COMMUN. MATH. SCI.
Vol. 16, No. 8, pp. 2199–2230
© 2018 International Press

MEAN FIELD LIMITS FOR NON-MARKOVIAN INTERACTING PARTICLES: CONVERGENCE TO EQUILIBRIUM, GENERIC FORMALISM, ASYMPTOTIC LIMITS AND PHASE TRANSITIONS*

MANH HONG DUONG† AND GRIGORIOS A. PAVLIOTIS‡

Abstract. In this paper, we study the mean field limit of weakly interacting particles with memory that are governed by a system of non-Markovian Langevin equations. Under the assumption of quasi-Markovianity (i.e. the memory in the system can be described using a **finite** number of auxiliary processes), we pass to the mean field limit to obtain the corresponding McKean-Vlasov equation in an extended phase space. For the case of a quadratic confining potential and a quadratic (Curie-Weiss) interaction, we obtain the fundamental solution (Green's function). For nonconvex confining potentials, we characterize the stationary state(s) of the McKean-Vlasov equation, and we show that the bifurcation diagram of the stationary problem is independent of the memory in the system. In addition, we show that the McKean-Vlasov equation for the non-Markovian dynamics can be written in the GENERIC formalism and we study convergence to equilibrium and the Markovian asymptotic limit.

Keywords. mean field limits; non-Markovian interacting particles; convergence to equilibrium; GENERIC; asymptotic limit; phase transitions.

AMS subject classifications. 82C31; 60H10; 35K10; 60J60; 60F17.

1. Introduction

Many systems in nature and in applications can be modeled using systems of interacting particles (or agents) that are possibly subject to thermal noise. Standard examples include plasma physics and galactic dynamics [5]; more recent applications are dynamical density functional theory [31,33], mathematical biology [27,43] and even mathematical models in the social sciences [35,50]. As examples of models of interacting agents in a noisy environment, that appear in the social sciences, we mention the modelling of cooperative behavior [10], risk management [34] and opinion formation [35].

For weakly interacting diffusions, i.e. when the strength of the interaction is inversely proportional to the number of particles, the mean field limit for the empirical measure has been studied rigorously, under quite general assumptions, see e.g. [12,28,52]. The mean field dynamics is described by the so-called McKean-Vlasov equation, which is a nonlinear, nonlocal Fokker-Planck-type equation [4,44,45]. This equation has been studied extensively and many things are known, including well-posedness of the evolution PDE [6], convergence to equilibrium, (non)uniqueness of the invariant measure and phase transitions. For example, it is by now well known that, for nonconvex confining potentials and a quadratic (Curie-Weiss type) interaction term–the so-called Desai-Zwanzig model [21], there exist more than one invariant measures at low temperatures [10,62,69]. Similar results of nonuniqueness of the stationary state at low temperatures have been also obtained for McKean-Vlasov equations modeling opinion formation [9,71] as well as for the Desai-Zwanzig model in a two-scale potential [32].

Most works on the study of the McKean-Vlasov equation are concerned with the equation that is obtained in the mean field limit of weakly interacting overdamped or un-

*Received: May 25, 2018; Accepted (in revised form): August 22, 2018. Communicated by Pierre-Emmanuel Jabin.

†School of Mathematics, University of Birmingham, Birmingham B15 2TT, UK (h.duong@bham.ac.uk).

‡Department of Mathematics, Imperial College London, London SW7 2AZ, UK (g.pavliotis@imperial.ac.uk).

derdamped Langevin dynamics. The McKean-Vlasov equation corresponding to weakly interacting overdamped Langevin dynamics is a (nonlinear and nonlocal) uniformly elliptic parabolic PDE and many of its properties are by now well understood, see for instance [7,8] and references therein. For weakly interacting underdamped Langevin diffusions, the corresponding McKean-Vlasov equation is not uniformly elliptic but, rather, hypoelliptic. However, many results are also known in this case, see for instance [20,70] and references therein. A recent result that is particularly relevant for the present work is that the invariant measure(s) of the McKean-Vlasov equation in phase space have a product measure structure and, in particular that the overdamped and underdamped McKean-Vlasov dynamics exhibit the same phase transitions [17]. One of the goals of this paper is to extend this result to McKean-Vlasov equations corresponding to non-Markovian Langevin dynamics.

To our knowledge, the mean field limit of interacting particles that are driven by colored noise, or that have memory is much less studied. Both problems are quite interesting from a modeling perspective, given that noise in many physical systems exhibits a nontrivial (spatiotemporal) correlation structure, and that many interesting dynamical systems are non-Markovian. As examples where colored noise and non-Markovianity play an important role, we mention colloidal systems and polymer dynamics [63] and nonequilibrium systems and active matter [25,47]. The presence of non-white noise and of memory renders the analysis more complicated, since, even for a finite number of interacting agents, the corresponding Fokker-Planck equation does not have a gradient structure (for colored noise) and it is necessarily degenerate (for systems with memory [54]). The main goal of this paper is to study the mean field limit and the properties of the resulting McKean-Vlasov equation for non-Markovian weakly interacting particle systems, under the assumption that the memory (or non-trivial correlation structure of the noise) can be described by means of a *finite* number of auxiliary processes–the so-called quasi-Markovian assumption [59][Ch. 8]. The study of mean field limits for weakly interacting particles driven by colored noise will be presented elsewhere.

Finite dimensional, and in particular, low dimensional stochastic systems with memory and or/colored noise have been studied extensively in the literature. We mention the work on noise-induced transitions for Langevin equations driven by colored noise [36,37] and the study of the generalized Langevin equation, including the rigorous derivation of such non-Markovian equations for particles coupled to one or more heat baths [61] and the references therein, [59][Ch. 8], the study of the ergodic properties of such systems [66], of their long-time behaviour [22] etc.

The starting point of the present work is the following particle system

$$dQ_i(t) = P_i(t)\,dt \tag{1.1a}$$

$$dP_i(t) = -\nabla_q V(Q_i(t))\,dt - \frac{1}{N}\sum_{j=1}^{N} \nabla_q U(Q_i(t)-Q_j(t))\,dt + \lambda^T Z_i(t)\,dt \tag{1.1b}$$

$$dZ_i(t) = -\lambda P_i(t)\,dt - AZ_i(t) + \sqrt{2\beta^{-1}A}dW_i(t). \tag{1.1c}$$

Here, $i=1,\dots,N$, $(Q_i,P_i,Z_i)\in\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^{dm} =: \mathbf{X}$; $\lambda\in\mathbb{R}^{dm\times d}$ and $A\in\mathbb{R}^{dm\times dm}$ are given matrices with A being symmetric positive definite; $\beta>0$ is the inverse temperature; and $V,U:\mathbb{R}^d\to\mathbb{R}$ are confining and interaction potentials, respectively. The notation ∇_q denotes the gradient operator with respect to q variable. Finally, $\{W_i\}_{i=1}^N$ are independent $\mathbb{R}^{dm}$-dimensional Wiener processes.

This diffusion process in the extended phase space $\mathbf{X}$ is equivalent to the system of weakly interacting generalized Langevin equations,

$$\ddot{Q}_i = -\nabla V(Q_i) - \frac{1}{N}\sum_{j=1}^{N}\nabla_q U(Q_i - Q_j) - \sum_{j=1}^{N}\int_0^t \gamma_{ij}(t-s)\dot{Q}_j(s)\,ds + F_i(t), \quad i=1,\dots N, \tag{1.2}$$

where $F(t) = (F_1(t),\dots F_N(t))$ is a mean zero, Gaussian, stationary process with autocorrelation function

$$E(F_i(t)F_j(s)) = \beta^{-1}\gamma_{ij}(t-s).$$

We can obtain the Markovian dynamics (1.1) from the generalized Langevin dynamics (1.2), under the assumption that the autocorrelation functions $[\gamma_{ij}(t-s)]_{i,j=1,\dots,m}$ are given by a linear combination of exponential functions [54]. In this case, the memory of the system can be described by adding a finite number of auxiliary stationary Ornstein-Uhlenbeck processes. Approximating arbitrary Gaussian stationary processes by a Markovian process in a rigorous, systematic and algorithmic way is still an open problem, related to the so-called stochastic realization problem [14,42], see also [40]. A recent approach to this problem, based on the representation of the noise by an infinite dimensional Ornstein-Uhlenbeck process is presented in [51], see also [58]. We plan to return to the combined Markovian approximation and mean field limit in a future work.

It is possible to study the hydrodynamic (mean field) limit, i.e. to show that the empirical measure

$$\rho_N := \frac{1}{N}\sum_{i=1}^{N}\delta_{(Q_i(t),P_i(t),Z_i(t))}$$

converges to the solution of

$$\begin{aligned}\partial_t\rho = &-\mathrm{div}_q(p\rho) + \mathrm{div}_p\Big[(\nabla_q V(q) + \nabla_q U*\rho(q) - \lambda^T z)\rho\Big] + \mathrm{div}_z\Big[(p\lambda + Az)\rho\Big] \\ &+ \beta^{-1}\mathrm{div}_z(A\nabla_z\rho),\end{aligned} \tag{1.3}$$

where the convolution $\nabla_q U*\rho$ is defined by $(\nabla_q U*\rho)(q) = \int_{\mathbf{X}}\nabla_q U(q-q')\rho(q',p',z')\,dq'dp'dz'$. We will refer to equation (1.3) as the *generalized McKean-Vlasov equation (GLMV)* for $\rho = \rho(t,q,p,z)$. It is the forward Kolmogorov (Fokker-Planck) equation associated to the following generalized McKean-Vlasov SDE

$$dQ(t) = P(t)\,dt, \tag{1.4a}$$
$$dP(t) = -\nabla V(Q(t))\,dt - \nabla_q U*\rho_t(Q_t)\,dt + \lambda^T Z(t)\,dt, \tag{1.4b}$$
$$dZ(t) = -P(t)\lambda\,dt - AZ(t) + \sqrt{2\beta^{-1}A}\,dW(t). \tag{1.4c}$$

where $\rho_t = \mathrm{Law}(Q(t),P(t),Z(t))$. The rigorous passage from the particle approximation (1.1) to the generalized McKean-Vlasov SDE (1.4) and the mean field limit for small Lipschitz interactions, can be justified, in principle, using the coupling approach developed in [3], see also [19]. The smallness restriction on the interaction can be removed using the results presented in [48]. These papers consider the case of interacting underdamped Langevin dynamics, but it should be possible to use similar techniques in order to study rigorously the mean field limit for the generalized Langevin interacting dynamics. We mention that the system of interacting particles that we study in this paper is similar to systems of interacting nonlinear oscillators, coupled to two heat baths at different temperatures [22]. Two related equations that also play an important role in this work are the *overdamped McKean-Vlasov equation*,

$$\partial_t \rho = \operatorname{div}\left[(\nabla V + \nabla U * \rho)\rho\right] + \beta^{-1}\Delta\rho,$$

and the *underdamped McKean-Vlasov equation*

$$\partial_t \rho = -\operatorname{div}_q(p\rho) + \operatorname{div}_p\left[(\nabla_q V(q) + \nabla_q U * \rho(q))\rho\right] + \gamma \operatorname{div}_p\left[p\rho\right] + \beta^{-1}\gamma\Delta_p\rho.$$

The generalized Langevin equation, i.e. equation (1.4) in the absence of an interaction potential ($U \equiv 0$), has been studied extensively. In particular, the ergodic properties and the hypoelliptic and hypocoercive structure of the dynamics and of the corresponding Fokker-Planck equation have been analyzed in [54] where the homogenization and overdamped limits are also studied. Another recent paper that has motivated this work is [13], where the overdamped limit of the underdamped McKean-Vlasov equation has been derived using variational and large deviations approach. In this paper, we generalize these aforementioned results to equation (1.4) for a quite broad class of confining and interaction potentials. We obtain the fundamental solution (Green's function) for the generalized McKean-Vlasov equation (1.3) for the case of a quadratic confining potential and a quadratic (Curie-Weiss) interaction. For nonconvex confining potentials, we characterize its stationary state(s), showing that the bifurcation diagram of the stationary problem is independent of the memory in the system. In addition, we study convergence to equilibrium and the Markovian asymptotic limit for both the finite system (1.1) and equation (1.3) using the formal perturbation expansions method developed in [59,60]. Furthermore, we demonstrate that equation (1.3) shares a similar GENERIC (General Equation for NonEquilibrium Reversible-Irreversible Coupling [57]) structure, which is a well-established formalism for non-equilibrium thermodynamical systems, unifying both reversible and irreversible dynamics, with the overdamped and underdamped McKean-Vlasov equations.

The rest of the paper is organized as follows. In Section 2, we calculate the spectrum of the Fokker-Planck operator associated to the finite system (1.1) and construct the fundamental solution to the mean field equation (1.3) in the case of quadratic confining and interaction potentials. In Section 3, we study nonconvex confining potentials. We provide a characterization of stationary states, study phase transition and show convergence to a stationary state. In Section 4, we recast the generalized McKean-Vlasov equation into the GENERIC framework, thereby obtaining an alternative derivation of its stationary solutions. In Section 5, we study the white noise limit for the generalized McKean-Vlasov dynamics, both for the particle system and for the mean field limit. Finally, a brief summary of the results obtained in this paper and discussion on future work are presented in Section 6.

2. Quadratic confining and interaction potentials: calculation of the fundamental solution, spectral analysis and convergence to equilibrium

In this section, we compute and compare the spectrum of the three operators associated with the overdamped McKean-Vlasov dynamics, the underdamped McKean-Vlasov dynamics and the generalized McKean-Vlasov (gMV) dynamics. In particular, in Section 2.1, we study the finite dimensional problem, and in Section 2.2, we calculate the spectrum and the fundamental solution of the mean field PDEs, for quadratic interaction and confining potentials.

2.1. Finitely many weakly interacting particles. We consider the following N-particle weakly interacting particle systems where the dynamics of the i-particle ($i=1,\dots,N$) depends on the average interaction with all other particles.

