gamma}_f\boldsymbol{\beta}_t^{(f)\epsilon}dt + \boldsymbol{\Sigma}_f d\boldsymbol{W}_t^{(d_f)}, \tag{127}$$

$$d\boldsymbol{\beta}_t^{(s)\epsilon} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)\epsilon}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}. \tag{128}$$

The heat $\mathcal{Q}_t^\epsilon$ and work $\mathcal{W}_t^\epsilon$ satisfy the following SDEs:

$$d\mathcal{Q}_t^\epsilon = m_0\epsilon\boldsymbol{v}_t^\epsilon \cdot d\boldsymbol{v}_t^\epsilon - \boldsymbol{F}(t,\boldsymbol{x}_t^\epsilon)\cdot d\boldsymbol{x}_t^\epsilon, \tag{129}$$

$$d\mathcal{W}_t^\epsilon = \frac{\partial U}{\partial t}dt + \boldsymbol{f}_{nc}(t,\boldsymbol{x}_t^\epsilon)\cdot d\boldsymbol{x}_t^\epsilon, \tag{130}$$

where $(\boldsymbol{x}_t^\epsilon, \boldsymbol{v}_t^\epsilon)$ solves the SDE system (124)-(128).

The dynamics in $\boldsymbol{v}^\epsilon$, $\boldsymbol{y}^\epsilon$ and $\boldsymbol{\beta}^{(f)\epsilon}$ are an order of magnitude faster than those in $\boldsymbol{x}^\epsilon$, $\boldsymbol{\beta}^{(s)\epsilon}$, $\mathcal{Q}^\epsilon$ and $\mathcal{W}^\epsilon$.

Consider the following system of five matrix equations for $\boldsymbol{J}_{11} = \boldsymbol{J}_{11}^T$, $\boldsymbol{J}_{21} = \boldsymbol{J}_{12}^T$ and $\boldsymbol{J}_{31} = \boldsymbol{J}_{13}^T$ (c.f. (Lim and Wehr, 2018)):

$$\boldsymbol{\gamma}_0\boldsymbol{J}_{11} + \boldsymbol{J}_{11}\boldsymbol{\gamma}_0^T + \boldsymbol{g}\boldsymbol{C}_1\boldsymbol{J}_{12}^T + \boldsymbol{J}_{12}\boldsymbol{C}_1^T\boldsymbol{g}^T = \boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{J}_{13}^T + \boldsymbol{J}_{13}\boldsymbol{C}_f^T\boldsymbol{\sigma}_f^T, \tag{131}$$

$$m_0\boldsymbol{J}_{11}\boldsymbol{h}^T\boldsymbol{C}_1\boldsymbol{M}_1 + \boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{J}_{23}^T = \boldsymbol{g}\boldsymbol{C}_1\boldsymbol{J}_{22} + m_0\boldsymbol{J}_{12}\boldsymbol{\Gamma}_1^T + \boldsymbol{\gamma}_0\boldsymbol{J}_{12}, \tag{132}$$

$$\boldsymbol{\gamma}_0\boldsymbol{J}_{13} + \boldsymbol{g}\boldsymbol{C}_1\boldsymbol{J}_{23} + m_0\boldsymbol{J}_{13}\boldsymbol{\Gamma}_f^T = \boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{M}_f, \tag{133}$$

$$\boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}\boldsymbol{J}_{12} + \boldsymbol{J}_{12}^T\boldsymbol{h}^T\boldsymbol{C}_1\boldsymbol{M}_1 = \boldsymbol{\Gamma}_1\boldsymbol{J}_{22} + \boldsymbol{J}_{22}\boldsymbol{\Gamma}_1^T, \tag{134}$$

$$\boldsymbol{M}_1\boldsymbol{C}_1^T\boldsymbol{h}\boldsymbol{J}_{13} = \boldsymbol{\Gamma}_1\boldsymbol{J}_{23} + \boldsymbol{J}_{23}\boldsymbol{\Gamma}_f^T. \tag{135}$$

We write $\mathcal{Q}_t^\epsilon = \frac{m_0}{2}\epsilon|\boldsymbol{v}_t^\epsilon|^2 - \frac{m_0}{2}\epsilon|\boldsymbol{v}_0^\epsilon|^2 - \mathcal{R}_t^\epsilon$, where $\mathcal{R}_t^\epsilon = \int_0^t \boldsymbol{F}(s,\boldsymbol{x}_s^\epsilon)\cdot d\boldsymbol{x}_s^\epsilon$. We expect that as $\epsilon \to 0$, the kinetic energy terms are of $O(1)$ and they tend to $\overline{\frac{m_0}{2}|\boldsymbol{v}_t|^2 - \frac{m_0}{2}|\boldsymbol{v}_0|^2}$, where the overline denotes average with respect to the invariant density of the stationary fast process (at a given slow ones), which is mean zero Gaussian with covariance matrix $\boldsymbol{J}_{11}$. Therefore, to study the asymptotic behavior of $\mathcal{Q}_t^\epsilon$ in the considered limit, it suffices to investigate the asymptotic behavior of $\mathcal{R}_t^\epsilon$.

One then has the following results.

Corollary 6.11 *The family of processes $\boldsymbol{x}_t^\epsilon$, satisfying the SDEs* (124)-(128), *converges, as $\epsilon \to 0$, to the solution of the following Itô SDE:*

$$d\boldsymbol{x}_t = \boldsymbol{\Gamma}^{-1}(\boldsymbol{x}_t)(\boldsymbol{F}(t,\boldsymbol{x}_t) + \boldsymbol{\sigma}_s(\boldsymbol{x}_t)\boldsymbol{C}_s\boldsymbol{\beta}_t^{(s)})dt + \boldsymbol{S}(\boldsymbol{x}_t)dt + \boldsymbol{\Gamma}^{-1}(\boldsymbol{x}_t)\boldsymbol{\Sigma}(\boldsymbol{x}_t)d\boldsymbol{W}_t^{(d_f)}, \quad (136)$$

$$d\boldsymbol{\beta}_t^{(s)} = -\boldsymbol{\Gamma}_s\boldsymbol{\beta}_t^{(s)}dt + \boldsymbol{\Sigma}_s d\boldsymbol{W}_t^{(d_s)}, \quad (137)$$

where $\boldsymbol{\Gamma} = \boldsymbol{\gamma}_0 + \boldsymbol{g}\boldsymbol{K}_1\boldsymbol{h}$, $\boldsymbol{\Sigma} = \boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}\boldsymbol{\Sigma}_f$, and $\boldsymbol{S}$ is the noise-induced drift whose expression is given by:

$$\begin{aligned}\boldsymbol{S} &= \boldsymbol{\nabla}\cdot\Big(\boldsymbol{\Gamma}^{-1}(m_0\boldsymbol{J}_{11} - \boldsymbol{g}(\boldsymbol{C}_1\boldsymbol{\Gamma}_1^{-1}\boldsymbol{J}_{21})^T + \boldsymbol{\sigma}_f(\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}\boldsymbol{J}_{31})^T)\Big) \\ &\quad + \boldsymbol{\Gamma}^{-1}\Big(\boldsymbol{g}\boldsymbol{\nabla}\cdot((\boldsymbol{C}_1\boldsymbol{\Gamma}_1^{-1}\boldsymbol{J}_{21})^T) - \boldsymbol{\sigma}_f\boldsymbol{\nabla}\cdot((\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}\boldsymbol{J}_{31})^T) - m_0\boldsymbol{\nabla}\cdot\boldsymbol{J}_{11}\Big), \quad (138)\end{aligned}$$

where the $\boldsymbol{J}_{ij}$ solve the system of matrix equations (131)-(135). *The convergence is in the strong pathwise sense as before.*

The presence of the noise-induced drift $\boldsymbol{S}$, due to the state-dependence of the coefficients $\boldsymbol{g}$, $\boldsymbol{h}$ and $\boldsymbol{\sigma}_f$, implies that the elimination of the fast degrees of freedom needs to be done carefully and naive procedure could lead to inconsistent result.

