MEAN FIELD MODEL FOR COLLECTIVE MOTION BISTABILITY

Josselin Garnier*

Centre de Mathématiques Appliquées, Ecole Polytechnique
91128 Palaiseau Cedex, France

George Papanicolaou

Department of Mathematics, Stanford University
Stanford, CA 94305, USA

Tzu-Wei Yang

School of Mathematics, University of Minnesota
Minneapolis, MN 55455, USA

(Communicated by Yuan Lou)

Abstract. We consider the Czirók model for collective motion of locusts along a one-dimensional torus. In the model, each agent's velocity locally interacts with other agents' velocities in the system, and there is also exogenous randomness to each agent's velocity. The interaction tends to create the alignment of collective motion. By analyzing the associated nonlinear Fokker-Planck equation, we obtain the condition for the existence of stationary order states and the conditions for their linear stability. These conditions depend on the noise level, which should be strong enough, and on the interaction between the agent's velocities, which should be neither too small, nor too strong. We carry out the fluctuation analysis of the interacting system and describe the large deviation principle to calculate the transition probability from one order state to the other. Numerical simulations confirm our analytical findings.

1. **Introduction.** Collective motion has become an emerging topic in biology and statistical physics and has a large potential for the applications in other fields, for example, opinion dynamics in social science [28] and autonomous robotic networks in engineering [32]. Collective motion describes the systemic behavior of a large group of animals, insects or bacteria. In a collective motion model, the system has a large number of individuals, expressed as their locations and velocities, (x_i, u_i), $i = 1, \ldots, N$. The movement of each individual is affected by the movements of other individuals in the system, mostly attractive interactions. Most biologically realistic and mathematically interesting models suppose *local* interactions, that is, any individual is affected only by its neighbors within a limited range. In addition, a collective motion model generally adds independent, internal or external randomness to the individuals' movements.

Two approaches are widely used to model collective motion: self-propelled particle (SPP) models and coarse-grained (CG) models. In an SPP model (also known as an individual-based model), the system is viewed as a large number of coupled stochastic processes describing the evolutions of the individuals in the system. In an SPP model we can easily build an elaborated interaction between an individual and its neighbors to match actual biological observations. However, an SPP model consists of numerous nonlinear equations and is difficult to analyze. To obtain the analytical tractability, further simplifications are usually required [41, 19]; otherwise the system behavior has to be verified by numerical simulations. On the other hand, a CG model (also known as a macroscopic or mean-field model) assumes infinitely many individuals in the system and considers the evolution of the distribution of the individuals. Most CG models in the literature lead to hydrodynamic equations so that the system behavior can be analyzed. CG models have some limitations. Although some CG models [3, 14, 23, 4, 26] are indeed constructed from microscopic dynamics, it is generally difficult to relate the models to actual individuals' properties [1]. Another limitation of CG models is that noise present at the microscopic level can be averaged out by coarse graining. As a result most hydrodynamic equations only describe the deterministic evolution and cannot capture and analyze some important stochastic events like transitions from one order state to another order state observed for instance in [7]. It is possible to incorporate noise in the CG model in a phenomenological way [5] but it is important to model noise at the macroscopic level in a consistent way and to clarify what type of noise in the CG model is generated by noise in the SSP model.

In the recent years, there is a huge amount of literature concerning the modeling of collective motion so we mention only a few papers guiding our own work. Vicsek *et al.* [39] propose the scalar noise model that is currently the most well-known SPP model. The Czirók model [11] is a one-dimensional model on a torus. It is a simplified version of the Vicsek model and is used to describe the collective motion of locusts for instance (see [40, 1] and references therein). Buhl *et al.* [7] experimentally and numerically study the switching of the alignments of locusts and find that the probability of switching is exponentially decaying with the number of locusts. In [41, 19] the authors assume global and uniform interactions in the Czirók model and they derive the nonlinear Fokker-Planck equations for the average velocity and consider overall stochastic behavior. Some interesting features, like the existences of leaders or predators are considered in [10] and [30], respectively. Other various individual-based models related to the Czirók model are discussed in [29, 9, 31, 2], and [36, 18, 37, 3, 4, 26, 25, 38, 33, 8, 15, 34, 35] construct and analyze CG models of collective motion. These papers show that the linear stability analysis of the homogeneous solutions is a simple but powerful technique, and that the formation of "bands" (regions with high density of particles moving in the same direction) can be observed in numerical simulations and related to traveling solutions of CG equations.