- The overdamped McKean-Vlasov (oMV) dynamics

$$dX_i = -\nabla V(X_i)\,dt - \frac{1}{N}\sum_{j=1}^{N}\nabla U(X_i - X_j)\,dt + \sqrt{2\beta^{-1}}dW_i. \tag{2.1}$$

- The underdamped McKean-Vlasov (uMV) dynamics

$$dQ_i = P_i\,dt, \tag{2.2a}$$
$$dP_i = -\nabla V(Q_i)\,dt - \frac{1}{N}\sum_{j=1}^{N}\nabla U(Q_i - Q_j)\,dt - \gamma P_i\,dt + \sqrt{2\gamma\eta^{-1}}dW_i. \tag{2.2b}$$

- The generalized McKean-Vlasov (gMV) dynamics

$$dQ_i = P_i\,dt, \tag{2.3a}$$
$$dP_i = -\nabla V(Q_i)\,dt - \frac{1}{N}\sum_{j=1}^{N}\nabla U(Q_i - Q_j)\,dt + \lambda^T Z_i\,dt, \tag{2.3b}$$
$$dZ_i = -P_i\lambda\,dt - AZ_i\,dt + \sqrt{2\beta^{-1}A}dW_i. \tag{2.3c}$$

In this section, we consider the case where both the external and interacting potentials are quadratic

$$V(\xi) = \frac{1}{2}\omega^2\xi^2 \quad \text{and} \quad U(\xi) = \frac{1}{2}\eta^2\xi^2. \tag{2.4}$$

In this case, all the dynamics can be written in the form of a (possibly degenerate) Ornstein-Ulenbeck process [41, 49, 56]

$$d\mathcal{Y}^N = \mathcal{B}^N\mathcal{Y}^N\,dt + \sqrt{2\beta^{-1}\mathcal{D}^N}\,dW^N, \tag{2.5}$$

with the corresponding state spaces, drift and diffusion matrices (written using block matrices):

- The quadratic oMV dynamics

$$\mathcal{Y}^N_{oMV} = (X_1,\dots,X_N) \in \mathbb{R}^{dN}, \quad \mathsf{Q}^N_{oMV} = I_{dN\times dN};$$
$$\mathcal{B}^N_{oMV} = \begin{pmatrix} -\left(\omega^2 + \frac{N-1}{N}\eta^2\right) & \dots & \frac{1}{N}\eta^2 \\ \vdots & \ddots & \vdots \\ \frac{1}{N}\eta^2 & \dots & -\left(\omega^2 + \frac{N-1}{N}\eta^2\right) \end{pmatrix} \in \mathbb{R}^{dN\times dN}.$$

- The quadratic uMV dynamics

$$\mathcal{Y}^N_{uMV} = (Q_1,\dots,Q_N,P_1,\dots,P_N) \in \mathbb{R}^{2dN}, \quad \mathcal{B}^N_{uMV} = \left(\begin{array}{c|c} 0 & I \\ \hline \mathcal{B}^N_{oMV} & -\gamma \end{array}\right),$$
$$\mathcal{D}^N_{uMV} = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & \gamma \end{array}\right).$$

- The quadratic gMV dynamics

$$\mathcal{Y}^N_{gMV} = (Q_1,\dots,Q_N,P_1,\dots,P_N,Z_1^1,\dots,Z_1^m,\dots,Z_N^1,\dots,Z_N^m) \in \mathbb{R}^{(2+m)dN},$$

$$\mathcal{B}^N_{gMV} = \begin{pmatrix} 0 & I & 0 & \dots & 0 \\ \mathcal{B}^N_{oMV} & 0 & \lambda_1 & \dots & \lambda_m \\ 0 & -\lambda_1 & -\alpha_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -\lambda_m & 0 & \dots & -\alpha_m \end{pmatrix}, \quad \mathcal{D}^N_{gMV} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \alpha_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \alpha_m \end{pmatrix}.$$

2.1.1. Calculation of the spectrum and of the fundamental solution. The quadratic cases are of special interest because we can characterize the spectrum of the dynamics explicitly, which enables us to compare their spectrum and rates of convergence to equilibrium. Our analysis will be based on the result in [49, 56] on the description of the spectrum of general hypoelliptic quadratic systems. Before introducing the result, we need to recall the relevant concept of hypoellipticity. We consider the following general Ornstein-Ulenbeck operator

$$P = Bx\cdot\nabla + \mathrm{div}(D\nabla), \quad x\in\mathbb{R}^d, \tag{2.6}$$

where B and D are real $d\times d$-matrices, with D symmetric positive semidefinite. We say that P is a hypoelliptic Ornstein-Uhlenbeck operator if one of the following equivalent conditions (thus all of them) holds

(i) The symmetric positive semidefinite matrices

$$D_t = \int_0^t e^{sB} D e^{sB^T}\,ds \tag{2.7}$$

are nonsingular for some (equivalently, for all) $t>0$, i.e., $\det D_t > 0$.

(ii) The Kalman rank condition holds

$$\mathrm{Rank}[B|D^{\frac12}] = d$$

where $[B|D^{\frac12}] = [D^{\frac12}, BD^{\frac12},\dots,B^{d-1}D^{\frac12}]$ is the $d\times d^2$ matrix obtained by writing consecutively the columns of the matrices $B^jD^{\frac12}$ $(j=0,\dots,d-1)$, with $D^{\frac12}$ being the symmetric positive semidefinite matrix given by the square root of D.

(iii) The Hörmander condition holds, i.e., the Lie algebra generated by $Y_0, X_1,\dots,X_d$ has full rank at every point in $\mathbb{R}^d$,

$$\forall x\in\mathbb{R}^d, \quad \mathrm{Rank}\,\mathcal{L}(X_1,\dots,X_d,Y_0)(x) = d,$$

with

$$Y_0 := Bx\cdot\nabla, \quad X_i = \sum_{j=1}^d D_{ij}\partial_{x_j}, i = 1,\dots,d.$$

It is known that [39, 41] when the Ornstein-Ulenbeck operator P in (2.6) is hypoelliptic, then it has a smooth fundamental solution $\Gamma(t,x,y)$ given explicitly by

$$\Gamma(t,x,y) = \frac{1}{(4\pi)^{d/2}\sqrt{\det D_t}} \exp\Big(-\frac14 (x-e^{-tB}y)^T D_t^{-1}(x-e^{-tB}y)\Big), \tag{2.8}$$

where D_t is given by equation (2.7). Furthermore, it admits a unique invariant measure that is absolutely continuous with respect to the Lebesgue measure $d\mu(x)=\rho(x)\,dx$ with

$$\rho(x)=\frac{1}{(4\pi)^{d/1}\sqrt{\det D_\infty}}\exp\Big(-\frac{1}{4}x^T D_\infty^{-1}x\Big),$$

where

$$D_\infty=\int_0^\infty e^{sB}De^{sB^T}\,ds.$$

We will make use of the following result.

PROPOSITION 2.1 ([49,56]). *Suppose that*

$$P=Bx\cdot\nabla+\mathrm{div}(D\nabla),\quad x\in\mathbb{R}^d,$$

is a hypoelliptic Ornstein-Ulenbeck operator that has the invariant measure $d\mu(x)=\mu(x)\,dx$. *Then the spectrum of the operator* $P:L^2_\mu\to L^2_\mu$ *equipped with the domain*

$$D(P)=\{u\in L^2_\mu:Pu\in L^2_\mu\},$$

is only composed of eigenvalues with finite algebraic multiplicities given by

$$\sigma(P)=\Big\{\sum_{\lambda\in\sigma(B)}\lambda k_\lambda:k_\lambda\in\mathbb{N}\Big\}. \tag{2.9}$$

We note that in [49] the spectrum in all L^p_μ spaces is calculated. By direct verification, see for instance [55, Section 4.2] for such a verification for similar models, one can show that all the quadratic systems considered in the previous section are hypoelliptic. To characterize their spectrum, according to Proposition 2.1, we need to calculate eigenvalues of their corresponding drift matrices B.

- *Spectrum of the quadratic oMV dynamics.* The eigenvalues of $\mathcal{B}^N_{oMV}$ can be found by direct computation:

$$\sigma(\mathcal{B}^N_{oMV})=\{-\omega^2,-(\omega^2+\eta^2)\}.$$

Thus

$$\sigma(L^N_{oMV})=\{-\omega^2k_1-(\omega^2+\eta^2)k_2:k_1,k_2\in\mathbb{N}\}.$$

- *Spectrum of the quadratic uMV dynamics.* Let α be an eigenvalue of $\mathcal{B}^N_{uMV}$, that is, there exist $\xi_1,\xi_2\in\mathbb{R}^{dN}$ such that

$$\begin{pmatrix}0 & I\\ \mathcal{B}^N_{oMV} & -\gamma\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}=\alpha\begin{pmatrix}\xi_1\\ xi_2.\end{pmatrix}$$

This implies that $\xi_2=\alpha\xi_1$ and $\mathcal{B}^N_{oMV}\xi_1-\gamma\xi_2=\lambda\xi_2$. Substituting the former into the latter, we get

$$\mathcal{B}^N_{oMV}\xi_1=\alpha(\alpha+\gamma)\xi_1.$$

Thus $\alpha(\alpha+\gamma)$ is an eigenvalue of $\mathcal{B}^N_{oMV}$, from which we deduce that

$$\alpha(\alpha+\gamma)=-\omega^2\quad\text{or}\quad\alpha(\alpha+\gamma)=-(\omega^2+\eta^2).$$

Solving these equations, we find

$$\alpha_{1,2} = \frac{-\gamma \pm \sqrt{\gamma^2 - 4\omega^2}}{2} \quad \text{and} \quad \alpha_{3,4} = \frac{-\gamma \pm \sqrt{\gamma^2 - 4(\omega^2 + \eta^2)}}{2}. \tag{2.10}$$

The spectrum of the quadratic uMV dynamics is then given by

$$\sigma(L^N_{uMV}) = \Big\{ \sum_{i=1}^{4} k_i \alpha_i : k_i \in \mathbb{N}, i = 1, \dots, 4 \Big\}.$$

- *Spectrum of the quadratic gMV dynamics.* Similarly, let ν be an eigenvalue of $\mathcal{B}^N_{gMV}$, that is, for some $\xi_1, \xi_2, \zeta_1, \dots, \zeta_m \in \mathbb{R}^{dN}$, it holds that

$$\begin{pmatrix} 0 & I & 0 & \dots & 0 \\ \mathcal{B}^N_{oMV} & 0 & \lambda_1 & \dots & \lambda_m \\ 0 & -\lambda_1 & -\alpha_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -\lambda_m & 0 & \dots & -\alpha_m \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \zeta_1 \\ \vdots \\ \zeta_m \end{pmatrix} = \nu \begin{pmatrix} \xi_1 \\ \xi_2 \\ \zeta_1 \\ \vdots \\ \zeta_m \end{pmatrix}.$$

This is equivalent to the system

$$\begin{cases} \xi_2 & = \nu \xi_1, \\ B^N_{MV} \xi_1 + \sum_{j=1}^m \lambda_j \zeta_j & = \nu \xi_2, \\ -\lambda_j \xi_2 - \alpha_j \zeta_j & = \nu \zeta_j \quad \text{for all} \quad j = 1, \dots, m. \end{cases}$$

By substituting the first and the last m equations into the second one, we get

$$\mathcal{B}^N_{MV} \xi_1 = \nu \Big(\nu + \sum_{j=1}^{m} \frac{\lambda_j^2}{\nu + \alpha_j} \Big) \xi_1.$$

It follows that $\nu\Big(\nu + \sum_{j=1}^m \frac{\lambda_j^2}{\nu + \alpha_j}\Big)$ is an eigenvalue of $\mathcal{B}^N_{MV}$. Thus, we have

$$\nu \Big(\nu + \sum_{j=1}^{m} \frac{\lambda_j^2}{\nu + \alpha_j} \Big) = -\omega^2 \quad \text{or} \quad \nu \Big(\nu + \sum_{j=1}^{m} \frac{\lambda_j^2}{\nu + \alpha_j} \Big) = -(\omega^2 + \eta^2).$$

In general, these equations, which are equivalent to two polynomial equations of degree $m+2$, can not be solved explicitly. Suppose that $\nu_1, \dots, \nu_{m+2}$ and $\nu_{m+3}, \dots, \nu_{2m+4}$ are solutions to the first and second equations, respectively. Then the spectrum of the quadratic gMV dynamics is given by

$$\sigma(L^N_{gMV}) = \Big\{ \sum_{j=1}^{2m+4} \nu_j k_j : k_j \in \mathbb{N}, j = 1, \dots, 2m+4 \Big\}.$$

REMARK 2.1. From the calculation of the spectrum of the Fokker-Planck operator and by applying [56, Theorem 2.7] it follows that we have exponentially fast convergence to equilibrium in the L^2-space weighted by the invariant measure. This is also true in L^1 at least for the overdamped McKean-Vlasov equation, see [65].

2.2. The mean field PDE. The mean-field limits of the quadratic systems considered in Section 2.1 can be written in a common form

$$dX(t) = BX(t)\,dt + K(X(t) - \langle X(t)\rangle)\,dt + \sqrt{2D^{\frac{1}{2}}}\,dW(t), \tag{2.11}$$

for suitable matrices B, K and D with D symmetric positive semidefinite. Here $\langle X(t)\rangle$ denotes the average of $X(t)$ with respect to its own law.

In this section, we will obtain the fundamental solution of the Fokker-Planck equation associated to the SDE (2.11), which is given by

$$\begin{aligned}&\partial_t\rho(x,t) = -\operatorname{div}\Big(Bx\rho(x,t)\Big) - \operatorname{div}\Big(\int_{\mathbb{R}^n} K(x-x')\rho(x',t)\,dx'\,\rho(x,t)\Big) + \operatorname{div}\Big(D\nabla\rho(x,t)\Big),\\ &\rho(x,0) = \delta(x-x_0).\end{aligned} \tag{2.12}$$

Proposition 2.2. *The solution of the initial value problem (2.12), i.e. the Green's function of the Fokker-Planck operator, is given by,*

$$\rho(x,t) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\mathcal{Q}(t)}}\exp\Big[-\frac{1}{2}\mathcal{Q}^{-1}(t)(x-m(t)x_0)\cdot(x-m(t)x_0)\Big], \tag{2.13}$$

where the mean vector and the covariance matrix are given, respectively, by

$$m(t) = e^{tB} \quad \text{and} \quad \mathcal{Q}(t) = 2\int_0^t e^{2s(B+K)}D\,ds. \tag{2.14}$$

Proof. Equation (2.12) is a nonlinear nonlocal degenerate partial differential equation. Since the law of the finite particle system corresponding to (2.11), when starting at a deterministic initial condition, is Gaussian, its mean field limit is also Gaussian. Therefore, we look for the fundamental solution to (2.12) of the form of a multivariate Gaussian distribution with time-dependent mean and covariance matrix

$$\rho(x,t) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\mathcal{Q}(t)}}\exp\Big[-\frac{1}{2}\mathcal{Q}^{-1}(t)(x-m(t)x_0)\cdot(x-m(t)x_0)\Big]. \tag{2.15}$$

We will find the mean $m(t)$ and the variance $\mathcal{Q}(t)$ by requiring that $\rho(x,t)$ satisfies equation (2.12). Let us define

$$g(x,t) := -\log\rho(x,t) = \frac{1}{2}\mathcal{Q}^{-1}(t)(x-m(t)x_0)\cdot(x-m(t)x_0) + s(t), \tag{2.16}$$

where $s(t) = \frac{1}{2}\log\det\mathcal{Q}(t) + \frac{n}{2}\log(2\pi)$. The logarithmic transformation has been used previously in the literature, particularly in optimal control [24]. Substituting $\rho = \exp(-g)$ in (2.12), we can rewrite the equation in terms of g as follows

$$\partial_t g = -(B+K)x\cdot\nabla g + D:\nabla^2 g - D\nabla g\cdot\nabla g + h(t)\cdot\nabla g + \operatorname{Tr}(B+K), \tag{2.17}$$

where the notation $:$ denotes the Frobenius inner product of two matrices $A:B = \sum_{ij}A_{ij}B_{ij} = \operatorname{Tr}(A^TB)$ and

$$h(t) = \int Kx'\rho(x',t)\,dx'$$

$$
\begin{aligned}
&= \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathcal{Q}(t)}} \int K x' \exp\Big[-\frac{1}{2}\mathcal{Q}^{-1}(t)(x'-m(t)x_0)\cdot(x'-m(t)x_0)\Big]\,dx' \\
&= K m(t) x_0.
\end{aligned}
$$

To find $\mathcal{Q}$ and m, we set equal the corresponding terms in two sides of (2.17). Using the explicit formula (2.16) of g, we compute the left-hand side of (2.17), which is the time-derivative of g

$$
\partial_t g = \frac{1}{2}\partial_t(\mathcal{Q}^{-1}(t))x\cdot x - \partial_t(\mathcal{Q}^{-1}(t)m(t))x_0\cdot x + \frac{1}{2}\partial_t\Big(m(t)^T\mathcal{Q}^{-1}(t)m(t)\Big)x_0\cdot x_0 + \partial_t s(t). \tag{2.18}
$$

Since $\nabla g = \mathcal{Q}^{-1}(t)(x-m(t)x_0)$, $\nabla^2 g = \mathcal{Q}^{-1}(t)$, the right-hand side of (2.17) is given by

$$
\begin{aligned}
&-(B+K)x\cdot\nabla g + D:\nabla^2 g - D\nabla g\cdot\nabla g + h(t)\cdot\nabla g + \mathrm{Tr}(B+K) \\
&= -\mathcal{Q}^{-1}(t)(B+K)x\cdot(x-m(t)x_0) - \mathcal{Q}^{-1}(t)D\mathcal{Q}^{-1}(t)(x-m(t)x_0)\cdot(x-m(t)x_0) \\
&\qquad + \mathcal{Q}^{-1}(t)Km(t)x_0\cdot(x-m(t)x_0) + D:\mathcal{Q}^{-1}(t) - \mathrm{Tr}(B+K) \\
&= \Big[-\mathcal{Q}^{-1}(t)(B+K) - \mathcal{Q}^{-1}(t)D\mathcal{Q}^{-1}(t)\Big]x\cdot x - \Big[-m(t)^T\mathcal{Q}^{-1}(t)(B+K) \\
&\quad - 2m(t)^T\mathcal{Q}^{-1}(t)D\mathcal{Q}^{-1}m(t) - \mathcal{Q}^{-1}Km(t)\Big]x_0\cdot x \\
&\quad - \Big[m(t)^T\mathcal{Q}^{-1}(t)D\mathcal{Q}^{-1}(t)m(t) + m(t)^T\mathcal{Q}^{-1}(t)Km(t)\Big]x_0\cdot x_0 \\
&\quad + D:\mathcal{Q}^{-1}(t) + \mathrm{Tr}(B+K).
\end{aligned} \tag{2.19}
$$