Corollary 6.12 *Let $\boldsymbol{\lambda}_A$ denote the antisymmetric part of $\boldsymbol{\lambda} = -m_0\boldsymbol{\Gamma}^{-1}\boldsymbol{J}_{11} + \boldsymbol{\Gamma}^{-1}\boldsymbol{g}\boldsymbol{C}_1\boldsymbol{\Gamma}_1^{-1}\boldsymbol{J}_{21} - \boldsymbol{\Gamma}^{-1}\boldsymbol{\sigma}_f\boldsymbol{C}_f\boldsymbol{\Gamma}_f^{-1}\boldsymbol{J}_{31}$, where the $\boldsymbol{J}_{ij}$ solve the system of matrix equation* (131)-(135).

The family of processes $(\mathcal{W}_t^\epsilon, \mathcal{R}_t^\epsilon)$ converges, as $\epsilon \to 0$, to the solution of the following SDEs:

$$d\mathcal{W}_t = \frac{\partial U}{\partial t}dt + \boldsymbol{f}_{nc}(t,\boldsymbol{x}_t)\circ d\boldsymbol{x}_t + dW_t^{anom}, \quad (139)$$

$$d\mathcal{R}_t = \boldsymbol{F}(t,\boldsymbol{x}_t)\circ d\boldsymbol{x}_t + d\mathcal{R}_t^{anom}, \quad (140)$$

where

$$d\mathcal{W}_t^{anom} = [\boldsymbol{\nabla}\cdot(\boldsymbol{f}_{nc}^T(t,\boldsymbol{x}_t)\boldsymbol{\lambda}_A(\boldsymbol{x}_t)) - \boldsymbol{f}_{nc}^T(t,\boldsymbol{x}_t)\boldsymbol{\nabla}\cdot\boldsymbol{\lambda}_A(\boldsymbol{x}_t)]dt, \quad (141)$$

$$d\mathcal{R}_t^{anom} = [\boldsymbol{\nabla}\cdot(\boldsymbol{F}^T(t,\boldsymbol{x}_t)\boldsymbol{\lambda}_A(\boldsymbol{x}_t)) - \boldsymbol{F}^T(t,\boldsymbol{x}_t)\boldsymbol{\nabla}\cdot\boldsymbol{\lambda}_A(\boldsymbol{x}_t)]dt, \quad (142)$$

and $\boldsymbol{x}_t$ solves the SDE (136)-(137). *The convergence is in the strong pathwise sense as before.*

Corollary 6.13 *$d\mathcal{W}_t^{anom} = d\mathcal{R}_t^{anom} = 0$ when one of the following conditions holds:*

(i) $\boldsymbol{\gamma}_0$, $\boldsymbol{g}$, $\boldsymbol{h}$, $\boldsymbol{\sigma}_f$, $\boldsymbol{C}_i$, $\boldsymbol{M}_i$, $\boldsymbol{\Gamma}_i$ $(i = 1, f)$ are diagonal;

(ii) $\boldsymbol{\gamma}_0$ and $\boldsymbol{\sigma}_0$ are zero, the fluctuation-dissipation relation of the second kind holds, and $\boldsymbol{\Gamma}^{-1}\boldsymbol{\sigma}_f\boldsymbol{K}_f^T\boldsymbol{\sigma}_f^T$ is symmetric (detailed balance).

In contrast to the Markovian limit case, it is generally not possible to express both the work and heat functional in terms of trajectory of the effective slow process without additional drift terms. This is possible for the work functional in the case where $\boldsymbol{f}_{nc}$ is independent of position. Also, the matrix $\boldsymbol{\lambda}$ loses the meaning as the time integral of the correlation function of a physical noise process.

We next discuss the above results in the case $\boldsymbol{\gamma}_0 = \mathbf{0}$. The limiting expression for $\mathcal{W}_t$ and $\mathcal{R}_t$ can be expressed in terms of trajectory of the slow process via Stratonovich discretization if and only if $\boldsymbol{\lambda}_A$ vanishes. In the one-dimensional setting, the Stratonovich procedure is justified even if the fluctuation-dissipation relation is violated. However, in contrast to the results obtained for the Markovian limit, a stricter condition is needed for $\boldsymbol{\lambda}_A$ to vanish in the general multi-dimensional case. Whether $\boldsymbol{\lambda}_A$ vanishes or not is not entirely attributed to the symmetry associated with the noise term, but it also depends on the properties of the memory function as well as the coefficients $\boldsymbol{g}$, $\boldsymbol{h}$ and $\boldsymbol{\sigma}_f$. The unifying message in the above discussion is that, in the multidimensional setting, higher level of coarse-graining or model reduction often leads to justification of use of Stratonovich procedure in defining thermodynamic functionals using equations for the effective dynamics for a smaller, more restricted class of systems. In the special one-dimensional setting, the Stratonovich procedure is always justified.

6.6 Discussions

We have considered the joint Markovian and small mass limit of the GLE (Procedure (5.5)) in the previous subsection, as well as the procedure where the small mass limit is taken after the Markovian limit is taken here (Procedure (5.2)). A natural question is how do the effective equations obtained via these two limiting procedures compare. To allow the comparison, we assume that $\boldsymbol{\gamma}_0 = \mathbf{0}$ and the detailed balance condition on the fast process holds, i.e. $\boldsymbol{\Theta}_A = \mathbf{0}$. First, note that the solution of (136) coincides, in law, with that of (84) if and only if the noise-induced drifts $\boldsymbol{S}$ (in (138)) and $\boldsymbol{H}$ (in (87)) coincide. A sufficient condition for this is when the fluctuation-dissipation relation of the second kind holds (Lim and Wehr, 2018). Second, the work functionals, satisfying (139) and (88) respectively, coincide, if in addition, $\mathcal{W}_t^{anom} = \mathcal{W}_t'$, i.e. if and only if $\boldsymbol{\lambda}_A = \boldsymbol{K}_A^T$. This occurs, for instance, in the very special case of one dimensions where the fluctuation-dissipation relation of the second kind, i.e. $\boldsymbol{g} = \boldsymbol{h}^T = \boldsymbol{\sigma}_f$ and $\boldsymbol{R}_f(t) = \boldsymbol{\kappa}(t)$ holds.