By comparing (2.18) and (2.19), we obtain the following equations

$$
\partial_t \mathcal{Q}^{-1}(t) = 2\Big[-\mathcal{Q}^{-1}(t)(B+K) - \mathcal{Q}^{-1}(t)D\mathcal{Q}^{-1}(t)\Big], \tag{2.20}
$$

$$
\partial_t(\mathcal{Q}^{-1}(t)m(t)) = -m(t)^T\mathcal{Q}^{-1}(t)(B+K) - 2m(t)^T\mathcal{Q}^{-1}(t)D\mathcal{Q}^{-1}m(t) - \mathcal{Q}^{-1}Km(t), \tag{2.21}
$$

$$
\partial_t(m(t)^T\mathcal{Q}^{-1}(t)m(t)) = -2\Big[m(t)^T\mathcal{Q}^{-1}(t)D\mathcal{Q}^{-1}(t)m(t) + m(t)^T\mathcal{Q}^{-1}(t)Km(t)\Big], \tag{2.22}
$$

$$
\partial_t s(t) = D:\mathcal{Q}^{-1}(t) + \mathrm{Tr}(B+K). \tag{2.23}
$$

We will solve these equations together with the initial value datum $\mathcal{Q}(0)=\mathbf{0}, m(0)=\mathbf{I}$ and $s(0)=\frac{n}{2}\log(2\pi)$. Equation (2.20) is a nonlinear Riccati-type equation for $\mathcal{Q}^{-1}(t)$. We use the following formula for the derivative of the inverse of a matrix

$$
\frac{dA^{-1}(t)}{dt} = -A^{-1}(t)\frac{dA(t)}{dt}A^{-1}(t).
$$

Applying this formula for $\mathcal{Q}(t)$ and using (2.20), we get

$$
\begin{aligned}
\partial_t\mathcal{Q}(t) &= -\mathcal{Q}(t)\partial_t(\mathcal{Q}^{-1}(t))\mathcal{Q}(t) = -2\mathcal{Q}(t)\Big[-\mathcal{Q}^{-1}(t)(B+K) - \mathcal{Q}^{-1}(t)D\mathcal{Q}^{-1}(t)\Big]\mathcal{Q}(t) \\
&= 2\Big[D + (B+K)\mathcal{Q}(t)\Big].
\end{aligned} \tag{2.24}
$$

Using Duhamel's formula, we obtain

$$
\mathcal{Q}(t) = 2\int_0^t e^{2(t-s)(B+K)}D\,ds = 2\int_0^t e^{2s(B+K)}D\,ds. \tag{2.25}
$$

From (2.20) and (2.21) we deduce that

$$\partial_t m(t) = Bm(t), \tag{2.26}$$

which yields $m(t) = e^{tB}$. From (2.20) and (2.26), we have

$$\begin{aligned}
&\partial_t(m(t)^T \mathcal{Q}^{-1}(t)m(t)) \\
=&\partial_t m(t)^T \mathcal{Q}^{-1}(t)m(t) + m(t)^T \partial_t(\mathcal{Q}^{-1}(t))m(t) + m(t)^T \mathcal{Q}^{-1}(t)\partial_t m(t) \\
=&-2\Big[m(t)^T \mathcal{Q}^{-1}(t)D\mathcal{Q}^{-1}(t)m(t) + m(t)^T \mathcal{Q}^{-1}(t)Km(t)\Big],
\end{aligned}$$

which is exactly (2.22). In other words, (2.22) is a consequence of (2.20) and (2.21). Similarly, we show now that (2.23) is also redundant as a consequence of (2.24). In fact, applying the formula for the derivative of the determinant of a matrix, we get

$$\begin{aligned}
\partial_t s(t) &= \frac{1}{2}\partial_t \log\det \mathcal{Q}(t) = \frac{1}{2}\frac{\partial_t \det \mathcal{Q}(t)}{\mathcal{Q}(t)} \\
&= \frac{1}{2}\mathrm{Tr}\Big(\mathcal{Q}^{-1}(t)\partial_t \mathcal{Q}(t)\Big) \\
&\overset{(2.24)}{=} \mathrm{Tr}\Big[\mathcal{Q}^{-1}(t)\big(D + (B+K)\mathcal{Q}(t)\big)\Big] \\
&= D : \mathcal{Q}^{-1}(t) + \mathrm{Tr}(B+K),
\end{aligned}$$

which is indeed (2.23), as claimed. In conclusion, we find that

$$\mathcal{Q}(t) = 2\int_0^t e^{2s(B+K)}D\,ds \quad \text{and} \quad m(t) = e^{tB}.$$

∎

Remark 2.2.
1) By taking $d=1$, $B=-\gamma$, $K=\kappa$ and $D=Q$, we recover the known result [26, Equation (3.166)] for the Shimizu-Yamada model (the overdamped McKean-Vlasov equation with quadratic interaction and confining potentials).
2) Suppose that $\mathscr{L} = \mathcal{S} + \mathcal{A}$ is an infinitesimal generator of a diffusion process that possesses an invariant measure ρ_∞ and satisfies

$$\mathcal{S} = \mathcal{S}^*, \quad \mathcal{A} = -\mathcal{A}^* \text{ in } L^2_{\rho_\infty} \text{ and } \mathcal{S}\rho_\infty = 0 = \mathcal{A}\rho_\infty.$$

Then the following relation holds [59, Chapter 4]:

$$\mathscr{L}^*(f\rho_\infty) = (\mathcal{S}f - \mathcal{A}f)\rho_\infty, \tag{2.27}$$

From this relation, one can find the fundamental solution to the forward Fokker-Plank equation, $\mathscr{L}^*\rho = 0$, from the fundamental solution to the backward Fokker-Planck equation, $\mathscr{L}f = 0$, and vice versa.

3. Nonconvex potentials: stationary solutions, phase transition and convergence to equilibrium

For non-convex confining potentials and for the Curie-Weiss quadratic interaction potential, it is well known that the the McKean-Vlasov equation exhibits a continuous phase transition [10]: at sufficiently high temperatures, only one stationary distribution exists, corresponding to zero mean (magnetization order parameter), whereas at low temperatures this mean zero state becomes unstable and two new branches emerge (for the Landau quadratic potential), corresponding to a nonzero magnetization. A natural question is whether the addition of inertia, i.e. the underdamped McKean-Vlasov equation, or the addition of memory in the system can change the structure of this pitchfork bifurcation. This problem was studied for the underdamped McKean-Vlasov equation in [17], where it was shown that the presence of inertia does not change the bifurcation diagram. In this section we show that this is the case also for the generalized McKean-Vlasov dynamics.

In this section, we make the following assumptions on the confining potential V, the interaction potential U and the initial data ρ_0. They are made in order to rely on the results in [18, 67, 68].

Assumption 3.1.

(i) $V \in C^\infty(\mathbb{R}^d)$; $V(q) \geq C_4|q|^4 - C_2|q|^2$ *for some positive constants* C_2 *and* C_4; $|\nabla V(q)| \leq K|q|^{2m}$ *for some positive constants* m *and* K; $\nabla^2 V(q) > 0$ *when* $q \notin \mathcal{K}$ *for some compact subset* $\mathcal{K} \subset \mathbb{R}^d$. *Finally,* $\lim_{|q|\to+\infty} \nabla^2 V(q) = +\infty$.

(ii) U *is nonnegative and there exists an even, positive and convex polynomial function* G *with* $\deg(G) =: 2n \geq 2$ *such that* $U(q) = G(|q|)$.

(iii) *The initial measure* ρ_0 *admits a* $\mathcal{C}^\infty$*-continuous density with respect to the Lebesgue measure and has a finite entropy. Furthermore, its* $8r^2$*-th moment with respect to the variable* q, *where* $r := \max\{m, n\}$, *and its second moment with respect to the variable* p *are finite.*

3.1. Stationary solutions: characterization and phase transition. We first provide a characterization of stationary solutions to equation (1.3), i.e., solution to the following equation

$$\mathsf{K}[\rho]\rho = 0, \tag{3.1}$$

where

$$\begin{aligned} \mathsf{K}[\mu]\rho = -\mathrm{div}_q(p\rho) + \mathrm{div}_p\Big[(\nabla_q V(q) + \nabla_q U * \mu(q) - \lambda^T z)\rho\Big] \\ + \mathrm{div}_z\Big[(p\lambda + Az)\rho\Big] + \beta^{-1}\mathrm{div}_z(A\nabla_z \rho), \end{aligned} \tag{3.2}$$

for a given $\mu \in L^1(\mathbf{X})$. Note that in writing $\mathsf{K}[\rho]\rho$ in (3.1) (and throughout the paper), by abuse of notation, we identify a measure that has a Lebesgue density with its density. For a given μ, the operator $\mathsf{K}[\mu]\rho$ is linear in ρ which can be seen as the linearization of the (quadratic) operator $\mathsf{K}[\rho]\rho$ around the state μ.

We start by showing that the stationary McKean-Vlasov equation (3.1) can be rewritten as a nonlinear integral equation. This reformulation of the stationary Fokker-Planck equation is well known for the overdamped McKean-Vlasov equation, see, e.g. [64][Lem. 4.1, Thm. 4.1].

Proposition 3.1. *If there exists a solution* $\rho_\infty \in L^1(\mathbf{X}) \cap L^\infty(\mathbf{X})$ *to equation* (3.1) *then*

$$\rho_\infty(q,p,z) = \frac{1}{Z}\exp\left[-\beta\Big(\frac{|p|^2}{2} + \frac{\|z\|^2}{2} + V(q) + (U * \rho_\infty)(q)\Big)\right], \tag{3.3}$$

where Z *is the normalization constant*

$$Z = \int_{\mathbf{X}} \exp\left[-\beta\Big(\frac{|p|^2}{2} + \frac{\|z\|^2}{2} + V(q) + (U * \rho_\infty)(q)\Big)\right] d\mathbf{x}. \tag{3.4}$$

Conversely, any probability measure whose density satisfies (3.3) *is a stationary solution to* (1.3)*.*

In the proof of Proposition 3.1 below we will need the following divergence theorem in the whole space [29, Section 4.5.2]:

THEOREM 3.1. *Let F be a C^1 vector field in $\mathbb{R}^n$. Assume that $\|F\|_1<\infty$ and $\|\operatorname{div}F\|_1<\infty$. Then*

$$\int_{\mathbb{R}^n}\operatorname{div}F\,d\mathbf{x}=0.$$

Proof. **(Proof of Proposition 3.1).** The converse statement is obviously verified by direct computations. We adapt the proof of a similar statement for the underdamped McKean-Vlasov equation from [17, Proposition 1] and [16, Proposition 1 & 2] to prove the necessary condition for the generalized McKean-Vlasov equation. It consists of two steps. Due to the presence of the additional variable z, Step 1 below will be more involved than those in aforementioned papers.

Step 1: The linearized equation. We first consider the linearized equation

$$\mathsf{K}\rho:=\mathsf{K}[\mu]\rho=0, \tag{3.5}$$

where $\mu\in L^1(\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^{dm})$ is given. We define

$$\mathbb{H}_\mu(q,p,z):=\frac{1}{2}|p|^2+V(q)+\frac{1}{2}\|z\|^2+(U*\mu)(q)\quad\text{and} \tag{3.6a}$$

$$\mathbb{J}=\begin{pmatrix}0&-I&0\\I&0&-\lambda^T\\0&\lambda&0\end{pmatrix}\in\mathbb{R}^{(2d+m)\times(2d+m)}. \tag{3.6b}$$

Then equation (3.5) can be written in a compact form as

$$\mathsf{K}[\mu]\rho=\operatorname{div}[\mathbb{J}\nabla\mathbb{H}_\mu\rho]+\operatorname{div}_z[A(z\rho+\beta^{-1}\nabla_z\rho)]=0,$$

where the div and ∇ in the first term are with respect to the full variable $\mathbf{x}=(q,p,z)$.

We prove the following claim: define

$$f(q,p,z):=\frac{1}{\bar{Z}}\exp\left[-\beta\left(\frac{|p|^2}{2}+\frac{\|z\|^2}{2}+V(q)+U*\mu(q)\right)\right]=\frac{1}{\bar{Z}}\exp(-\beta\mathbb{H}_\mu),$$

where $\bar{Z}$ is the normalization constant, $\bar{Z}=\int_{\mathbf{X}}e^{-\beta\mathbb{H}_\mu}\,dx$, and

$$\begin{aligned}\mathcal{A}:=\Big\{u:\mathbf{X}\to\mathbb{R}:g:=uf^{-1/2}\ \text{ satisfies that }\ g\in C^1,\|\mathbb{J}\nabla\mathbb{H}_\mu g^2\|_1<\infty,\\ \|\operatorname{div}(\mathbb{J}\nabla\mathbb{H}_\mu g^2)\|_1<\infty,\ \|f^{1/2}gA\nabla_z(f^{-1/2}g)\|_1<\infty,\\ \&\ \ \|\operatorname{div}(f^{1/2}gA\nabla_z(f^{-1/2}g))\|_1<\infty\Big\}.\end{aligned}$$

Then f is the unique solution in $\mathcal{A}$ to the linearized equation (3.5). The conditions in $\mathcal{A}$ will ensure later the application of Theorem 3.1. To verify $f\in\mathcal{A}$, it is sufficient to show that $\|\mathbb{J}\nabla\mathbb{H}f\|_1<\infty$. This is true under Assumption 3.1 on the confining and interaction potentials. In addition, straightforward computations give

$$\operatorname{div}[\mathbb{J}\nabla\mathbb{H}_\mu f]=\mathbb{J}\nabla\mathbb{H}_\mu\cdot\nabla f+f\operatorname{div}(\mathbb{J}\nabla\mathbb{H}_\mu)=-\beta\mathbb{J}\nabla\mathbb{H}_\mu\cdot\nabla\mathbb{H}f=0\quad\text{and}$$

$$\mathrm{div}_z[A(zf+\beta^{-1}\nabla_z f]=0,$$

where we have used the antisymmetry of $\mathbb{J}$ and the fact that $\nabla_z f=-\beta z f$. Therefore, $\mathsf{K}[\mu]f=0$. Suppose that there exists another solution $\bar{f}\in\mathcal{A}$ to equation (3.5). We need to prove that $\bar{f}=f$. Let $g:=\bar{f}f^{-1/2}$, i.e., $\bar{f}=gf^{1/2}$. We have

$$\begin{aligned}\mathrm{div}[\mathbb{J}\nabla\mathbb{H}_\mu\bar{f}]&=\mathbb{J}\nabla\mathbb{H}_\mu\cdot\nabla(gf^{1/2})\\&=f^{1/2}\mathbb{J}\nabla\mathbb{H}_\mu\cdot\nabla g+g\mathbb{J}\nabla\mathbb{H}_\mu\cdot\nabla f^{1/2}\\&=f^{1/2}\mathbb{J}\nabla\mathbb{H}_\mu\cdot\nabla g\\&=f^{1/2}\,\mathrm{div}[\mathbb{J}\nabla\mathbb{H}_\mu g],\end{aligned}$$

where we have used the fact that $\mathbb{J}\mathbb{H}_\mu\cdot\nabla f^{1/2}=-\frac{\beta}{2}f^{1/2}\mathbb{J}\mathbb{H}_\mu\cdot\nabla\mathbb{H}_\mu=0$. Furthermore,

$$\begin{aligned}\mathrm{div}_z[A(z\bar{f}+\beta^{-1}\nabla_z\bar{f})]&=\mathrm{div}_z[A\Big(z\bar{f}+\beta^{-1}\nabla_z(ff^{-1/2}g))]\\&=\mathrm{div}_z\Big[A(z\bar{f}+\beta^{-1}f\nabla_z(f^{-1/2}g)+\beta^{-1}f^{-1/2}g\nabla_z f\Big)\Big]\\&=\beta^{-1}\,\mathrm{div}_z\Big[Af\nabla_z(f^{-1/2}g)\Big],\end{aligned}$$

since $z\bar{f}+\beta^{-1}gf^{-1/2}\nabla_z f=z\bar{f}-gf^{-1/2}fz=z\bar{f}-z\bar{f}=0$. Therefore, we obtain that

$$\mathsf{K}[\mu]\bar{f}=f^{1/2}\,\mathrm{div}[J\nabla\mathbb{H}_\mu g]+\beta^{-1}\,\mathrm{div}_z\Big[Af\nabla_z(f^{-1/2}g)\Big].$$