Similar, albeit slightly more tedious, comparison can also be performed for the results obtained via Procedure (5.5) and those via Procedure (5.4). In general, convergence of the dynamical and functional paths depends on regularity of the approximating sequence. Different homogenization procedures give rise to approximating sequences of different regularity and thus different limiting behavior where different forms of area anomaly appear, so the commutativity of the procedures is not guaranteed unless one restricts to special cases – these cases invoke symmetry in the form of a detailed balance as well as the relation between dissipation and fluctuation driving the fast dynamics.

7. Conclusions

We have explored and performed various multiple time scale analysis (homogenization) for a class of generalized Langevin dynamics together with the stochastic processes describing the heat-like and work-like functionals in stochastic thermodynamics. We have addressed and discussed the important problem of justifying the use of Stratonovich convention in the definition of these functionals in the situations where there exists wide separation of time scales of various levels in the systems. We find that, unless certain symmetry is present in the GLE system, it is generally not possible to express the effective evolution of these functionals solely in terms of trajectory of the effective process describing the system dynamics via the standard Stratonovich convention, and additional information of the full process is needed to do so.

Depending on the level of coarse graining, one needs to impose appropriate symmetry conditions in such a way that the area anomaly, encoded by the antisymmetric part of the Onsager matrix associated with the fast dynamics, vanishes, in order to make this possible. In the case where these functionals are thermodynamic, the absence of these symmetry conditions gives rise to anomalous thermodynamics in the homogenized systems. Our results can be applied to concrete physical systems, including the ones described in Appendix A, in various time scale separation scenarios.

Acknowledgements. The author gratefully acknowledges the support provided by the NORDITA Fellowship during the year 2018-2020. He is also grateful to Peter Hänggi, Ralf Eichhorn and Stefano Bo for critical reading of the manuscript and suggestions for improving an earlier version of the manuscript.

Appendix A. Examples of GLE Systems in Nonequilibrium Statistical Mechanics

In this section, we provide three examples of physical system that can be modeled by (special cases of) the GLEs studied in this paper.

Example A.1 A Brownian particle in a temperature gradient. *We consider a Brownian particle immersed in a nonequilibrium heat bath where a temperature gradient is present. For this system, the temperature of the heat bath varies with the position of the particle and a generalized fluctuation-dissipation relation holds. We model the system by the GLE defined in Section 2, with* $\boldsymbol{\gamma}_0 = \mathbf{0}$, $\boldsymbol{\sigma}_0 = \mathbf{0}$, $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x}) = \sqrt{\gamma(\boldsymbol{x})}\boldsymbol{I}$ *and* $\boldsymbol{\sigma}(\boldsymbol{x}) = \sqrt{k_B T(\boldsymbol{x})\gamma(\boldsymbol{x})}\boldsymbol{I}$, *where* $\boldsymbol{x} \in \mathbb{R}^d$ $(d = 1, 2, 3)$, $\gamma > 0$ *is a scalar function,* k_B *is the Boltzmann constant,* T *is the state-dependent temperature of the bath and* $\boldsymbol{I}$ *is the identity matrix. The resulting GLE is then:*

$$d\boldsymbol{x}_t = \boldsymbol{v}_t dt, \tag{143}$$

$$m d\boldsymbol{v}_t = \boldsymbol{F}(t, \boldsymbol{x}_t)dt - \sqrt{\gamma(\boldsymbol{x}_t)}\left(\int_0^t \boldsymbol{\kappa}(t-s)\sqrt{\gamma(\boldsymbol{x}_s)}\boldsymbol{v}_s ds\right)dt + \sqrt{k_B T(\boldsymbol{x}_t)\gamma(\boldsymbol{x}_t)}\boldsymbol{\xi}_t dt, \tag{144}$$

where $\boldsymbol{\xi}_t \in \mathbb{R}^d$ *is a mean-zero, stationary Gaussian colored noise with covariance function equals to* $\boldsymbol{\kappa}(t)$. *The memory function and the colored noise are defined in a similar way as before:* $\boldsymbol{\kappa}(t) = \boldsymbol{C}_1 e^{-\boldsymbol{\Gamma}_1 t}\boldsymbol{M}_1\boldsymbol{C}_1^T$ *and* $\boldsymbol{\xi}_t = \boldsymbol{C}_1\boldsymbol{\beta}_t$, *where* $d\boldsymbol{\beta}_t = -\boldsymbol{\Gamma}_1\boldsymbol{\beta}_t dt + \boldsymbol{\Sigma}_1 d\boldsymbol{W}_t$. *The above model has been used to study the phenomena of thermophoresis in (Lim and Wehr, 2018) (see also the discussions and references related to the GLE* (143)-(144) *there).*

Example A.2 **Active matter systems with spatially inhomogeneous activity.** *We consider a small system in an equilibrium (passive) heat bath at the constant temperature*

T subject to an external force field described by $\boldsymbol{F}(t,\boldsymbol{x}) = -\boldsymbol{\nabla}_{\boldsymbol{x}} U(t,\boldsymbol{x}) + \boldsymbol{f}_{nc}(t,\boldsymbol{x})$ and an active force field described by $\boldsymbol{\sigma}_a(\boldsymbol{x})\boldsymbol{\eta}$, where $\boldsymbol{x} \in \mathbb{R}^d$ ($d = 1,2,3$), U is the potential, $\boldsymbol{f}_{nc}$ is a non-conservative force field, $\boldsymbol{\sigma}_a : \mathbb{R}^d \to \mathbb{R}^{d\times a}$ is a state-dependent coefficient, and $\boldsymbol{\eta} \in \mathbb{R}^a$ is a mean-zero stationary Ornstein-Uhlenbeck process. We model this system by the GLE in Section 2 with $\boldsymbol{\gamma}_0 = \boldsymbol{0}$, $\boldsymbol{\sigma}_0 = \boldsymbol{0}$, $\boldsymbol{g} = \boldsymbol{h}^T = \boldsymbol{\sigma}_p \in \mathbb{R}^{d\times d_1}$ (constant matrix), $\boldsymbol{\sigma}(\boldsymbol{x}) = [\sqrt{k_B T}\boldsymbol{\sigma}_p \;\; \boldsymbol{\sigma}_a(\boldsymbol{x})] \in \mathbb{R}^{d\times(d_1+a)}$, $\boldsymbol{\xi}_t = \boldsymbol{C}_2\boldsymbol{\beta}_t$, with $\boldsymbol{C}_2 = \boldsymbol{I}$, $\boldsymbol{\beta}_t = (\boldsymbol{\zeta}_t, \boldsymbol{\eta}_t) \in \mathbb{R}^{d_1+a}$, $\boldsymbol{\zeta}_t = \boldsymbol{C}_p\boldsymbol{\theta}_t$. More precisely:

$$d\boldsymbol{x}_t = \boldsymbol{v}_t dt, \tag{145}$$

$$m d\boldsymbol{v}_t = \boldsymbol{F}(t,\boldsymbol{x}_t)dt - \boldsymbol{\sigma}_p\left(\int_0^t \boldsymbol{\kappa}(t-s)\boldsymbol{\sigma}_p^T \boldsymbol{v}_s ds\right)dt + \sqrt{k_B T}\boldsymbol{\sigma}_p\boldsymbol{C}_p\boldsymbol{\theta}_t dt + \boldsymbol{\sigma}_a(\boldsymbol{x}_t)\boldsymbol{\eta}_t dt, \tag{146}$$