We now define

$$\mathsf{Q}g:=-f^{-1/2}\mathsf{K}[\mu]\bar{f}=-\mathrm{div}[J\nabla\mathbb{H}_\mu g]-\beta^{-1}f^{-1/2}\,\mathrm{div}_z\Big[Af\nabla_z(f^{-1/2}g)\Big].$$

We have $Qg=0$. On the other hand, we have

$$\begin{aligned}0=\langle\mathsf{Q}g,g\rangle_{L^2}&=-\int_{\mathbf{X}}\Big(\mathrm{div}[\mathbb{J}\nabla\mathbb{H}_\mu g]+\beta^{-1}f^{-1/2}\,\mathrm{div}_z\Big[Af\nabla_z(f^{-1/2}g)\Big]\Big)g\,d\mathbf{x}\\&=\frac{1}{2}\int_{\mathbf{X}}\mathrm{div}[\mathbb{J}\nabla\mathbb{H}_\mu g^2]\,d\mathbf{x}+\int_{\mathbf{X}}\mathrm{div}[f^{1/2}gA\nabla_z(f^{-1/2}g)]\,d\mathbf{x}\\&\quad+\beta^{-1}\int_{\mathbf{X}}fA\nabla_z(f^{-1/2}g)\cdot\nabla_z(f^{-1/2}g)\,d\mathbf{x}\\&=\beta^{-1}\int_{\mathbf{X}}fA\nabla_z(f^{-1/2}g)\cdot\nabla_z(f^{-1/2}g)\,d\mathbf{x}.\end{aligned}$$

We note that the first two integrals vanish because $f\in\mathcal{A}$ and Theorem 3.1. Since the matrix A is positive definite, it follows that $\nabla_z(f^{-1/2}g)=0$, i.e., $f^{-1/2}g=h(q,p)$ for some function h. Hence $\bar{f}=gf^{1/2}=fh(q,p)$ and

$$\begin{aligned}0&=\mathsf{K}[\mu]\bar{f}=\mathsf{K}[\mu](fh)\\&=\mathrm{div}[\mathbb{J}\nabla\mathbb{H}_\mu(fh)]+\mathrm{div}_z[A(zfh+\beta^{-1}\nabla_z(fh))]\\&=f\mathbb{J}\nabla\mathbb{H}_\mu\cdot\nabla h+h\mathsf{K}[\mu]f.\end{aligned}$$

This implies that

$$0=\mathbb{J}\nabla\mathbb{H}_\mu\cdot\nabla h=\nabla_p\mathbb{H}_\mu\cdot\nabla_q h+(-\nabla_q\mathbb{H}_\mu+\lambda\nabla_z\mathbb{H}_\mu)\cdot\nabla_p h,$$

hence $\nabla_p h = \nabla_q h = 0$, i.e., h is a constant. Since $\|\bar{f}\|_{L^1} = \|f\|_{L^1} = 1$, h must be 1 and therefore $\bar{f} = f$ as required.

Step 2: The nonlinear equation. Now suppose that ρ_∞ satisfies the (nonlinear) stationary equation (3.1), i.e., $\mathsf{K}[\rho_\infty]\rho_\infty = 0$. According to Step 1, ρ_∞ must satisfy

$$\rho_\infty = \frac{\exp\left[-\beta\left(\frac{|p|^2}{2} + \frac{\|z\|^2}{2} + V(q) + U * \rho_\infty(q)\right)\right]}{\int_{\mathbf{X}} \exp\left[-\beta\left(\frac{|p|^2}{2} + \frac{\|z\|^2}{2} + V(q) + U * \rho_\infty(q)\right)\right] d\mathbf{x}}, \tag{3.7}$$

which is the claimed statement. □

As a direct consequence of Proposition 3.1 we obtain the following corollary, which is a natural generalization of the fact that the invariant measures of the finite dimensional underdamped Langevin and generalized Langevin dynamics are product measures.

Corollary 3.1. *The number of stationary states of the overdamped, underdamped and generalized Mckean-Vlasov equations is the same.*

Proof. Stationary solutions of the overdamped, the underdamped and the generalized McKean-Vlasov equations satisfy the following integral equations

$$\rho_\infty^{oMV} = \frac{\exp\left[-\beta\left(V(q) + U * \rho_\infty(q)\right)\right]}{\int_{\mathbf{X}} \exp\left[-\beta\left(V(q) + U * \rho_\infty(q)\right)\right] dq},$$

$$\rho_\infty^{uMV} = \frac{\exp\left[-\beta\left(\frac{|p|^2}{2} + V(q) + U * \rho_\infty(q)\right)\right]}{\int_{\mathbf{X}} \exp\left[-\beta\left(\frac{|p|^2}{2} + V(q) + U * \rho_\infty(q)\right)\right] d\mathbf{x}} = \frac{\exp\left(-\frac{|p|^2}{2}\right)}{\int_{\mathbb{R}^d} \exp\left(-\frac{|p|^2}{2}\right) dp} \rho_\infty^{MV},$$

and

$$\begin{aligned}\rho_\infty^{gMV} &= \frac{\exp\left[-\beta\left(\frac{|p|^2}{2} + \frac{\|z\|^2}{2} + V(q) + U * \rho_\infty(q)\right)\right]}{\int_{\mathbf{X}} \exp\left[-\beta\left(\frac{|p|^2}{2} + \frac{\|z\|^2}{2} + V(q) + U * \rho_\infty(q)\right)\right] d\mathbf{x}} \\ &= \frac{\exp\left(-\frac{\|z\|^2}{2}\right)}{\int_{\mathbb{R}^{md}} \exp\left(-\frac{\|z\|^2}{2}\right) dz} \frac{\exp\left(-\frac{|p|^2}{2}\right)}{\int_{\mathbb{R}^d} \exp\left(-\frac{|p|^2}{2}\right) dp} \rho_\infty^{MV},\end{aligned}$$

respectively. Due to the separability of the variables p and z in the above formulas, the number of stationary solutions of the underdamped and the generalized McKean-Vlasov equations are the same as that of the overdamped Mckean-Vlasov equation. □

Remark 3.1. A consequence of Corollary 3.1 is that, for nonconvex confining potentials, and for the Curie-Weiss quadratic interaction potential, the bifurcation diagrams and even the critical temperature (see, e.g. [10][Result II, Remark 3.3.1]) are the same for the overdamped, underdamped and generalized McKean-Vlasov dynamics.

Remark 3.2 (Effect of colored noise on the structure of phase transitions).
The McKean-Vlasov dynamics with colored noise is given by

$$dX(t) = -\nabla V(X(t))\,dt - \nabla U(X(t)) * \rho_t\,dt + \delta n\,dt, \tag{3.8a}$$

$$dn(t) = \alpha n(t)\,dt + \sqrt{2\alpha\beta^{-1}}\,dW(t), \tag{3.8b}$$

where $\rho_t = \mathrm{Law}(X(t), n(t))$, δ, α are given positive constants. The Fokker-Planck equation associated to the SDE (3.8) is given by

$$\frac{\partial \rho}{\partial t} = \mathrm{div}_x \Big[(\nabla V(x) + \nabla U(q) * \rho + \delta n)\rho \Big] + \alpha \Big(\mathrm{div}_n(n\rho) + \beta^{-1}\Delta_n \rho \Big). \tag{3.9}$$

There does not appear to be an obvious gradient structure for equation (3.9). An invariant measure $\rho_\infty = \rho_\infty(x,n)\,dxdn$ to (3.9) satisfies

$$\mathrm{div}_x \Big[(\nabla V(x) + \nabla U(x) * \rho_\infty + \delta n)\rho_\infty \Big] + \alpha \Big(\mathrm{div}_n(n\rho_\infty) + \beta^{-1}\Delta_n \rho_\infty \Big) = 0.$$

It implies that ρ_∞ can not be written in the form

$$\rho_\infty(x,n) = Z^{-1} \exp\Big[-\beta \Big(V(x) + U * \rho_\infty(x) + \frac{\|n\|^2}{2} \Big) \Big],$$

For this reason, the effect of colored noise on the structure of phase transitions turns out to be nontrivial. We will study this problem in a forthcoming work.

3.2. Convergence to equilibrium. In this section, we prove that (sufficiently smooth) solutions to the generalized McKean-Vlasov equation converge to a stationary solution using the free energy approach developed in [1], see also [18]. In this paper we do not address the problem of well-posedness and regularity of solutions to the generalized McKean-Vlasov equation. However, we note that under Assumption 3.1, the McKean-Vlasov equation has a smooth solution [38, Lemma 2.1] (see also [45, 64]). In the absence of an interaction potential, the (linear) underdamped McKean-Vlasov and the generalized McKean-Vlasov equations are hypoelliptic [54] and, for smooth confining potentials, it is possible to prove regularity results for their solution [54]. In the presence of an interaction potential, for, say, smooth and bounded interaction potentials and for smooth initial data, regularity of solutions follows from the results in, e.g. [54]. For more general classes of sufficiently smooth interaction potentials, it should be possible, in principle, to adapt the techniques of [1, 2] to show that a solution to the generalised McKean-Vlasov equation is sufficiently smooth so that the computations in this section can be fully justified; but this issue is out of the scope of the present paper and we leave it for future study.

We consider the following *free energy functional*

$$\begin{aligned} F(\rho) &= \beta^{-1} \int \rho \log \rho + \int \Big[\frac{1}{2}|p|^2 + V(q) + \frac{1}{2}\|z\|^2 \Big] \rho \\ &\quad + \frac{1}{2} \int U(q-q')\rho(q',p',z')\rho(q,p,z)\, dq'dp'dz'\, dqdpdz \\ &= \int \Big[\frac{1}{2}|p|^2 + V(q) + \frac{1}{2}\|z\|^2 + \frac{1}{2} U * \rho + \beta^{-1}\log \rho \Big] \rho, \end{aligned} \tag{3.10}$$

and define $h(t) := F(\rho(\cdot,t))$.

The following lemma establishes useful properties of the free energy functional $F(\cdot)$ and of its time derivative along the flow $\rho(\cdot)$.

LEMMA 3.1. *The following assertions hold*

(1) F is bounded from below.

(2) h is a decreasing function of time and its time derivative is given by

$$\dot{h}(t) = -\int A(z\sqrt{\rho} + 2\beta^{-1}\nabla_z \sqrt{\rho}) \cdot (z\sqrt{\rho} + 2\beta^{-1}\nabla_z\sqrt{\rho}) \le 0. \tag{3.11}$$

(3) The limit $\lim_{t\to+\infty} h(t)$ *is well-defined.*

Proof. We start with proving that the free energy functional is bounded from below. In fact, since $U\geq 0$ and $\inf_{q\in\mathbb{R}^d}\Big(V(q)-\frac{|q|^2}{2}\Big)\geq C_1>-\infty$, we have

$$\begin{aligned} F(\rho) &\geq C_1+\int\Big[\frac{|p|^2}{2}+\frac{|q|^2}{2}+\frac{\|z\|^2}{2}\Big]\rho+\beta^{-1}\int\rho\log\rho \\ &= C_1+\beta^{-1}\mathcal{H}(\rho||\mu)-C_2 \\ &\geq C_1-C_2>-\infty, \end{aligned}$$

where $\mu:=\frac{1}{Z_1}\exp\Big(-\beta\big(\frac{|p|^2}{2}+\frac{|q|^2}{2}+\frac{\|z\|^2}{2}\big)\Big)$ with Z_1 being the normalization constant, $C_2=\ln Z_1$, and $\mathcal{H}(\rho||\mu)\geq 0$ is the relative entropy between $\rho(\mathbf{x})\,dx$ and μ, and is defined by

$$\mathcal{H}(\rho||\mu):=\int\log\Big(\frac{\rho(\mathbf{x})}{\mu(\mathbf{x})}\Big)\rho(\mathbf{x})\,d\mathbf{x}.$$

It is well known that $\mathcal{H}(\rho||\mu)\geq 0$, see for instance [11, Lemma 1.4.1]. Next, we prove the second assertion. Using $\mathbb{H}$ and $\mathbb{J}$ defined in (3.6a), equation (1.3) can be written as

$$\partial_t\rho=-\operatorname{div}\Big[\mathbb{J}\nabla\mathbb{H}\rho\Big]+\operatorname{div}_z\Big[A(z\rho+\beta^{-1}\nabla_z\rho)\Big],$$

where div on the right-hand side of the above equation denotes the divergence with respect to $\mathbf{x}=(q,p,z)$. This representation is more convenient for the following computations:

$$\begin{aligned} \dot{h}(t) &= \frac{d}{dt}F(\rho(t)) \\ &= \int\Big[\mathbb{H}+\beta^{-1}(\log\rho+1)\Big]\partial_t\rho \\ &= \int\Big[\mathbb{H}+\beta^{-1}\log\rho\Big]\cdot\Big[-\operatorname{div}(\mathbb{J}\nabla\mathbb{H}\rho)+\operatorname{div}_z(Az\rho+\beta^{-1}A\nabla_z\rho)\Big] \\ &= \int\Big[\mathbb{J}\mathbb{H}\cdot\mathbb{H}\rho+\beta^{-1}\mathbb{J}\nabla\mathbb{H}\cdot\nabla\rho\Big]-(Az\rho+\beta^{-1}A\nabla_z\rho)\cdot\nabla_z(H+\beta^{-1}\log\rho) \\ &\overset{(*)}{=} -\int\frac{1}{\rho}A(z\rho+\beta^{-1}\nabla_z\rho)\cdot(z\rho+\beta^{-1}\nabla_z\rho) \\ &= -\int A(z\sqrt{\rho}+2\beta^{-1}\nabla_z\sqrt{\rho})\cdot(z\sqrt{\rho}+2\beta^{-1}\nabla_z\sqrt{\rho}), \end{aligned}$$

where, to obtain the equality $(*)$, we have used the fact that the square bracket in the line above vanishes due to the antisymmetry of $\mathbb{J}$ and integration by parts, while the last equality follows from $(*)$ using the property that $\nabla_z\sqrt{\rho}=\frac{\nabla_z\rho}{2\sqrt{\rho}}$. Finally, the last assertion is a direct consequence of the first and second ones. ∎

Corollary 3.2. *For any $T>0$, we have*

$$\lim_{t\to\infty}\int_0^T\int_{\mathbf{X}}A\big(z\sqrt{\rho^{(t)}}+2\beta^{-1}\nabla_z\sqrt{\rho^{(t)}}\big)\cdot\big(z\sqrt{\rho^{(t)}}+2\beta^{-1}\nabla_z\sqrt{\rho^{(t)}}\big)\,d\mathbf{x}\,ds=0 \qquad (3.12)$$

where $\rho^{(t)}$ is defined by $\rho^{(t)}(s,\mathbf{x}):=\rho(t+s,\mathbf{x})$.

Proof. For any given $T>0$, by part (iii) of Lemma 3.1, we have

$$\lim_{t\to\infty}\int_0^T \dot h(t+s)\,ds=\lim_{t\to\infty}(h(t+T)-h(t))=0.$$

On the other hand, by part (ii), we get

$$\begin{aligned}0&=\lim_{t\to\infty}\int_0^T \dot h(t+s)\,ds\\&=\lim_{t\to\infty}\int_0^T\int_{\mathbf{X}} A\big(z\sqrt{\rho^{(t)}}+2\beta^{-1}\nabla_z\sqrt{\rho^{(t)}}\big)\cdot\big(z\sqrt{\rho^{(t)}}+2\beta^{-1}\nabla_z\sqrt{\rho^{(t)}}\big)\,d\mathbf{x}\,ds,\end{aligned}$$

which proves the statement of the corollary. ∎

Given any sequence $\{t_n\}\to\infty$, we will denote by ρ^n the function $\rho^{(t_n)}$ (cf. the definition of $\rho^{(t)}$ for a given time t given in Corollary 3.2). The following lemma establishes compactness properties of the sequence $\{\rho^n\}$.

LEMMA 3.2. *The following assertions hold:*
(1) The two sequences $\{Q_1^n, n\in\mathbb{N}\}$ and $\{Q_2^n, n\in\mathbb{N}\}$, defined in

$$Q_1^n:=\int_{\mathbf{X}}\rho^n\log\rho^n\mathbb{1}_{\rho^n\ge 1}\,d\mathbf{x},$$

and

$$Q_2^n:=\int_{\mathbf{X}}\Big(\frac{|p|^2}{2}+\frac{\|z\|^2}{2}+V(q)+\frac12 U*\rho^n\Big)\rho^n\,d\mathbf{x}$$

are uniformly bounded in time.
(2) The sequence of densities $\{\rho^n, n\in\mathbb{N}\}$ is relatively compact in $C([0;T],L^1(\mathbf{X}))$.

Proof. The proof of this lemma follows along the same lines as that of [18, Lemma 3.5] and [Lemma 5.2] [1]; hence we omit it here. ∎

LEMMA 3.3. *Let $\xi\in C^2(\mathbb{R})$ with $\xi''\in L^\infty(\mathbb{R})$ and $\xi(0)=0$. Then the following equation holds in the sense of distribution on $(0,T)\times\mathbf{X}$*

$$\frac{\partial\xi(\rho)}{\partial t}-\mathsf{K}[\rho]\xi(\rho)=\mathrm{Tr}(A)\Big(\rho\xi'(\rho)-\xi(\rho)\Big)-\beta^{-1}\xi''(\rho)A\nabla_z\rho\cdot\nabla_z\rho, \tag{3.13}$$

where the operator K *is defined in* (3.2).

Proof. The proof follows along the lines of [1, Lemma 3.9]. We sketch the main idea here. Multiplying the equation $\frac{\partial\rho}{\partial t}\rho-\mathsf{K}[\rho]\rho=0$ with $\xi'(\rho)$, we obtain

$$\frac{\partial\xi(\rho)}{\partial t}-\Big(\mathsf{K}[\rho]\rho\Big)\xi'(\rho)=0. \tag{3.14}$$

Direct computations and using the following identities

$$\mathrm{div}_z(Az\rho)\xi'(\rho)=\mathrm{div}_z(Az\xi(\rho))-\mathrm{Tr}(A)(\xi(\rho)-\rho\xi'(\rho)),$$

and

$$\mathrm{div}_z(A\nabla_z\rho)\xi'(\rho)=\mathrm{div}_z(A\nabla_z\xi(\rho))-\xi''(\rho)A\nabla_z\rho\cdot\nabla_z\rho,$$

yield:

$$\Big(\mathsf{K}[\rho]\rho\Big)\xi'(\rho)=\mathsf{K}[\rho]\xi(\rho)-\mathrm{Tr}(A)(\xi(\rho)-\rho\xi'(\rho))-\xi''(\rho)A\nabla_z\rho\cdot\nabla_z\rho. \tag{3.15}$$

Substituting (3.15) into (3.14), we obtain (3.13). ◻

PROPOSITION 3.2. *There exists a function* $\rho^\infty\in C([0,T];L^1(\mathbf{X}))$ *such that*

$$\lim_{n\to+\infty}\rho^n=\rho^\infty \quad in \quad C([0,T],L^1(\mathbf{X})),$$

and the following assertions hold
(1) $\sqrt{\rho^\infty}$ *solves the following equation*

$$\frac{\partial\sqrt{\rho^\infty}}{\partial t}-\mathsf{K}[\rho^\infty]\sqrt{\rho^\infty}=-\frac{\mathrm{Tr(A)}}{2}\sqrt{\rho^\infty}+\frac{\beta}{4}Az\cdot z\sqrt{\rho^\infty}. \tag{3.16}$$

(2) ρ^∞ *is a stationary solution to equation* (1.3) *and is of the form*

$$\rho^\infty(q,p,z)=\frac{\exp\Big[-\beta\Big(\frac{\|z\|^2}{2}+\frac{|p|^2}{2}+V(q)+U*h^\infty(q)\Big)\Big]}{\int_{\mathbf{X}}\exp\Big[-\beta\Big(\frac{\|z\|^2}{2}+\frac{|p|^2}{2}+V(q)+U*h^\infty(q)\Big)\Big]\,d\mathbf{x}}, \tag{3.17}$$

where h^∞ *satisfies*

$$h^\infty=\frac{\exp\Big[-\beta\Big(V(q)+U*h^\infty(q)\Big)\Big]}{\int_{\mathbb{R}^d}\exp\Big[-\beta\Big(V(q)+U*h^\infty(q)\Big)\Big]\,dq},$$

Proof. By Lemma 3.2, there exists a subsequence of $\{t_n\}$, which we denote with the same index, and a function $\rho^\infty\in C([0,T];L^1(\mathbf{X}))$ such that

$$\rho^n\xrightarrow{n\to\infty}\rho^\infty \quad \text{in} \quad C([0,T],L^1(\mathbf{X})).$$

As a consequence

$$\sqrt{\rho^n}\xrightarrow{n\to\infty}\sqrt{\rho^\infty} \quad \text{in} \quad L^2([0,T]\times\mathbf{X}).$$

We now derive the limiting equation for $\sqrt{\rho^\infty}$. To focus on the main idea, we provide a formal derivation here. Applying Lemma 3.3 to $\xi(\rho^n)=\sqrt{\rho^n}$ we obtain

$$\frac{\partial\sqrt{\rho^n}}{\partial t}-\mathsf{K}[\rho^n]\sqrt{\rho^n}=-\frac{\mathrm{Tr(A)}}{2}\sqrt{\rho^n}+\frac{\beta^{-1}}{4}\frac{A\nabla_z\rho^n\cdot\nabla_z\rho^n}{(\rho^n)^{3/2}}, \tag{3.18}$$

in the sense of distribution. We wish to pass to the limit $n\to\infty$ in this equation, for which we need appropriate compactness properties. Equation (3.12) implies that

$$\lim_{t\to\infty}\left\|\frac{\beta^2}{4}Az\cdot z\rho^n-A\nabla_z\sqrt{\rho^n}\cdot\nabla_z\sqrt{\rho^n}\right\|_{L^1([0,T]\times\mathbf{X})}=0.$$

In addition, similarly as in [67,68], we can prove that $\{\nabla F*\rho^n(q)\}$ converges uniformly on each compact set towards $\nabla F*\rho^\infty(q)$. We now pass to the limit $n\to\infty$ in equation (3.18) to get

$$\frac{\partial\sqrt{\rho^\infty}}{\partial t}-\mathsf{K}[\rho^\infty]\sqrt{\rho^\infty}=-\frac{\mathrm{Tr(A)}}{2}\sqrt{\rho^\infty}+\frac{\beta}{4}Az\cdot z\sqrt{\rho^\infty},$$

which proves (3.16). This derivation can be made rigorous following along the lines of [1, Proof of Theorem 1.2] by applying Lemma 3.3 to $\xi_\varepsilon(\rho^n)=\sqrt{\rho^n+\varepsilon}-\sqrt{\varepsilon}$ instead of the square root function; then passing first to $n\to\infty$ and then to $\varepsilon\to 0$. We omit the details here and proceed to the second part of the proposition. From (3.12), we get that

$$\frac{\beta}{2}z\sqrt{\rho^\infty}+\nabla_z\sqrt{\rho^\infty}=0, \tag{3.19}$$

in the sense of distributions on $[0,T]\times\mathbf{X}$. Multiplying (3.19) with $\exp\left(\frac{\beta\|z\|^2}{4}\right)$ we obtain

$$\nabla_z\left[\exp\left(\frac{\beta\|z\|^2}{4}\right)\sqrt{\rho^\infty}\right]=0,$$

in the sense of distribution on $[0,T]\times\mathbf{X}$, which implies that there exists a function $g^\infty(t,q,p)\in L^1_{\text{loc}}([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$ such that

$$\sqrt{\rho^\infty}=\sqrt{g^\infty(t,q,p)}\exp\left(-\frac{\beta\|z\|^2}{4}\right),\quad \text{i.e.,}\quad \rho^\infty(t,q,p,z)=g^\infty(t,q,p)\exp\left(-\frac{\beta\|z\|^2}{2}\right). \tag{3.20}$$

Substituting this representation to equation (3.16), we obtain the following equation for $\sqrt{g^\infty}$

$$\begin{aligned}\frac{\partial\sqrt{g^\infty}}{\partial t}=&-\mathrm{div}_q\left(p\sqrt{g^\infty}\right)+\mathrm{div}_p\left((\nabla V(q)+\nabla_q U*g^\infty(q))\sqrt{g^\infty}\right)\\&-\lambda^T z\cdot\left(\nabla_p\sqrt{g^\infty}+\frac{\beta}{2}p\sqrt{g^\infty}\right).\end{aligned} \tag{3.21}$$

Since g^∞ is independent of z, we must have

$$\frac{\beta}{2}p\sqrt{g^\infty}+\nabla_p\sqrt{g^\infty}=0.$$

This implies that there is a function $h^\infty(t,q)\in L^1_{\text{loc}}([0,T]\times\mathbb{R}^d)$ such that

$$g^\infty(t,q,p)=h^\infty(t,q)\exp\left(-\frac{\beta|p|^2}{2}\right). \tag{3.22}$$

Substituting this back into (3.21), we get

$$\frac{\partial\sqrt{h^\infty}}{\partial t}=-p\cdot\left(\nabla_q\sqrt{h^\infty}+\frac{\beta}{2}(\nabla V(q)+\nabla U*h^\infty(q))\sqrt{h^\infty}\right).$$

Since h^∞ is independent of p, we must have that

$$\frac{\partial\sqrt{h^\infty}}{\partial t}=\nabla_q\sqrt{h^\infty}+\frac{\beta}{2}(\nabla V(q)+\nabla U*h^\infty(q))\sqrt{h^\infty}=0,$$

which implies that h^∞ solves

$$h^\infty=\frac{\exp\left[-\beta(V(q)+U*h^\infty(q))\right]}{\int_{\mathbb{R}^d}\exp\left[-\beta(V(q)+U*h^\infty(q))\right]dq}. \tag{3.23}$$

Substituting (3.23) and (3.22) back to (3.20), we obtain that

$$\rho^\infty(q,p,z) = \frac{\exp\Big[-\beta\Big(\frac{\|z\|^2}{2}+\frac{|p|^2}{2}+V(q)+U*h^\infty(q)\Big)\Big]}{\int_{\mathbf{X}}\exp\Big[-\beta\Big(\frac{\|z\|^2}{2}+\frac{|p|^2}{2}+V(q)+U*h^\infty(q)\Big)\Big]\,d\mathbf{x}},$$

where h^∞ solves (3.23). According to Proposition 3.1, ρ^∞ is a stationary solution to equation (1.3). □

4. GENERIC formulation of the GLMV

In this section, we recast the overdamped, underdamped and generalized McKean-Vlasov equations into the GENERIC (General Equation for Non-Equilibrium Reversible-Irreversible Coupling) formalism, thus putting them in a common framework. In addition, using the GENERIC formulation, we can rederive formulas for their stationary states obtained in Section 3.

4.1. The GENERIC framework. As suggested by its name, in the GENERIC framework, an evolution equation for an unknown z in a state space Z is decomposed into the sum of a reversible dynamics and an irreversible dynamics as follows

$$\frac{\partial \mathsf{z}}{\partial t} = \underbrace{\mathsf{L}(\mathsf{z})\frac{\delta \mathsf{E}(\mathsf{z})}{\delta \mathsf{z}}}_{\text{reversible dynamics}} + \underbrace{\mathsf{M}(\mathsf{z})\frac{\delta \mathsf{S}(\mathsf{z})}{\delta \mathsf{z}}}_{\text{irreversible dynamics}}. \tag{4.1}$$

In the above equation, L,M are two operators, $\mathsf{E},\mathsf{S}:\mathsf{Z}\to\mathbb{R}$ are two functionals that are respectively called energy and entropy functionals, $\frac{\delta\mathsf{E}}{\delta\mathsf{z}},\frac{\delta\mathsf{S}}{\delta\mathsf{z}}$ are appropriate derivatives of E and S (such as either the Fréchet derivative or a gradient with respect to some inner product). The GENERIC framework imposes the following conditions on $\{\mathsf{L},\mathsf{M},\mathsf{E},\mathsf{S}\}$, which are called the GENERIC building blocks,

(C1) $\mathsf{L}=\mathsf{L}(\mathsf{z})$ is, for each z, an antisymmetric operator satisfying the Jacobi identity

$$\{\{\mathsf{F}_1,\mathsf{F}_2\}_\mathsf{L},\mathsf{F}_3\}_\mathsf{L}+\{\{\mathsf{F}_2,\mathsf{F}_3\}_\mathsf{L},\mathsf{F}_1\}_\mathsf{L}+\{\{\mathsf{F}_3,\mathsf{F}_1\}_\mathsf{L},\mathsf{F}_2\}_\mathsf{L}=0, \tag{4.2}$$

for all functions $\mathsf{F}_i:\mathsf{Z}\to\mathbb{R}$, $i=1,2,3$, where the Poisson bracket $\{\cdot,\cdot\}_\mathsf{L}$ is defined via

$$\{\mathsf{F},\mathsf{G}\}_\mathsf{L} := \frac{\delta\mathsf{F}}{\delta\mathsf{z}}\cdot\mathsf{L}\frac{\delta\mathsf{G}}{\delta\mathsf{z}}. \tag{4.3}$$

(C2) $\mathsf{M}=\mathsf{M}(\mathsf{z})$ is symmetric and positive semidefinite.