$$d\boldsymbol{\theta}_t = -\boldsymbol{\Gamma}_p\boldsymbol{\theta}_t dt + \boldsymbol{\Sigma}_p d\boldsymbol{W}_t, \tag{147}$$

$$d\boldsymbol{\eta}_t = -\boldsymbol{\Gamma}_a\boldsymbol{\eta}_t dt + \boldsymbol{\Sigma}_a d\boldsymbol{U}_t, \tag{148}$$

where $\boldsymbol{\zeta}_t$ is a mean-zero, stationary Gaussian colored noise with covariance function equals to $\boldsymbol{\kappa}(t) = \boldsymbol{C}_p e^{-\boldsymbol{\Gamma}_p t}\boldsymbol{M}_p\boldsymbol{C}_p^T \in \mathbb{R}^{d_1\times d_1}$, and $\boldsymbol{U}_t$, $\boldsymbol{W}_t$ are independent Wiener processes.

In the absence of $\boldsymbol{\sigma}_a(\boldsymbol{x}_t)\boldsymbol{\eta}_t$, the model can be derived from a microscopic Hamiltonian model describing a particle interacting with an equilibrium heat bath at temperature T. Therefore, the above model describes a system driven out of equilibrium by the active force $\boldsymbol{\sigma}_a(\boldsymbol{x}_t)\boldsymbol{\eta}_t$. The above model can be viewed as a closely related variant of the ones studied in (Leyman et al., 2018). In the joint limit where $\boldsymbol{\kappa}(t)$ tends to a Dirac delta function (memoryless limit), $\boldsymbol{\zeta}_t$ tends to a white noise (white noise limit) and $m \to 0$ (small mass limit), we recover the active Ornstein-Uhlenbeck model for active matter systems studied in (Dabelow et al., 2019) but with inhomogeneous activity due to the state-dependence of $\boldsymbol{\eta}_a$ here.

***Example A.3* A charged particle in a spatially inhomogeneous magnetic field.** *We consider an electrically charged particle of charge q in an equilibrium homogeneous heat bath. It is subject to a position-dependent magnetic field $\boldsymbol{B}(\boldsymbol{x})$ ($\boldsymbol{x} \in \mathbb{R}^3$) (Vuijk et al., 2019) and time-dependent force field, $\boldsymbol{F} = -\boldsymbol{\nabla}_{\boldsymbol{x}} U(t,\boldsymbol{x}) + q\boldsymbol{E}(t,\boldsymbol{x})$, consisting of forces from conservative potential and electric field. Assuming that the magnetic field is pointing along the unit vector $\boldsymbol{n}$ and $B(\boldsymbol{x})$ is the magnitude (i.e. $\boldsymbol{B}(\boldsymbol{x}) = B(\boldsymbol{x})\boldsymbol{n}$), the Lorentz force $q\boldsymbol{v}_t \times \boldsymbol{B}(\boldsymbol{x}_t)$ can be written as $qB(\boldsymbol{x}_t)\boldsymbol{Z}\boldsymbol{v}_t$, where $\boldsymbol{Z}$ is a matrix with elements given by $Z_{ij} = -\epsilon_{ijk}n_k$, where ϵ_{ijk} is the totally antisymmetric Levi-Civita symbol in 3D and n_k is the kth component of $\boldsymbol{n}$. This system can be described by the GLE with $\boldsymbol{\gamma}_0(\boldsymbol{x}) = -qB(\boldsymbol{x})\boldsymbol{Z}$, $\boldsymbol{\sigma}_0 = \boldsymbol{0}$, $\boldsymbol{g} = \boldsymbol{h}^T = \boldsymbol{\sigma}_b$, $\boldsymbol{\sigma} = \sqrt{k_B T}\boldsymbol{\sigma}_b$, $\boldsymbol{\xi}_t$ is the same colored noise as introduced in Section 2 but with its covariance function equals to $\boldsymbol{\kappa}(t)$:*

$$d\boldsymbol{x}_t = \boldsymbol{v}_t dt, \tag{149}$$

$$m d\boldsymbol{v}_t = \boldsymbol{F}(t,\boldsymbol{x}_t)dt - \boldsymbol{\sigma}_b\left(\int_0^t \boldsymbol{\kappa}(t-s)\boldsymbol{\sigma}_b^T \boldsymbol{v}_s ds\right)dt + qB(\boldsymbol{x}_t)\boldsymbol{Z}\boldsymbol{v}_t dt + \sqrt{k_B T}\boldsymbol{\sigma}_b\boldsymbol{\xi}_t dt. \tag{150}$$

In the Markovian limit (i.e. joint memoryless and white noise limit), one obtain a Langevin-Kramers equation with a state-dependent damping term (with a positive stable but not positive definite effective "damping" matrix) and an additive white noise term *(c.f. (Pavliotis, 2010)). The source of the state-dependence in the "damping" comes solely from the magnetic field. Different variants of model for such system have been studied in (Hidalgo-Gonzalez et al., 2016; Lisy and Tothova, 2013; Harko and Mocanu, 2016; Cui and Zaccone, 2018; Vuijk et al., 2019; Chun et al., 2018).*

Appendix B. Homogenization for a Class of SDEs with State-Dependent Coefficients

In this section, we recall a homogenization result that will be needed for studying homogenization for our GLEs and their functionals. This result is a special case of the main theorem in (Lim et al., 2020).

Let n_1, n_2, k_1, k_2 be positive integers. Let $\epsilon > 0$ be a small parameter and $\boldsymbol{X}^\epsilon(t) \in \mathbb{R}^{n_1}$, $\boldsymbol{Y}^\epsilon(t) \in \mathbb{R}^{n_2}$ for $t \in [0,T]$, where $T > 0$ is a constant. Let $\boldsymbol{W}^{(k_1)}$ and $\boldsymbol{W}^{(k_2)}$ denote independent Wiener processes, which are $\mathbb{R}^{k_1}$-valued and $\mathbb{R}^{k_2}$-valued respectively, on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ satisfying the usual conditions (Karatzas and Shreve, 2014).

With respect to the standard bases of $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ respectively, we write:

$$\boldsymbol{X}^\epsilon(t) = (X_1^\epsilon(t), X_2^\epsilon(t), \dots, X_{n_1}^\epsilon(t)), \tag{151}$$

$$\boldsymbol{Y}^\epsilon(t) = (Y_1^\epsilon(t), Y_2^\epsilon(t), \dots, Y_{n_2}^\epsilon(t)). \tag{152}$$

We consider the following family of singularly perturbed SDE systems[5] for $(\boldsymbol{X}^\epsilon(t), \boldsymbol{Y}^\epsilon(t)) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$:

$$d\boldsymbol{X}^\epsilon(t) = \boldsymbol{A}_1(t, \boldsymbol{X}^\epsilon(t))\boldsymbol{Y}^\epsilon(t)dt + \boldsymbol{B}_1(t, \boldsymbol{X}^\epsilon(t))dt + \boldsymbol{\Sigma}_1(t, \boldsymbol{X}^\epsilon(t))d\boldsymbol{W}^{(k_1)}(t), \tag{153}$$

$$\epsilon d\boldsymbol{Y}^\epsilon(t) = \boldsymbol{A}_2(t, \boldsymbol{X}^\epsilon(t))\boldsymbol{Y}^\epsilon(t)dt + \boldsymbol{B}_2(t, \boldsymbol{X}^\epsilon(t))dt + \boldsymbol{\Sigma}_2(t, \boldsymbol{X}^\epsilon(t))d\boldsymbol{W}^{(k_2)}(t), \tag{154}$$

with the initial conditions, $\boldsymbol{X}^\epsilon(0) = \boldsymbol{X}^\epsilon$ and $\boldsymbol{Y}^\epsilon(0) = \boldsymbol{Y}^\epsilon$, where $\boldsymbol{X}^\epsilon$ and $\boldsymbol{Y}^\epsilon$ are random variables that possibly depend on ϵ. In the SDEs (153)-(154), the coefficients $\boldsymbol{A}_1 : \mathbb{R}^+ \times \mathbb{R}^{n_1} \to \mathbb{R}^{n_1 \times n_2}$, $\boldsymbol{A}_2 : \mathbb{R}^+ \times \mathbb{R}^{n_1} \to \mathbb{R}^{n_2 \times n_2}$, $\boldsymbol{\Sigma}_2 : \mathbb{R}^+ \times \mathbb{R}^{n_1} \to \mathbb{R}^{n_2 \times k_2}$ are non-zero matrix-valued functions, whereas $\boldsymbol{B}_1 : \mathbb{R}^+ \times \mathbb{R}^{n_1} \to \mathbb{R}^{n_1}$, $\boldsymbol{B}_2 : \mathbb{R}^+ \times \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$, $\boldsymbol{\Sigma}_1 : \mathbb{R}^+ \times \mathbb{R}^{n_1} \to \mathbb{R}^{n_1 \times k_1}$ are (possibly zero) matrix-valued or vector-valued functions. They may depend on $\boldsymbol{X}^\epsilon$, as well as on t explicitly, as indicated by the parenthesis $(t, \boldsymbol{X}^\epsilon(t))$.

We are interested in the limit as $\epsilon \to 0$ of the SDEs (153)-(154), in particular the limiting behavior of the process $\boldsymbol{X}^\epsilon(t)$, under appropriate assumptions[6] on the coefficients. We make the following assumptions concerning the SDEs (153)-(154) and (155).

Assumption B.1 *The global solutions, defined on $[0,T]$, to the pre-limit SDEs (153)-(154) and to the limiting SDE (155) a.s. exist and are unique for all $\epsilon > 0$ (i.e. there are no explosions).*

5. Note that here the variables $\boldsymbol{X}^\epsilon(t)$ and $\boldsymbol{Y}^\epsilon(t)$ are general and they do not necessarily represent position and velocity variables of a physical system.
6. We forewarn the readers that our assumptions can be relaxed in various directions (see the relevant remarks in (Lim et al., 2020)) but we will not pursue these generalizations here. This approach may not be too appealing from a mathematical point of view but we stress that the main goal of the paper is to communicate, in the simplest yet rigorous manner, the consequences of the homogenization results to a broad range of audience and therefore some sacrifices in the completeness are unavoidable.

Assumption B.2 *The matrix-valued functions*

$$\{-\boldsymbol{A}_2(t,\boldsymbol{X}); t \in [0,T], \boldsymbol{X} \in \mathbb{R}^{n_1}\}$$

are uniformly positive stable, i.e. all real parts of the eigenvalues of $-\boldsymbol{A}_2(t,\boldsymbol{X})$ are bounded from below, uniformly in t and $\boldsymbol{X}$, by a positive constant (or, equivalently, the matrix-valued functions $\{\boldsymbol{A}_2(t,\boldsymbol{X}); t \in [0,T], \boldsymbol{X} \in \mathbb{R}^{n_1}\}$ are uniformly Hurwitz stable).

Assumption B.3 *For $t \in [0,T]$, $\boldsymbol{X} \in \mathbb{R}^{n_1}$, and $i = 1,2$, the functions $\boldsymbol{B}_i(t,\boldsymbol{X})$ and $\boldsymbol{\Sigma}_i(t,\boldsymbol{X})$ are continuous and bounded in t and $\boldsymbol{X}$, and Lipschitz in $\boldsymbol{X}$, whereas the functions $\boldsymbol{A}_i(t,\boldsymbol{X})$ and $(\boldsymbol{A}_i)_{\boldsymbol{X}}(t,\boldsymbol{X})$ are continuous in t, continuously differentiable in $\boldsymbol{X}$, bounded in t and $\boldsymbol{X}$, and Lipschitz in $\boldsymbol{X}$. Moreover, the functions $(\boldsymbol{A}_i)_{\boldsymbol{X}\boldsymbol{X}}(t,\boldsymbol{X})$ $(i = 1,2)$ are bounded for every $t \in [0,T]$ and $\boldsymbol{X} \in \mathbb{R}^{n_1}$.*

Assumption B.4 *The initial condition $\boldsymbol{X}^\epsilon(0) = \boldsymbol{X}^\epsilon \in \mathbb{R}^{n_1}$ is an $\mathcal{F}_0$-measurable random variable that may depend on ϵ, and we assume that $\mathbb{E}[|\boldsymbol{X}^\epsilon|^p] = O(1)$ as $\epsilon \to 0$ for all $p > 0$. Also, $\boldsymbol{X}^\epsilon$ converges, in the limit as $\epsilon \to 0$, to a random variable $\boldsymbol{X}$ as follows: $\mathbb{E}\left[|\boldsymbol{X}^\epsilon - \boldsymbol{X}|^p\right] = O(\epsilon^{pr_0})$, where $r_0 > 1/2$ is a constant, as $\epsilon \to 0$. The initial condition $\boldsymbol{Y}^\epsilon(0) = \boldsymbol{Y}^\epsilon \in \mathbb{R}^{n_2}$ is an $\mathcal{F}_0$-measurable random variable that may depend on ϵ, and we assume that for every $p > 0$, $\mathbb{E}[|\epsilon \boldsymbol{Y}^\epsilon|^p] = O(\epsilon^\alpha)$ as $\epsilon \to 0$, for some $\alpha \geq p/2$.*

We now state the homogenization theorem.