(C3) The following *degeneracy condition* holds

$$\mathsf{L}\frac{\delta\mathsf{S}}{\delta\mathsf{z}}=0,\quad \mathsf{M}\frac{\delta\mathsf{E}}{\delta\mathsf{z}}=0. \tag{4.4}$$

Being formulated in the GENERIC framework, an evolution equation automatically justifies the first and second laws of thermodynamics, i.e., along its solutions, energy is conserved while entropy is nondecreasing. Indeed, they are direct consequences of the above conditions:

$$\frac{d\mathsf{E}(\mathsf{z}(t))}{dt}=\frac{\delta\mathsf{E}}{\delta\mathsf{z}}\cdot\frac{d\mathsf{z}}{dt}=\frac{\delta\mathsf{E}}{\delta\mathsf{z}}\cdot\left(\mathsf{L}\frac{\delta\mathsf{E}}{\delta\mathsf{z}}+\mathsf{M}\frac{\delta\mathsf{S}}{\delta\mathsf{z}}\right)=\frac{\delta\mathsf{E}}{\delta\mathsf{z}}\cdot\mathsf{L}\frac{\delta\mathsf{E}}{\delta\mathsf{z}}+\frac{\delta\mathsf{E}}{\delta\mathsf{z}}\cdot\mathsf{M}\frac{\delta\mathsf{S}}{\delta\mathsf{z}}=0,$$

$$\frac{d\mathsf{S}(\mathsf{z}(t))}{dt}=\frac{\delta\mathsf{S}}{\delta\mathsf{z}}\cdot\frac{d\mathsf{z}}{dt}=\frac{\delta\mathsf{S}}{\delta\mathsf{z}}\cdot\left(\mathsf{L}\frac{\delta\mathsf{E}}{\delta\mathsf{z}}+\mathsf{M}\frac{\delta\mathsf{S}}{\delta\mathsf{z}}\right)=\frac{\delta\mathsf{S}}{\delta\mathsf{z}}\cdot\mathsf{M}\frac{\delta\mathsf{S}}{\delta\mathsf{z}}\geq 0.$$

In addition, in the GENERIC framework, equilibria can be obtained by the maximum entropy principle: they are the maximizers of the entropy S under the constraint that the energy is constant $\mathsf{E}(\mathsf{z}) = \mathsf{E}_0$ [46,53]. Define $\Phi(\mathsf{z}) = \mathsf{S}(\mathsf{z}) - \lambda(\mathsf{E}(\mathsf{z}) - \mathsf{E}_0)$ for some $\lambda \in \mathbb{R}$, which plays the role of a Lagrange multiplier. The states z_∞ that maximize the entropy under the constraint $\mathsf{E}(\mathsf{z}) = E_0$ are solutions to

$$\begin{cases} \frac{\delta\Phi(\mathsf{z})}{\delta\mathsf{z}} = 0, \\ \mathsf{E}(\mathsf{z}) = \mathsf{E}_0, \end{cases}$$

which is equivalent to

$$\begin{cases} \frac{\delta\mathsf{S}(\mathsf{z})}{\delta\mathsf{z}} = \lambda\frac{\delta\mathsf{E}(\mathsf{z})}{\delta\mathsf{z}}, \\ \mathsf{E}(\mathsf{z}) = \mathsf{E}_0. \end{cases} \tag{4.5}$$

Furthermore, since $\Phi(z_t)$ is also a nondecreasing function of time, it follows that all solutions z_t of the GENERIC equation (4.1) converge to z_∞ as $t \to \infty$. We refer to [53, Property 3] for a more detailed discussion.

In the next sections, we recast the MV dynamics, the VFP dynamics and the GLMV dynamics into the GENERIC framework and we use equation (4.5) to establish formulas for their stationary solutions and recovering the results in the previous section.

4.2. GENERIC formulation of the generalized McKean-Vlasov equation. It has been known that both the overdamped and underdamped McKean-Vlasov equations are Wasserstein gradient flows [7,8]. More recently, in [15], it was shown that the underdamped McKean-Vlasov equation can be cast into the GENERIC framework. In this section, we apply the technique therein for the generalized McKean-Vlasov equation. To this end, we need to construct the building block $\{\mathsf{Z},\mathsf{L},\mathsf{M},\mathsf{E},\mathsf{S}\}$ and verify Conditions $(C1)-(C3)$ given in the previous section.

Let ρ be a smooth solution of the the generalized McKean-Vlasov equation (1.3). We define

$$\mathcal{H}(\rho) := \int \Big(\frac{|p|^2}{2} + V(q) + \frac{1}{2}(U*\rho) + \frac{\|z\|^2}{2}\Big)\rho\, d\mathbf{x}. \tag{4.6}$$

Let us introduce an auxiliary time-dependent variable e whose evolution is given by

$$\frac{d}{dt}e = -\frac{d}{dt}\mathcal{H}(\rho_t) = \int Az\cdot z\rho\, d\mathbf{x} - \beta^{-1}\mathrm{Tr}(\mathrm{A}). \tag{4.7}$$

We now show that the coupled system for the variables (ρ, e),

$$\partial_t\begin{pmatrix}\rho\\ e\end{pmatrix} = \begin{pmatrix}\mathrm{div}(\rho\mathbb{J}\nabla\mathbb{H}_\rho) + \beta^{-1}\mathrm{div}(\rho A\nabla_z\log\rho) + \mathrm{div}(\rho Az)\\ -\beta^{-1}\mathrm{Tr}(\mathrm{A}) + \int \mathrm{Az}\cdot\mathrm{z}\rho\, d\mathbf{x}\end{pmatrix}, \tag{4.8}$$

can be formulated in the GENERIC framework. We construct the building blocks as follows

$$\mathsf{z} = (\rho, e), \quad \mathsf{Z} = \mathcal{P}_2(\mathbf{X}) \times \mathbb{R}, \quad \mathsf{E}(\rho, e) = \mathcal{H}(\rho) + e, \quad \mathsf{S}(\rho, e) = -\beta^{-1}\int \rho\log\rho + e,$$

$$\mathsf{L}(\rho, e) = \begin{pmatrix}\mathsf{L}_{\rho\rho} & 0\\ 0 & 0\end{pmatrix}, \quad \text{and} \quad \mathsf{M}(\rho, e) = \begin{pmatrix}\mathsf{M}_{\rho\rho} & \mathsf{M}_{\rho e}\\ \mathsf{M}_{e\rho} & \mathsf{M}_{ee},\end{pmatrix}$$

where, the operators defining L and M are given by

$$\mathsf{L}_{\rho\rho}\xi = \mathrm{div}\left(\rho\mathbb{J}\nabla\xi\right), \quad \mathsf{M}_{\rho\rho}\xi = -\mathrm{div}(\rho A\nabla_z\xi), \quad \mathsf{M}_{\rho e}r = r\,\mathrm{div}\left(\rho Az\right)$$
$$\mathsf{M}_{e\rho}\xi = -\int Az\cdot\nabla_z\xi\,\rho\,d\mathbf{x}, \quad \mathsf{M}_{ee}r = r\int Az\cdot z\,\rho\,d\mathbf{x}$$

with $\mathbb{J}$ being the antisymmetric matrix defined in 3.6b. We compute the first variations of E and S:

$$\frac{\delta\mathsf{E}}{\delta\mathsf{z}} = \begin{pmatrix} V(q)+\frac{|p|^2}{2}+U*\rho+\frac{\|z\|^2}{2} \\ 1\end{pmatrix} = \begin{pmatrix}\mathbb{H}_\rho\\ 1\end{pmatrix}, \qquad \frac{\delta\mathsf{S}}{\delta\mathsf{z}} = \begin{pmatrix}-\beta^{-1}(\log\rho+1)\\ 1\end{pmatrix}. \tag{4.9}$$

Thus the GENERIC equation associated with these blocks is given by

$$\begin{aligned}\partial_t\begin{pmatrix}\rho\\ e\end{pmatrix} &= \mathsf{L}\frac{\delta\mathsf{E}}{\delta\mathsf{z}}+\mathsf{M}\frac{\delta\mathsf{S}}{\delta\mathsf{z}} = \begin{pmatrix}\mathsf{L}_{\rho\rho}\mathbb{H}_\rho\\ 0\end{pmatrix} + \begin{pmatrix}-\beta^{-1}\mathsf{M}_{\rho\rho}(\log\rho+1)+\mathsf{M}_{\rho e}1\\ -\beta^{-1}\mathsf{M}_{e\rho}(\log\rho+1)+\mathsf{M}_{ee}1\end{pmatrix}\\
&= \begin{pmatrix}\mathrm{div}(\rho\mathbb{J}\nabla\mathbb{H}_\rho)\\ 0\end{pmatrix} + \begin{pmatrix}\beta^{-1}\mathrm{div}(\rho A\nabla_z\log\rho)+\mathrm{div}(\rho Az))\\ \beta^{-1}\int\rho Az\cdot\nabla_z\log\rho\,d\mathbf{x}+\int Az\cdot z\rho\,d\mathbf{x}\end{pmatrix}\\
&= \begin{pmatrix}\mathrm{div}(\rho\mathbb{J}\nabla\mathbb{H}_\rho)+\beta^{-1}\mathrm{div}(\rho A\nabla_z\log\rho)+\mathrm{div}(\rho Az)\\ \beta^{-1}\int\rho Az\cdot\nabla_z\log\rho\,d\mathbf{x}+\int Az\cdot z\rho\,d\mathbf{x}\end{pmatrix},\end{aligned}$$

which is indeed the system (4.8). By straight but lengthy computations, one can verify all the conditions imposed on the GENERIC framework for the building blocks constructed above.

4.3. Invariant measures. We now look for stationary solutions of the coupled system (4.8) using (4.5). A stationary solution $\mathsf{z}_\infty = (\rho_\infty, e_\infty)$, where ρ_∞ is a probability measure, of the system (1.3) & (4.7) maximizes the entropy $\mathsf{S}(\mathsf{z})$ under the constraints that $\mathsf{E}(\mathsf{z})$ is a constant and that $\int\rho(dqdp)=1$. Define $\Phi(\mathsf{z}) = \mathsf{S}(\mathsf{z}) - \lambda_1(\mathsf{E}(\mathsf{z})-\mathsf{E}_0) - \lambda_2(\int_{\mathbb{R}^{2d}}\rho(dqdp)-1)$ for some $\lambda_1,\lambda_2\in\mathbb{R}$, then (ρ_∞, e_∞) satisfies the following equation

$$\begin{cases}\frac{\delta\Phi(\mathsf{z})}{\delta\mathsf{z}} = 0,\\ \mathsf{E}(\mathsf{z}) = \mathsf{E}_0,\\ \int\rho_\infty(dqdpdz) = 1.\end{cases} \tag{4.10}$$

Using the computations in (4.9), we have

$$\frac{\delta\Phi(\mathsf{z})}{\delta\mathsf{z}} = \frac{\delta\mathsf{S}(\mathsf{z})}{\delta\mathsf{z}} - \lambda_1\frac{\delta\mathsf{E}(\mathsf{z})}{\delta\mathsf{z}} - \begin{pmatrix}\lambda_2\\ 0\end{pmatrix} = \begin{pmatrix}-\beta^{-1}(\log\rho+1)-\lambda_1\mathbb{H}_\rho-\lambda_2\\ 1-\lambda_1\end{pmatrix}.$$

Therefore, equation (4.10) becomes

$$\begin{cases}-\beta^{-1}(\log\rho_\infty+1) = \lambda_1\mathbb{H}_{\rho_\infty}+\lambda_2,\\ 1=\lambda_1,\\ \mathbb{H}_\rho+e=\mathsf{E}_0,\\ \int\rho_\infty(dqdp)=1.\end{cases}$$

Substituting $\lambda_1=1$ into the first equation and then combining with the last equation, we obtain that ρ_∞ satisfies

$$\rho_\infty = \frac{\exp\left(-\beta\mathbb{H}_{\rho_\infty}\right)}{\int\exp\left(-\beta\mathbb{H}_{\rho_\infty}\right)d\mathbf{x}} = \frac{\exp\left(-\beta(V(q)+\frac{|p|^2}{2}+U*\rho+\frac{\|z\|^2}{2})\right)}{\int\exp\left(-\beta(V(q)+\frac{|p|^2}{2}+U*\rho+\frac{\|z\|^2}{2})\right)}.$$

This is exactly equation (3.3) obtained in Section 3. Similarly, we can derive the formula in Corollary 3.1 for a stationary solution ρ_∞^{VFP} of the VFP dynamics using its GENERIC formulation in [15].

5. White noise (Markovian) limits

In this section, we derive the underdamped McKean-Vlasov dynamics from the white noise (Markovian) limit of the generalized McKean-Vlasov dynamics both for the particle system and for the mean-field PDE. We apply the formal perturbation expansions method developed in [59, 60].

5.1. From the generalized to the underdamped Langevin equations. We will derive (2.2) from (2.3) in the limit of vanishing correlation time of the noise, which corresponds to rescaling λ and A in (2.3) according to $\lambda \mapsto \lambda/\varepsilon$ and $A \mapsto A/\varepsilon^2$. The SDE (2.3) becomes

$$dQ_i^\varepsilon = P_i^\varepsilon\, dt, \tag{5.1a}$$

$$dP_i^\varepsilon = -\nabla V(Q_i^\varepsilon)\, dt - \frac{1}{N}\sum_{j=1}^N \nabla U(Q_i^\varepsilon - Q_j^\varepsilon)\, dt + \frac{1}{\varepsilon}\lambda^T Z_i^\varepsilon\, dt, \tag{5.1b}$$

$$dZ_i^\varepsilon = -\frac{1}{\varepsilon}P_i^\varepsilon \lambda\, dt - \frac{1}{\varepsilon^2}AZ_i\, dt + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}A}dW_i. \tag{5.1c}$$

PROPOSITION 5.1. *Let $\{Q_i^\varepsilon, P_i^\varepsilon, Z_i^\varepsilon\}$ be the solution of (5.1), and assume that the matrix A is invertible. Then $(Q_i^\varepsilon(t), P_i^\varepsilon(t))$ converges weakly to the solution of the Vlasov-Fokker-Planck equation (2.2), where the friction coefficient is given by the formula*

$$\gamma = \langle \lambda, A^{-1}\lambda\rangle. \tag{5.2}$$

Proof. We rewrite (5.1) in a compact matrix form

$$d\hat{Q}^\varepsilon = \hat{P}^\varepsilon\, dt, \tag{5.3a}$$

$$d\hat{P}^\varepsilon = F(\hat{Q})\, dt + \frac{1}{\varepsilon}\lambda^T \hat{Z}^\varepsilon\, dt, \tag{5.3b}$$

$$d\hat{Z}^\varepsilon = -\frac{1}{\varepsilon}\hat{P}^\varepsilon \lambda\, dt - \frac{1}{\varepsilon^2}A\hat{Z}^\varepsilon\, dt + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}A}d\hat{W}. \tag{5.3c}$$

where we have used the following notation

$$\hat{Q}^\varepsilon = (Q_1^\varepsilon, \dots, Q_N^\varepsilon)^T, \quad \hat{P}^\varepsilon = (P_1^\varepsilon, \dots, P_N^\varepsilon)^T, \quad \hat{Z}^\varepsilon = (Z_1^\varepsilon, \dots, Z_N^\varepsilon)^T, \quad \hat{W} = (W_1, \dots, W_N)^N,$$

and

$$F(\hat{Q}^\varepsilon) = \Big(-\nabla V(Q_1^\varepsilon) - \frac{1}{N}\sum_{j=1}^N \nabla U(Q_1^\varepsilon - Q_j^\varepsilon), \dots, -\nabla V(Q_N^\varepsilon) - \frac{1}{N}\sum_{j=1}^N \nabla U(Q_N^\varepsilon - Q_j^\varepsilon)\Big)^T.$$

The backward Kolmogorov equation associated with (5.3) is

$$\frac{\partial u^\varepsilon}{\partial t} = \Big(\frac{1}{\varepsilon^2}\mathscr{L}_0 + \frac{1}{\varepsilon}\mathscr{L}_1 + \mathscr{L}_2\Big)u, \tag{5.4}$$

with

$$\mathscr{L}_0 = -A\hat{z}\cdot\nabla_{\hat{z}} + \beta^{-1}A : D_{\hat{z}}^2,$$

$$\mathscr{L}_1 = \lambda^T \hat{z}\cdot\nabla_{\hat{p}} - \hat{p}\lambda\cdot\nabla_{\hat{z}},$$
$$\mathscr{L}_2 = \hat{p}\cdot\nabla_{\hat{q}} + F(\hat{q})\cdot\nabla_{\hat{p}}.$$

We seek a solution to (5.4) in the form of a power series expansion in ε:

$$u^\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots.$$

By substituting this expansion into (5.4) and equating powers of ε, we obtain the following sequence of equations

$$\mathscr{L}_0 u_0 = 0, \tag{5.5a}$$
$$-\mathscr{L}_0 u_1 = \mathscr{L}_1 u_0, \tag{5.5b}$$
$$-\mathscr{L}_0 u_2 = \mathscr{L}_1 u_1 + \mathscr{L}_2 u_0 - \frac{\partial u_0}{\partial t}, \tag{5.5c}$$
$$\dots = \dots.$$