Theorem B.1 *Suppose that the family of SDE systems* (153)-(154) *satisfies Assumption B.1-B.4. Let $(\boldsymbol{X}^\epsilon(t), \boldsymbol{Y}^\epsilon(t)) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ be their solutions, with the initial conditions $(\boldsymbol{X}^\epsilon, \boldsymbol{Y}^\epsilon)$. Let $\boldsymbol{X}(t) \in \mathbb{R}^{n_1}$ be the solution to the following Itô SDE with the initial position $\boldsymbol{X}(0) = \boldsymbol{X}^\epsilon$:*

$$\begin{aligned} d\boldsymbol{X}(t) = {} & [\boldsymbol{B}_1(t,\boldsymbol{X}(t)) - \boldsymbol{A}_1(t,\boldsymbol{X}(t))\boldsymbol{A}_2^{-1}(t,\boldsymbol{X}(t))\boldsymbol{B}_2(t,\boldsymbol{X}(t))]dt \\ & + \boldsymbol{S}(t,\boldsymbol{X}(t))dt + \boldsymbol{\Sigma}_1(t,\boldsymbol{X}(t))d\boldsymbol{W}^{(k_1)}(t) \\ & - \boldsymbol{A}_1(t,\boldsymbol{X}(t))\boldsymbol{A}_2^{-1}(t,\boldsymbol{X}(t))\boldsymbol{\Sigma}_2(t,\boldsymbol{X}(t))d\boldsymbol{W}^{(k_2)}(t). \end{aligned} \tag{155}$$

In the above $\boldsymbol{S}(t,\boldsymbol{X}(t))$ is the noise-induced drift vector whose ith component is given by

$$S^i(t,\boldsymbol{X}) = -\frac{\partial}{\partial X^l}\bigg((A_1 A_2^{-1})^{ij}(t,\boldsymbol{X})\bigg) \cdot A_1^{lk}(t,\boldsymbol{X}) \cdot J^{jk}(t,\boldsymbol{X}), \tag{156}$$

where $i,l = 1,\ldots,n_1$, $j,k = 1,\ldots,n_2$, or in index-free notation,

$$\boldsymbol{S} = \boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{\nabla}\cdot(\boldsymbol{J}\boldsymbol{A}_1^T) - \boldsymbol{\nabla}\cdot(\boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{J}\boldsymbol{A}_1^T), \tag{157}$$

and $\boldsymbol{J} \in \mathbb{R}^{n_2 \times n_2}$ is the unique solution to the Lyapunov equation:

$$\boldsymbol{J}\boldsymbol{A}_2^T + \boldsymbol{A}_2\boldsymbol{J} = -\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^T. \tag{158}$$

Then the process $\boldsymbol{X}^\epsilon(t)$ converges, as $\epsilon \to 0$, to the solution $\boldsymbol{X}(t)$, of the Itô SDE (155)*, in the following sense: for all finite $T > 0$,*

$$\sup_{t\in[0,T]} |\boldsymbol{X}^\epsilon(t) - \boldsymbol{X}(t)| \to 0, \tag{159}$$

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

Remark B.1 *If $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are independent of $\boldsymbol{X}$, then the Itô equation* (155) *is equivalent to the equation:*

$$d\boldsymbol{X}(t) = [\boldsymbol{B}_1(t,\boldsymbol{X}(t)) - \boldsymbol{A}_1(t,\boldsymbol{X}(t))\boldsymbol{A}_2^{-1}(t,\boldsymbol{X}(t))\boldsymbol{B}_2(t,\boldsymbol{X}(t))]dt + \boldsymbol{H}_\alpha(t,\boldsymbol{X}(t))dt \\ + \boldsymbol{\Sigma}_1(t)d\boldsymbol{W}^{(k_1)}(t) - \boldsymbol{A}_1(t,\boldsymbol{X}(t))\boldsymbol{A}_2^{-1}(t,\boldsymbol{X}(t))\boldsymbol{\Sigma}_2(t) \circ^\alpha d\boldsymbol{W}^{(k_2)}(t), \tag{160}$$

where $\circ^\alpha$, $\alpha \in [0,1]$, specifies the rule of stochastic integration, whereby the stochastic integral is evaluated at $t_n = (1-\alpha)t_n + \alpha t_{n+1}$ on the discretization intervals $[t_n, t_{n+1}]$ (so $\alpha = 0$ corresponds to Itô integral, $\alpha = 1/2$ to Stratonovich, and $\alpha = 1$ to anti-Itô), and $\boldsymbol{H}_\alpha$ is the corresponding noise-induced drift term whose ith component is:

$$H^i_\alpha = S^i - \alpha \frac{\partial(A_1 A_2^{-1}\Sigma_2)^{ik}}{\partial X^j}(A_1 A_2^{-1}\Sigma_2)^{jk}, \tag{161}$$

with S^i given by (156)*.*

After some algebraic manipulations and using the Lyapunov equation $\boldsymbol{A}_2\boldsymbol{J} + \boldsymbol{J}\boldsymbol{A}_2^T = -\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^T$, one can rewrite H^i_α as:

$$H^i_\alpha = \frac{1}{2}Q^{qj}(\alpha)[G_q, G_j]^i, \tag{162}$$

where $\boldsymbol{G}_q$ denotes the vector field associated to the qth column of the matrix $\boldsymbol{A}_1\boldsymbol{A}_2^{-1}$, $[G_q, G_j]^i$ denotes the ith component of Lie bracket[7] *of the vector fields $\boldsymbol{G}_q$ and $\boldsymbol{G}_j$ (i.e. the derivative of $\boldsymbol{G}_j$ along the flow generated by $\boldsymbol{G}_q$), and*

$$\boldsymbol{Q}(\alpha) = \alpha \boldsymbol{J}\boldsymbol{A}_2^T - (1-\alpha)\boldsymbol{A}_2\boldsymbol{J}. \tag{163}$$

Provided that $\boldsymbol{A}_2$ is Hurwitz stable, $\boldsymbol{Q}(\alpha)$ can be represented as the solution to the Lyapunov equation (Bellman, 1997):

$$\boldsymbol{A}_2\boldsymbol{Q}(\alpha) + \boldsymbol{Q}(\alpha)\boldsymbol{A}_2^T = antisym(\boldsymbol{A}_2\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^T), \tag{164}$$

where $antisym(\boldsymbol{A})$ denotes the antisymmetric part of the matrix $\boldsymbol{A}$.

Now, let us consider the Stratonovich case $\alpha = 1/2$. In this case, $\boldsymbol{Q} := \boldsymbol{Q}(1/2)$ is the antisymmetric part of the Onsager matrix $-\boldsymbol{A}_2\boldsymbol{J}$, i.e. $\boldsymbol{Q} = (\boldsymbol{J}\boldsymbol{A}_2^T - \boldsymbol{A}_2\boldsymbol{J})/2$ (see also Remark 4.1). Therefore, when the detailed balance condition (i.e. when $\boldsymbol{A}_2\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^T$ is symmetric) holds, $\boldsymbol{Q}$ (physically a measure of irreversibility of the fast process, and mathematically a matrix encoding stochastic area of the limiting process) vanishes and the resulting limiting SDE for $\boldsymbol{X}(t)$ is a Stratonovich SDE without additional drift correction terms. On the other hand, if $\alpha = 0$ (Itô), $\boldsymbol{Q}(\alpha = 0)$ is simply the (non-zero) Onsager matrix, whereas if $\alpha = 1$ (anti-Itô), $\boldsymbol{Q}(\alpha = 1)$ equals to negative transpose of the Onsager matrix.

7. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are the first order differential operators corresponding to the vector fields $\boldsymbol{A}(\boldsymbol{x})$ and $\boldsymbol{B}(\boldsymbol{x})$, i.e. $\boldsymbol{A} = \sum_i A^i(\boldsymbol{x})\frac{\partial}{\partial x^i}$ and $\boldsymbol{B} = \sum_j B^j(\boldsymbol{x})\frac{\partial}{\partial x^j}$, then the Lie bracket (commutator) between $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as the operator $[\boldsymbol{A},\boldsymbol{B}] = \boldsymbol{A}\boldsymbol{B} - \boldsymbol{B}\boldsymbol{A}$.

References

Aykut Argun, Jalpa Soni, Lennart Dabelow, Stefano Bo, Giuseppe Pesce, Ralf Eichhorn, and Giovanni Volpe. Experimental realization of a minimal microscopic heat engine. Physical Review E*, 96(5):052106, 2017.*

Erik Aurell. Unified picture of strong-coupling stochastic thermodynamics and time reversals. Physical Review E*, 97(4):042112, 2018.*

Richard Bellman. Introduction to Matrix Analysis*, volume 19. SIAM, 1997.*

Jeremiah Birrell. Entropy anomaly in Langevin–Kramers dynamics with a temperature gradient, matrix drag, and magnetic field. Journal of Statistical Physics*, 173(6):1549–1586, Dec 2018.*

Stefano Bo and Antonio Celani. White-noise limit of nonwhite nonequilibrium processes. Physical Review E*, 88(6):062150, 2013.*

Stefano Bo and Antonio Celani. Entropy production in stochastic systems with fast and slow time-scales. Journal of Statistical Physics*, 154(5):1325–1351, 2014.*

Stefano Bo and Antonio Celani. Multiple-scale stochastic processes: Decimation, averaging and beyond. Physics Reports*, 670:1–59, 2017.*

Stefano Bo, Soon Hoe Lim, and Ralf Eichhorn. Functionals in stochastic thermodynamics: How to interpret stochastic integrals. Journal of Statistical Mechanics: Theory and Experiment*, 2019(8):084005, Aug 2019.*

Yvain Bruned, Ilya Chevyrev, and Peter K Friz. Examples of renormalized SDEs. In International Conference on Stochastic Partial Differential Equations and Related Fields*, pages 303–317. Springer, 2016.*

Antonio Celani, Stefano Bo, Ralf Eichhorn, and Erik Aurell. Anomalous thermodynamics at the microscale. Physical Review Letters*, 109(26):260603, 2012.*

Raphaël Chetrite and Krzysztof Gawędzki. Fluctuation relations for diffusion processes. Communications in Mathematical Physics*, 282(2):469–518, 2008.*

Ilya Chevyrev, Peter K Friz, Alexey Korepanov, Ian Melbourne, and Huilin Zhang. Multiscale systems, homogenization, and rough paths. In International Conference in Honor of the 75th Birthday of SRS Varadhan*, pages 17–48. Springer, 2016.*

Hyun-Myung Chun, Xavier Durang, and Jae Dong Noh. Emergence of nonwhite noise in Langevin dynamics with magnetic Lorentz force. Phys. Rev. E*, 97:032117, Mar 2018.*

Andrés Córdoba, Tsutomu Indei, and Jay D Schieber. Elimination of inertia from a Generalized Langevin Equation: Applications to microbead rheology modeling and data analysis. Journal of Rheology*, 56(1):185–212, 2012.*

Bingyu Cui and Alessio Zaccone. Generalized Langevin equation and fluctuation-dissipation theorem for particle-bath systems in external oscillating fields. Phys. Rev. E*, 97:060102, Jun 2018.*

Lennart Dabelow, Stefano Bo, and Ralf Eichhorn. Irreversibility in active matter systems: Fluctuation theorem and mutual information. Physical Review X*, 9(2):021009, 2019.*

Donald L Ermak and J Andrew McCammon. Brownian dynamics with hydrodynamic interactions. The Journal of Chemical Physics*, 69(4):1352–1360, 1978.*

Oded Farago and Niels Grønbech-Jensen. Langevin dynamics in inhomogeneous media: Re-examining the Itô-Stratonovich dilemma. Physical Review E*, 89(1):013301, 2014.*

Étienne Fodor, Denis S Grebenkov, Paolo Visco, and Frédéric van Wijland. Generalized Langevin equation with hydrodynamic backflow: Equilibrium properties. Physica A: Statistical Mechanics and its Applications*, 422:107–112, 2015.*

Peter Friz, Paul Gassiat, and Terry Lyons. Physical Brownian motion in a magnetic field as a rough path. Transactions of the American Mathematical Society*, 367(11):7939–7955, 2015.*

Peter K Friz and Martin Hairer. A Course on Rough Paths*, second edition, Springer, 2020.*

Peter K Friz and Nicolas B Victoir. Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, volume 120. Cambridge University Press, 2010.*

Crispin Gardiner. Stochastic Methods*, volume 4. Springer Berlin, 2009.*

Hao Ge and Xiao Jin. Anomalous contribution and fluctuation theorems in singular perturbed diffusion processes. arXiv preprint arXiv:1811.04529*, 2018.*

Nathan Glatt-Holtz, David Herzog, Scott McKinley, and Hung Nguyen. The generalized Langevin equation with power-law memory in a nonlinear potential well. Nonlinearity*, 33(6):2820–2852, 2020.*

Claude Godrèche and Jean-Marc Luck. Characterising the nonequilibrium stationary states of Ornstein–Uhlenbeck processes. Journal of Physics A: Mathematical and Theoretical*, 52(3):035002, 2018.*

Georg A. Gottwald, Daan T. Crommelin, and Christian L. E. Franzke. Stochastic Climate Theory*, page 209–240. Cambridge University Press, 2017.*

Igor Goychuk. Viscoelastic subdiffusion: Generalized Langevin equation approach. Advances in Chemical Physics*, 150:187, 2012.*

Peter Hänggi. Generalized Langevin equations: A useful tool for the perplexed modeller of nonequilibrium fluctuations? In Stochastic Dynamics*, pages 15–22. Springer, 1997.*

Peter Hänggi. Private communications. 2019.

Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: Fifty years after Kramers. Reviews of Modern Physics*, 62(2):251, 1990.*

Tiberiu Harko and Gabriela Mocanu. Electromagnetic radiation of charged particles in stochastic motion. The European Physical Journal C*, 76(3):160, Mar 2016.*

JC Hidalgo-Gonzalez, JI Jiménez-Aquino, and M Romero-Bastida. Non-Markovian Brownian motion in a magnetic field and time-dependent force fields. Physica A: Statistical Mechanics and its Applications*, 462:1128–1147, 2016.*

Scott Hottovy, Giovanni Volpe, and Jan Wehr. Noise-induced drift in stochastic differential equations with arbitrary friction and diffusion in the Smoluchowski-Kramers limit. Journal of Statistical Physics*, 146(4):762–773, 2012.*

Scott Hottovy, Austin McDaniel, Giovanni Volpe, and Jan Wehr. The Smoluchowski-Kramers limit of stochastic differential equations with arbitrary state-dependent friction. Communications in Mathematical Physics*, 336(3):1259–1283, 2015.*

Nobuyuki Ikeda and Shinzo Watanabe. Stochastic Differential Equations and Diffusion Processes*, volume 24. Elsevier, 2014.*

Tsutomu Indei, Jay D Schieber, Andrés Córdoba, and Ekaterina Pilyugina. Treating inertia in passive microbead rheology. Physical Review E*, 85(2):021504, 2012.*

Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. Journal of Basic Engineering*, 82(1):35–45, 1960.*

I. Karatzas and S. Shreve. Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 2014.*

Raz Kupferman. Fractional kinetics in Kac–Zwanzig heat bath models. Journal of Statistical Physics*, 114(1):291–326, 2004.*

Benedict Leimkuhler and Matthias Sachs. Ergodic properties of quasi-Markovian generalized Langevin equations with configuration dependent noise and non-conservative force. In Giambattista Giacomin, Stefano Olla, Ellen Saada, Herbert Spohn, and Gabriel Stoltz, editors, Stochastic Dynamics Out of Equilibrium*, pages 282–330, Cham, 2019. Springer International Publishing.*

Antoine Lejay and Terry Lyons. On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. Application to homogenization. In Current Trends in Potential Theory*, volume 7. The Theta foundation/American Mathematical Society, 2003.*

Antoine Lejay et al. On the convergence of stochastic integrals driven by processes converging on account of a homogenization property. Electronic Journal of Probability*, 7, 2002.*

Paul Lévy et al. Wiener's random function, and other Laplacian random functions. In Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. The Regents of the University of California, 1951.*

Maximilian Leyman, Freddie Ogemark, Jan Wehr, and Giovanni Volpe. Tuning phototactic robots with sensorial delays. Physical Review E*, 98(5):052606, 2018.*

Soon Hoe Lim and Jan Wehr. Homogenization for a class of generalized Langevin equations with an application to thermophoresis. Journal of Statistical Physics*, Nov 2018.*

Soon Hoe Lim, Jan Wehr, and Maciej Lewenstein. Homogenization for generalized Langevin equations with applications to anomalous diffusion. Annales Henri Poincaré*, 2020.*

Vladimir Lisy and Jana Tothova. Brownian motion of charged particles driven by correlated noise in magnetic field. Transport Theory and Statistical Physics*, 42(6-7):365–380, 2013.*

Olga Lopusanschi and Damien Simon. Area anomaly and generalized drift of iterated sums for hidden Markov walks. arXiv preprint arXiv:1709.04288*, 2017.*

Olga Lopusanschi and Damien Simon. Lévy area with a drift as a renormalization limit of Markov chains on periodic graphs. Stochastic Processes and their Applications*, 128(7): 2404–2426, 2018.*

Terry J Lyons. Differential equations driven by rough signals. Revista Matemática Iberoamericana*, 14(2):215–310, 1998.*

Martin Lysy, Natesh S Pillai, David B Hill, M Gregory Forest, John WR Mellnik, Paula A Vasquez, and Scott A McKinley. Model comparison and assessment for single particle tracking in biological fluids. Journal of the American Statistical Association*, 111(516): 1413–1426, 2016.*

Katarzyna Macieszczak, Kay Brandner, and Juan P. Garrahan. Unified thermodynamic uncertainty relations in linear response. Phys. Rev. Lett.*, 121:130601, Sep 2018.*

Simon JA Malham and Anke Wiese. An introduction to SDE simulation. arXiv preprint arXiv:1004.0646*, 2010.*

Scott A McKinley, Lingxing Yao, and M Gregory Forest. Transient anomalous diffusion of tracer particles in soft matter. Journal of Rheology (1978-present)*, 53(6):1487–1506, 2009.*

Hazime Mori. Transport, collective motion, and Brownian motion. Progress of Theoretical Physics*, 33(3):423–455, 1965.*

Hung D Nguyen. The small-mass limit and white-noise limit of an infinite dimensional generalized Langevin equation. Journal of Statistical Physics*, 173(2):411–437, 2018.*

M Ottobre and GA Pavliotis. Asymptotic analysis for the generalized Langevin equation. Nonlinearity*, 24(5):1629, 2011.*

Rui Pan, Thai M. Hoang, Zhaoyu Fei, Tian Qiu, Jonghoon Ahn, Tongcang Li, and H. T. Quan. Quantifying the validity and breakdown of the overdamped approximation in stochastic thermodynamics: Theory and experiment. Phys. Rev. E*, 98:052105, Nov 2018.*

GA Pavliotis. Asymptotic analysis of the Green–Kubo formula. IMA Journal of Applied Mathematics*, 75(6):951–967, 2010.*

Grigorios A Pavliotis. Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, volume 60. Springer, 2014.*

Grigoris Pavliotis and Andrew Stuart. Multiscale Methods: Averaging and Homogenization*. Springer Science & Business Media, 2008.*

Luc Rey-Bellet. Open classical systems. In Open Quantum Systems II*, pages 41–78. Springer, 2006.*

Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. Reports on Progress in Physics*, 75(12):126001, 2012.*

Ken Sekimoto. Stochastic Energetics*, volume 799. Springer, 2010.*

Francisco J Sevilla. The non-equilibrium nature of active motion. In Quantitative Models for Microscopic to Macroscopic Biological Macromolecules and Tissues*, pages 59–86. Springer, 2018.*

P Siegle, I Goychuk, and P Hänggi. Markovian embedding of fractional superdiffusion. EPL (Europhysics Letters)*, 93(2):20002, 2011.*

IM Sokolov. Itô, Stratonovich, Hänggi and all the rest: The thermodynamics of interpretation. Chemical Physics*, 375(2-3):359–363, 2010.*

Héctor J Sussmann. Limits of the Wong-Zakai type with a modified drift term. In Stochastic Analysis*, pages 475–493. Elsevier, 1991.*

Peter Talkner and Peter Hänggi. Colloquium: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical. Rev. Mod. Phys.*, 92:041002, Oct 2020.*

Harry L Trentelman, Anton A Stoorvogel, and Malo Hautus. Control Theory for Linear Systems. Communications and Control Engineering Series, Springer*, 2002.*

Hidde Derk Vuijk, Joseph Michael Brader, and Abhinav Sharma. Anomalous fluxes in overdamped Brownian dynamics with Lorentz force. Journal of Statistical Mechanics: Theory and Experiment*, 2019(6):063203, 2019.*

JC Willems and JH Van Schuppen. 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, Mass., June 18–29, 1979*, volume 62, page 283. Springer, 1980.*

Eugene Wong and Moshe Zakai. On the convergence of ordinary integrals to stochastic integrals. The Annals of Mathematical Statistics*, 36(5):1560–1564, 1965.*

Xiang Yang, Chang Liu, Yunyun Li, Fabio Marchesoni, Peter Hänggi, and HP Zhang. Hydrodynamic and entropic effects on colloidal diffusion in corrugated channels. Proceedings of the National Academy of Sciences*, 114(36):9564–9569, 2017.*

Robert Zwanzig. Nonlinear generalized Langevin equations. Journal of Statistical Physics*, 9(3):215–220, 1973.*