It follows from the first equation (5.5a) that to leading order, the solution of the Kolmogorov equation is independent of the auxiliary variables $\hat{z}$, $u_0 = u(\hat{q},\hat{p},t)$. The solvability of the second equation (5.5b) reads

$$\int_{\mathbb{R}^{dmN}} \mathscr{L}_1 u_0 \, e^{-\frac{\beta}{2}\|\hat{z}\|^2}\, d\hat{z} = 0,$$

which is fulfilled since $\mathscr{L}_1 u_0 = \lambda^T \hat{z}\cdot\nabla_{\hat{p}} u$, thus

$$\int_{\mathbb{R}^{dmN}} \mathscr{L}_1 u_0 e^{-\frac{\beta}{2}\|\hat{z}\|^2}\, d\hat{z} = \int_{\mathbb{R}^{dmN}} \left(\lambda^T \hat{z}\cdot\nabla_{\hat{p}} u\right) e^{-\frac{\beta}{2}\|\hat{z}\|^2}\, d\hat{z} = 0.$$

Equation (5.5b) becomes

$$-\mathscr{L}_0 u_1 = \lambda^T \hat{z}\cdot\nabla_{\hat{p}} u$$

The solution to this equation is

$$u_1(q,p,t) = \hat{z} A^{-1}\lambda\cdot\nabla_{\hat{p}} u.$$

Note that in the above equation, we ignore the part of the solution that lies in the null space of $\mathscr{L}_0$ since it will not affect the limiting equation. The solvability condition for the third equation (5.5c) gives

$$\int \left(\mathscr{L}_1 u_1 + \mathscr{L}_2 u - \frac{\partial u}{\partial t}\right)(2\pi\beta^{-1})^{dmN} e^{-\frac{\beta}{2}\|\hat{z}\|^2}\, d\hat{z} = 0.$$

Since $\mathscr{L}_2 u - \frac{\partial u}{\partial t}$ is independent of $\hat{z}$, we deduce from the above solvability equation that

$$\frac{\partial u}{\partial t} = \mathscr{L}_2 u + \int (\mathscr{L}_1 u_1)(2\pi\beta^{-1})^{dmN} e^{-\frac{\beta}{2}\|\hat{z}\|^2}\, d\hat{z}. \tag{5.6}$$

We compute the second term on the right-hand side of (5.6) using the properties of Gaussian distributions:

$$\int (\mathscr{L}_1 u_1)(2\pi\beta^{-1})^{dmN} e^{-\frac{\beta}{2}\|\hat{z}\|^2}\, d\hat{z} = \langle \lambda, A^{-1}\lambda\rangle \Big(-\hat{p}\cdot\nabla_{\hat{p}} u + \beta^{-1}\Delta_{\hat{p}} u\Big).$$

Substituting this calculation back into (5.6), we obtain that u satisfies the following PDE

$$\frac{\partial u}{\partial t} = \hat{p}\cdot\nabla_{\hat{q}} u + F(\hat{q})\cdot\nabla_{\hat{p}} u + \langle \lambda, A^{-1}\lambda\rangle \Big(-\hat{p}\cdot\nabla_{\hat{p}} u + \beta^{-1}\Delta_{\hat{p}} u\Big).$$

This is exactly the backward Kolmogorov equation of the Vlasov-Fokker-Planck system (2.2) with $\gamma = \langle \lambda, A^{-1}\lambda\rangle$. ∎

5.2. From the generalized McKean-Vlasov equation to the underdamped Mckean-Vlasov equation at the mean-field level. Under the same rescaling, $\lambda \mapsto \lambda/\varepsilon$ and $A \mapsto A/\varepsilon^2$, as in the previous section, system (1.4) becomes

$$dQ^\varepsilon(t) = P^\varepsilon(t)\,dt, \tag{5.7a}$$

$$dP^\varepsilon(t) = -\nabla V(Q^\varepsilon(t))\,dt - \nabla_q U * \rho_t^\varepsilon(Q_t^\varepsilon)\,dt + \frac{1}{\varepsilon}\lambda^T Z^\varepsilon(t)\,dt, \tag{5.7b}$$

$$dZ^\varepsilon(t) = -\frac{1}{\varepsilon}P^\varepsilon(t)\lambda\,dt - \frac{1}{\varepsilon^2}AZ^\varepsilon(t) + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}A}dW(t). \tag{5.7c}$$

PROPOSITION 5.2. *Let $(Q^\varepsilon, P^\varepsilon, Z^\varepsilon)$ be the solution of (5.7), and assume that the matrix A is invertible. Then $(Q^\varepsilon, P^\varepsilon)$ converges weakly to the solution of the Vlasov-Fokker-Planck equation*

$$dQ(t) = P(t)\,dt, \tag{5.8a}$$

$$dP(t) = -\nabla V(Q(t))\,dt - \nabla_q U * \rho_t(Q(t))\,dt - \gamma P(t)\,dt + \sqrt{2\gamma\beta^{-1}}dW(t), \tag{5.8b}$$

where the friction coefficient is given by formula (5.2).

The Fokker-Planck equation associated with the McKean SDE (5.7) is

$$\begin{aligned}
\frac{\partial \rho^\varepsilon}{\partial t} &= \mathscr{L}^* \rho^\varepsilon \\
&= -p\cdot\nabla_q \rho^\varepsilon + (\nabla V(q) + \nabla_q U(q) * \rho_t^\varepsilon)\cdot\nabla_p \rho^\varepsilon + \frac{1}{\varepsilon}\Big(-\lambda^T z\cdot\nabla_p \rho^\varepsilon + p\lambda\cdot\nabla_z \rho^\varepsilon\Big) \\
&\qquad + \frac{1}{\varepsilon^2}\Big(\mathrm{div}_p(Az\rho^\varepsilon) + \beta^{-1}\mathrm{div}(A\nabla_p \rho^\varepsilon)\Big) \\
&:= \left(\mathscr{L}_2^* + \frac{1}{\varepsilon}\mathscr{L}_1^* + \frac{1}{\varepsilon^2}\mathscr{L}_0^*\right)\rho^\varepsilon.
\end{aligned} \tag{5.9}$$

According to Proposition 3.1, an invariant measure ρ_∞ of the equation (5.7), if it exists, satisfies

$$\rho_\infty(q,p,z) = \frac{\exp\left[-\beta\left(\frac{|p|^2}{2} + \frac{\|z\|^2}{2} + V(q) + U * \rho_\infty(q)\right)\right]}{\int \exp\left[-\beta\left(\frac{|p|^2}{2} + \frac{\|z\|^2}{2} + V(q) + U * \rho_\infty(q)\right)\right]dqdpdz}. \tag{5.10}$$

We define the function $f^\varepsilon(q,p,z,t)$ through

$$\rho^\varepsilon(q,p,z,t) = \rho_\infty(q,p,z) f^\varepsilon(q,p,z,t). \tag{5.11}$$

LEMMA 5.1. *The function $f^\varepsilon(q,p,z,t)$ defined in (5.11) satisfies the equation*

$$\begin{aligned}
\frac{\partial f^\varepsilon}{\partial t} &= -p\cdot\nabla_q f^\varepsilon + (\nabla V(q) + \nabla_q U(q) * (f^\varepsilon \rho_\infty))\cdot\nabla_p f^\varepsilon + \beta f^\varepsilon p\cdot\nabla U(q) * \rho_\infty(1 - f^\varepsilon) \\
&\qquad + \frac{1}{\varepsilon}\Big(-\lambda^T z\cdot\nabla_p f^\varepsilon + p\lambda\cdot\nabla_z f^\varepsilon\Big) + \frac{1}{\varepsilon^2}\Big(-Az\cdot\nabla_z f^\varepsilon + \beta^{-1}\mathrm{div}_z(A\nabla_z f^\varepsilon)\Big) \\
&=: \Big(\hat{\mathscr{L}}_2 + \frac{1}{\varepsilon}\hat{\mathscr{L}}_1 + \frac{1}{\varepsilon^2}\hat{\mathscr{L}}_0\Big) f^\varepsilon.
\end{aligned} \tag{5.12}$$

Proof. For simplicity of notation, we drop the superscript ε on f^ε. From the definition of f and ρ_∞, we compute

$$\frac{\nabla_z\rho}{\rho}=\frac{\nabla_z(f\rho_\infty)}{f\rho_\infty}=\frac{f\nabla_z\rho_\infty+\rho_\infty\nabla_z f}{f\rho_\infty}=\frac{-\beta f\rho_\infty z+\rho_\infty\nabla_z f}{f\rho_\infty}=-\beta z+\frac{\nabla_z f}{f}. \tag{5.13}$$

Therefore, we obtain

$$\begin{aligned}\mathscr{L}_0^*\rho=\mathscr{L}_0^*(\rho_\infty f)&=\mathrm{div}_z\left[A\rho\Big(z+\beta^{-1}\frac{\nabla_z\rho}{\rho}\Big)\right]\\ &\overset{(5.13)}{=}\beta^{-1}\mathrm{div}_z\Big(\rho_\infty A\nabla_z f\Big)\\ &=\beta^{-1}\Big(\nabla_z\rho_\infty\cdot A\nabla_z f+\rho_\infty\,\mathrm{div}_z(A\nabla_z f)\Big)\\ &=\rho_\infty\Big(-Az\cdot\nabla_z f+\beta^{-1}\mathrm{div}_z(A\nabla_z f)\Big).\end{aligned} \tag{5.14}$$

We proceed with $\mathscr{L}_1^*\rho$:

$$\begin{aligned}\mathscr{L}_1^*\rho=\mathscr{L}_1^*(f\rho_\infty)&=-\lambda^T z\cdot\nabla_p(f\rho_\infty)+p\lambda\cdot\nabla_z(f\rho_\infty)\\ &=\rho_\infty\Big(-\lambda^T z\cdot\nabla_p f+p\lambda\cdot\nabla_z f\Big)+f(-\lambda^T z\cdot\nabla_p\rho_\infty+p\lambda\cdot\nabla_z\rho_\infty)\\ &=\rho_\infty\Big(-\lambda^T z\cdot\nabla_p f+p\lambda\cdot\nabla_z f\Big).\end{aligned} \tag{5.15}$$

Finally we compute $\mathscr{L}_2^*\rho$:

$$\begin{aligned}\mathscr{L}_2^*\rho&=\mathscr{L}_2^*(f\rho_\infty)\\ &=-p\cdot\nabla_q(f\rho_\infty)+(\nabla V(q)+\nabla_q U(q)*\rho_t)\cdot\nabla_p(f\rho_\infty)\\ &=-p\cdot(f\nabla_q\rho_\infty+\rho_\infty\nabla_q f)+(\nabla V(q)+\nabla_q U(q)*\rho_t)\cdot(f\nabla_p\rho_\infty+\rho_\infty\nabla_p f)\\ &=\Big[-p\cdot\big(-\beta f(\nabla V(q)+\nabla U(q)*\rho_\infty)+\nabla_q f\big)+(\nabla V(q)+\nabla_q U(q)*\rho_t)\times\\ &\qquad\big(-\beta f p+\nabla_p f\big)\Big]\rho_\infty\\ &=\Big[-p\cdot\nabla_q f+(\nabla V(q)+\nabla_q U(q)*\rho_t)\cdot\nabla_p f+\beta f p\cdot\nabla U(q)*(\rho_\infty-\rho_t)\Big]\rho_\infty.\end{aligned} \tag{5.16}$$

Substituting (5.14), (5.15) and (5.16) into (5.9), we obtain

$$\begin{aligned}\frac{\partial f}{\partial t}&=-p\cdot\nabla_q f+(\nabla V(q)+\nabla_q U(q)*(f\rho_\infty))\cdot\nabla_p f+\beta f p\cdot\nabla U(q)*\rho_\infty(1-f)\\ &\quad+\frac{1}{\varepsilon}\Big(-\lambda^T z\cdot\nabla_p f+p\lambda\cdot\nabla_z f\Big)+\frac{1}{\varepsilon^2}\Big(-Az\cdot\nabla_z f+\beta^{-1}\mathrm{div}_z(A\nabla_z f)\Big),\end{aligned}$$

which is equation (5.12). This completes the proof of the lemma. ∎

We are now ready to prove Proposition 5.2.

Proof. **(Proof of Proposition 5.2.)** We use the perturbation expansion method similarly as in the proof of Proposition 5.1. We look for a solution to (5.12) of the form

$$f^\varepsilon=f_0+\varepsilon f_1+\varepsilon^2 f_2+\dots.$$

Substituting this expansion into (5.12) and equating powers of ε, we obtain the following sequence of equations

$$\hat{\mathscr{L}}_0 f_0=0, \tag{5.17a}$$

$$-\hat{\mathscr{L}}_0 f_1 = \hat{\mathscr{L}}_1 f_0, \tag{5.17b}$$

$$-\hat{\mathscr{L}}_0 f_2 = \hat{\mathscr{L}}_2 f_0 + \hat{\mathscr{L}}_1 f_1 - \frac{\partial f_0}{\partial t}, \tag{5.17c}$$

$$\dots = \dots$$

It follows from equation (5.17a) that f_0 is independent of z:

$$f_0 = f(q,p,t).$$

We compute the right-hand side of (5.17b):

$$\hat{\mathscr{L}}_1 f_0 = -\lambda^T z \cdot \nabla_p f.$$

Thus equation (5.17b) becomes

$$\hat{\mathscr{L}}_0 f_1 = \lambda^T z \cdot \nabla_p f.$$

This equation satisfies the solvability condition since $\lambda^T z \cdot \nabla_p f$ is orthogonal to the null space of $\hat{\mathscr{L}}_0^*$ which consists of functions of the form $e^{-\beta \frac{\|z\|^2}{2}} u(q,p)$. Therefore, it has a unique solution, up to a term in the null space of $\hat{\mathscr{L}}_0^*$,

$$f_1 = -z A^{-1} \lambda \cdot \nabla_p f.$$

From this expression we can compute

$$\hat{\mathscr{L}}_1 f_1 = \lambda^T A^{-1} \lambda \Big(\|z\|^2 \Delta_p f - p \cdot \nabla_p f \Big). \tag{5.18}$$

The solvability condition for equation (5.17c) is that its right-hand side must be orthogonal to the null of $\hat{\mathscr{L}}_0^*$, i.e.,

$$\int \Big(\hat{\mathscr{L}}_2 f + \hat{\mathscr{L}}_1 f_1 - \frac{\partial f}{\partial t} \Big) Z^{-1} e^{-\beta \frac{\|z\|^2}{2}} \, dz = 0. \tag{5.19}$$

Since $\hat{\mathscr{L}}_2 f - \frac{\partial f}{\partial t}$ does not depends on z, we conclude from (5.19) that

$$\frac{\partial f}{\partial t} = \hat{\mathscr{L}}_2 f + \int (\hat{\mathscr{L}}_1 f_1) Z^{-1} e^{-\beta \frac{\|z\|^2}{2}} \, dz. \tag{5.20}$$

Since f does not depend on z and due to the separability of variables in ρ_∞, the first term in the right-hand side of (5.20) can be simplified to

$$\hat{\mathscr{L}}_2 f = -p \cdot \nabla_q f + (\nabla V(q) + \nabla_q U(q) * (f \hat{\rho}_\infty)) \cdot \nabla_p f + \beta f p \cdot \nabla U(q) * \hat{\rho}_\infty (1 - f),$$

where $\hat{\rho}_\infty$ satisfies

$$\hat{\rho}_\infty(q,p) = \frac{\exp\Big[-\beta \Big(\frac{|p|^2}{2} + V(q) + U * \hat{\rho}_\infty(q) \Big) \Big]}{\int \exp\Big[-\beta \Big(\frac{|p|^2}{2} + V(q) + U * \hat{\rho}_\infty(q) \Big) \Big] \, dq dp}. \tag{5.21}$$

The second term in the right-hand side of (5.20) can be computed using (5.18)

$$\int (\hat{\mathscr{L}}_1 f_1) Z^{-1} e^{-\beta \frac{\|z\|^2}{2}} \, dz = \lambda^T A^{-1} \lambda \Big(\beta^{-1} \Delta_p f - p \cdot \nabla_p f \Big).$$

Substituting the above computation back to (5.20), we obtain that

$$\frac{\partial f}{\partial t} = -p\cdot\nabla_q f + (\nabla V(q) + \nabla_q U(q) * (f\hat{\rho}_\infty))\cdot\nabla_p f + \beta f p\cdot\nabla U(q) * \hat{\rho}_\infty(1-f) \\ + \lambda^T A^{-1}\lambda\Big(\beta^{-1}\Delta_p f - p\cdot\nabla_p f\Big).$$

We define $\hat{\rho}(q,p,t)$ through

$$\hat{\rho}(q,p,t) = f(q,p,t)\hat{\rho}_\infty(q,p).$$

Then analogously, as in Lemma 5.1, we conclude that $\hat{\rho}(q,p,t)$ satisfies the Fokker-Planck equation associated with the Vlasov-Fokker-Planck dynamics (5.8) with the friction coefficient γ given by $\gamma = \lambda^T A^{-1}\lambda$. We thus complete the proof of Proposition 5.2. □

6. Discussion and future work

In this paper, we studied the generalized McKean-Vlasov equation that arises as the mean field limit of a system of interacting non-Markovian Langevin equations, under the assumption that the memory in the system can be described by introducing a finite number of auxiliary variables. We provided explicit formulas for the fundamental solution of the McKean-Vlasov equation for quadratic confining and interaction potentials, we studied the form of stationary states and we showed that no additional stationary states appear due to the memory in the system. Furthermore, we showed, under appropriate assumptions on the confining and interaction potentials, an exponentially fast convergence to an equilibrium; we showed how the generalized McKean-Vlasov equation can be written in the GENERIC form and we studied the white noise limit.

As mentioned in the introduction section, the starting point of the present work is a system of weakly interacting diffusions in an extended phase space, where a finite number of auxiliary processes describe the memory of the system. Perhaps, a more natural starting point would be a system of weakly interacting particles that is coupled to one or more heat baths, e.g. the model studied in [22], but for weakly interacting nonlinear oscillators. Passing to the mean field limit and eliminating the heat bath variables are two operations that do not commute, in general. The rigorous study of the combined mean field and thermodynamic limits in, for example, a Kac-Zwanzig-type model [23,30] is an interesting and difficult problem that we will leave for further study.

Other possible extensions of the present work include: first, the rigorous proof of a propagation of chaos result for non-Markovian interacting Langevin equations in particular without the quasi-Markovianity assumption; second, the development of a complete existence and uniqueness of solutions theory for the generalized McKean-Vlasov equation; in addition, the study of non-Markovian interacting processes without the quasi-Markovianity assumption; and finally the study of the effect of coloured noise on the McKean-Vlasov dynamics, and in particular on the number and type of phase transitions.

Acknowledgments. GP is partially supported by the EPSRC under Grants No. EP/P031587/1, EP/L024926/1, and EP/L020564/1.

REFERENCES

[1] L. L. Bonilla, J. A. Carrillo, and J. Soler, *Asymptotic behavior of an initial-boundary value problem for the Vlasov-Poisson-Fokker-Planck system*, SIAM J. Appl. Math., 57(5):1343–1372, 1997. 3.2, 3.2, 3.2, 3.2

[2] F. Bouchut and J. Dolbeault, *On long time asymptotics of the Vlasov-Fokker-Planck equation and of the Vlasov-Poisson-Fokker-Planck system with Coulombic and Newtonian potentials*, Differential Integral Equations, 8(3):487–514, 1995. 3.2

[3] F. Bolley, A. Guillin, and F. Malrieu, *Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation*, M2AN Math. Model. Numer. Anal., 44(5):867–884, 2010. 1

[4] V. I. Bogachev, N. V. Krylov, M. Röckner, and S. V. Shaposhnikov, *Fokker-Planck-Kolmogorov Equations*, Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 207, 2015. 1

[5] J. Binney and S. Tremaine, *Galactic Dynamics*, Second Edition, Princeton University Press, 2008. 1

[6] B. Chazelle, Q. Jiu, Q. Li, and C. Wang, *Well-posedness of the limiting equation of a noisy consensus model in opinion dynamics*, J. Diff. Eqs., 263(1):365–397, 2017. 1

[7] J. A. Carrillo, R. J. McCann, and C. Villani, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Rev. Mat. Iberoamericana, 19(3):971–1018, 12, 2003. 1, 4.2

[8] J. A. Carrillo, R. J. McCann, and C. Villani, *Contractions in the 2-Wasserstein length space and thermalization of granular media*, Arch. Ration. Mech. Anal., 179(2):217–263, 2006. 1, 4.2

[9] L. Chayes and V. Panferov, *The McKean-Vlasov equation in finite volume*, J. Stat. Phys., 138(1-3):351–380, 2010. 1

[10] D. A. Dawson, *Critical dynamics and fluctuations for a mean-field model of cooperative behavior*, J. Statist. Phys., 31(1):29–85, 1983. 1, 3, 3.1

[11] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, 902, 1997. 3.2

[12] D. A. Dawson and J. Gärtner, *Large deviations from the McKean-Vlasov limit for weakly interacting diffusions*, Stochastics, 20(4):247–308, 1987. 1

[13] M. H. Duong, A. Lamacz, M. A. Peletier, and U. Sharma, *Variational approach to coarse-graining of generalized gradient flows*, Calc. Var. Partial Differential Equations, 56(4):100, 2017. 1

[14] H. Dym and H. P. McKean, *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*, Probability and Mathematical Statistics, Academic Press, New York, 31, 1976. 1

[15] M. H. Duong, M. A. Peletier, and J. Zimmer, *GENERIC formalism of a Vlasov-Fokker-Planck equation and connection to large-deviation principles*, Nonlinearity, 26(11):2951–2971, 2013. 4.2, 4.3

[16] K. Dressler, *Stationary solutions of the Vlasov-Fokker-Planck equation*, Math. Methods Appl. Sci., 9(2):169–176, 1987. 3.1

[17] M. H. Duong and J. Tugaut, *Stationary solutions of the Vlasov-Fokker-Planck equation: existence, characterization and phase-transition*, Appl. Math. Lett., 52:38–45, 2016. 1, 3, 3.1

[18] M. H. Duong and J. Tugaut, *The Vlasov-Fokker-Planck equation in non-convex landscapes: convergence to equilibrium*, Electron. Commun. Probab., 23:10, 2018. 3, 3.2, 3.2

[19] M. H. Duong, *Long time behaviour and particle approximation of a generalised Vlasov dynamic*, Nonlinear Anal., 127:1–16, 2015. 1

[20] L. Desvillettes and C. Villani, *On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: the linear Fokker-Planck equation*, Comm. Pure Appl. Math., 54(1):1–42, 2001. 1

[21] R. C. Desai and R. Zwanzig, *Statistical mechanics of a nonlinear stochastic model*, J. Statist. Phys., 19(1):1–24, 1978. 1

[22] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet, *Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures*, Comm. Math. Phys., 201(3):657–697, 1999. 1, 1, 6

[23] G. W. Ford, M. Kac, and P. Mazur, *Statistical mechanics of assemblies of coupled oscillators*, J. Math. Phys., 6:504–515, 1965. 6

[24] W. H. Fleming, *Logarithmic transformations and stochastic control*, W.H. Fleming and L.G. Gorostiza (eds.), Advances in Filtering and Optimal Stochastic Control, Springer Berlin Heidelberg, 131–141, 1982.

[25] E. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, *How far from equilibrium is active matter?*, Phys. Rev. Lett., 117:038103, 2016. 2.2

[26] T. D. Frank, *Nonlinear Fokker-Planck Equations*, Springer Series in Synergetics, Springer-Verlag, Berlin, 2005. 1

[27] F. Farkhooi and W. Stannat, *Complete mean-field theory for dynamics of binary recurrent networks*, Phys. Rev. Lett., 119:208301, 2017. 2.2

[28] J. Gärtner, *On the McKean-Vlasov limit for interacting diffusions*, Math. Nachr., 137:197–248, 1988. 1

[29] M. H. Giga, Y. Giga, and J. Saal, *Nonlinear Partial Differential Equations: Asymptotic Behavior of Solutions and Self-Similar Solutions*, Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser Boston, 2010. 1

[30] D. Givon, R. Kupferman, and A. Stuart, *Extracting macroscopic dynamics: model problems and algorithms*, Nonlinearity, 17(6):R55–R127, 2004. 3.1

[31] B. D. Goddard, A. Nold, N. Savva, G. A. Pavliotis, and S. Kalliadasis, *General dynamical density functional theory for classical fluids*, Phys. Rev. Lett., 109:120603, 2012. 6

[32] S. N. Gomes and G. A. Pavliotis, *Mean field limits for interacting diffusions in a two-scale potential*, J. Nonlinear Sci., 28(3):905–941, 2018. 1

[33] B. D. Goddard, G. A. Pavliotis, and S. Kalliadasis, *The overdamped limit of dynamic density functional theory: rigorous results*, Multiscale Model. Simul., 10(2):633–663, 2012. 1

[34] J. Garnier, G. Papanicolaou, and T.-W. Yang, *Large deviations for a mean field model of systemic risk*, SIAM J. Financial Math., 4(1):151–184, 2013.

[35] J. Garnier, G. Papanicolaou, and T.-W. Yang, *Consensus convergence with stochastic effects*, Vietnam J. Math., 45(1-2):51–75, 2017. 1

[36] P. Hanggi and P. Jung, *Colored noise in dynamical systems*, in I Prigogine, and SA Rice (eds.), Advances in Chemical Physics, 89:239–326, 1995. 1

[37] W. Horsthemke and R. Lefever, *Noise-Induced Transitions*, Springer Series in Synergetics, Springer-Verlag, Berlin, 15, 1984. 1
1

[38] S. Herrmann and J. Tugaut, *Non-uniqueness of stationary measures for self-stabilizing processes*, Stochastic Process. Appl., 120(7):1215–1246, 2010. 1

[39] A. Kolmogoroff, *Zufallige bewegungen (zur theorie der brownschen bewegung)*, Ann. of Math., 35(1):116–117, 1934. 3.2

[40] R. Kupferman, *Fractional kinetics in Kac-Zwanzig heat bath models*, J. Statist. Phys., 114(1-2):291–326, 2004. 2.1.1

[41] L. Lorenzi and M. Bertoldi, *Analytical Methods for Markov Semigroups*, Chapman & Hall/CRC, 2007. 1

[42] A. Lindquist and G. Picci, *Realization theory for multivariate stationary Gaussian processes*, SIAM J. Control Optim., 23(6):809–857, 1985. 2.1, 2.1.1

[43] E. Luçon and W. Stannat, *Transition from Gaussian to non-Gaussian fluctuations for mean-field diffusions in spatial interaction*, Ann. Appl. Probab., 26(6):3840–3909, 2016. 1

[44] H. P. McKean, Jr., *A class of Markov processes associated with nonlinear parabolic equations*, Proc. Nat. Acad. Sci. U.S.A., 56:1907–1911, 1966. 1

[45] H. P. McKean, Jr., *Propagation of chaos for a class of non-linear parabolic equations*, in Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967), Air Force Office Sci. Res., Arlington, Va., 41–57, 1967. 1

[46] A. Mielke, *Formulation of thermoelastic dissipative material behavior using GENERIC*, Contin. Mech. Thermodyn., 23(3):233–256, 2011. 1, 3.2

[47] D. Mandal, K. Klymko, and M. R. DeWeese, *Entropy production and fluctuation theorems for active matter*, Phys. Rev. Lett., 119:258001, 2017. 4.1

[48] P. Monmarché, *Long-time behaviour and propagation of chaos for mean field kinetic particles*, Stochastic Process. Appl., 127(6):1721–1737, 2017. 1

[49] G. Metafune, D. Pallara, and E. Priola, *Spectrum of Ornstein-Uhlenbeck operators in L^p spaces with respect to invariant measures*, J. Funct. Anal., 196(1):40–60, 2002. 1

[50] S. Motsch and E. Tadmor, *Heterophilious dynamics enhances consensus*, SIAM Rev., 56(4):577–621, 2014. 2.1, 2.1.1, 2.1, 2.1.1

[51] H. D. Nguyen, *The small-mass limit and white-noise limit of an infinite dimensional generalized Langevin equation*, J. Stat. Phy., 173(2):411–437, 2018. 1

[52] K. Oelschläger, *A martingale approach to the law of large numbers for weakly interacting stochastic processes*, Ann. Probab., 12(2):458–479, 1984. 1

[53] H. C. Öttinger and M. Grmela, *Dynamics and thermodynamics of complex fluids. II. Illustrations of a general formalism*, Phys. Rev. E (3), 56(6):6633–6655, 1997. 1

[54] M. Ottobre and G. A. Pavliotis, *Asymptotic analysis for the generalized Langevin equation*, Nonlinearity, 24(5):1629–1653, 2011. 4.1, 4.1

[55] M. Ottobre, G. A. Pavliotis, and K. Pravda-Starov, *Exponential return to equilibrium for hypoelliptic quadratic systems*, J. Funct. Anal., 262(9):4000–4039, 2012. 1, 1, 1, 3.2

[56] M. Ottobre, G. A. Pavliotis, and K. Pravda-Starov, *Some remarks on degenerate hypoelliptic Ornstein-Uhlenbeck operators*, J. Math. Anal. Appl., 429(2):676–712, 2015. 2.1.1

[57] H. C. Öttinger, *Beyond Equilibrium Thermodynamics*, Wiley-Interscience, 2005. 2.1, 2.1.1, 2.1, 2.1

[58] M. Ottobre, *Asymptotic analysis for Markovian models in non-equilibrium statistical mechanics*,

PhD thesis, Imperial College London, 2011.

[59] G. A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Texts in Applied Mathematics, Springer, New York, 60, 2014. 1
1

[60] G. A. Pavliotis and A. M. Stuart, *Multiscale Methods: Averaging and Homogenization*, Texts in Applied Mathematics, Springer, New York, 53, 2008. 1, 1, 2.2, 5
1, 5

[61] L. Rey-Bellet, *Open classical systems*, in Open Quantum Systems. II, Lecture Notes in Math., Springer, Berlin, 1881:41–78, 2006. 1

[62] M. Shiino, *Dynamical behavior of stochastic systems of infinitely many coupled nonlinear oscillators exhibiting phase transitions of mean-field type: H theorem on asymptotic approach to equilibrium and critical slowing down of order-parameter fluctuations*, Phys. Rev. A, 36:2393–2412, 1987. 1

[63] I. Snook, *The Langevin and Generalized Langevin Approach to the Dynamics of Atomic, Polymeric and Colloidal Systems*, Elsevier, Amsterdam, 2007. 1

[64] Y. Tamura, *On asymptotic behaviors of the solution of a nonlinear diffusion equation*, J. Fac. Sci. Univ. Tokyo Sect. IA Math., 31(1):195–221, 1984. 3.1, 3.2

[65] Y. Tamura, *Free energy and the convergence of distributions of diffusion processes of McKean type*, J. Fac. Sci. Univ. Tokyo Sect. IA Math., 34(2):443–484, 1987. 2.1

[66] M. M. Tropper, *Ergodic and quasideterministic properties of finite-dimensional stochastic systems*, J. Statist. Phys., 17(6):491–509, 1977. 1

[67] J. Tugaut, *Convergence to the equilibria for self-stabilizing processes in double-well landscape*, Ann. Probab., 41(3A):1427–1460, 2013. 3, 3.2

[68] J. Tugaut, *Self-stabilizing processes in multi-wells landscape in $\mathbb{R}^d$-convergence*, Stochastic Process. Appl., 123(5):1780–1801, 2013. 3, 3.2

[69] J. Tugaut, *Phase transitions of McKean-Vlasov processes in double-wells landscape*, Stochastics, 86(2):257–284, 2014. 1

[70] C. Villani, *Hypocoercivity*, Mem. Amer. Math. Soc., 202(950):141, 2006. 1

[71] C. Wang, Q. Li, W. E, and B. Chazelle, *Noisy Hegselmann-Krause systems: phase transition and the $2R$-conjecture*, J. Stat. Phys., 166(5):1209–1225, 2017. 1