One of the most important goals of the collective motion modeling is to mathematically reproduce and analyze the *ordering phenomenon* where the individuals show a coherence in their movements, such as swarms of locusts, flocks of birds, or schools of fish. Such an ordering phenomenon is called an order state of the system and there are generally multiple order states that represent different directions and rotations. The counterpart of the order states is the disorder state, where the system does not show such a pattern. From the perspective of statistical physics, in

this paper we address several interesting questions that are also considered in many of the literature:

1. Existence and number of order states.
2. Quantities that control the phase transition between the disorder state and the order states.
3. Transition probability from one order state to another order state.

Our contributions in this paper are the following. We interpret the Czirók model as an interacting particle system in the McKean-Vlasov framework [12, 13]. In this sense, we characterize the state of the system by the empirical probability measure of the locations and velocities of all the individuals, and as the number of individuals goes to infinity, the random, empirical measure converges to a deterministic, probability measure whose density is the solution of a nonlinear Fokker-Planck equation. Therefore, when the number of individuals is large, we can consider the Fokker-Planck equation instead of all the individuals. In this way, we are also able to provide a clear connection between the SPP model and the CG model. This procedure was proved to be very efficient to study the stationary states and their stability properties in the case of opinion dynamics models [21]. It turns out that it is still efficient to address collective motion models, although the framework is different: In the opinion dynamics model addressed in [21], the Fokker-Planck equation is non-degenerate while it is degenerate in the collective motion model that we addresse here, as the external noise affects only the velocities and not the locations. As a result noise plays a very different role which strongly affects the conditions for the existence and stability of the stationary states. For instance, the noise has to be large enough to ensure the existence of the non-trivial stationary states in the opinion dynamics model. In dramatic contrast, we will see that the existence of the order states in the collective motion model does not depend on the noise level and that their stability depends in a nontrivial way of the noise level. Noise has to be large enough to ensure the linear stability of the order states, but large noise also triggers frequent transitions from one order state to another one.

By analyzing the Fokker-Planck equation obtained in the Czirók model, we find that the system may have one disorder state and two order states (clockwise and counterclockwise rotations at constant velocity). We find the necessary and sufficient condition for the existence of the order states. This existence condition quantitatively describes the phase transition between the disorder state and the order states. We then perform the linear stability analysis to examine the stabilities of the order and disorder states, and we find that the order states are stable when the external noises are sufficiently strong. The noise threshold value depends in a nontrivial way of the interaction between individuals, the smallest threshold value is reached when the interaction is neither too weak, nor too strong. Such a phenomenon that the randomness of movement improves the alignment of collective motion coincides with the observation in [41]. When noise is small, the order states are unstable and the complex modulational instability generates moving clusters whose velocities and sizes can be identified. Finally, when the number of individuals is large but finite, the empirical measure is still stochastic and has a small probability to transit from one order state to the other one. This small probability can be described by the large deviation principle. It increases with the noise level. It is exponentially decaying as a function of the number of individuals; this fact is also observed in [7, 41]. We confirm our analysis by extensive numerical simulations.

The paper is organized as follows. The interacting particle model and its mean-field limit is presented in section 2. The equilibria of the nonlinear Fokker-Planck equation are analyzed in section 3. In section 4, the linearized stability analysis of the nonlinear Fokker-Planck equation is performed. We provide the fluctuation analysis and the large deviation principle in sections 5 and 6, respectively. In section 7, we verify our analytical findings by numerical simulations. We briefly consider the nonsymmetric Czirók model with normalized influence functions in section 8. We end with a brief summary and conclusions in section 9.

2. **Model and mean field limit.** We consider the Czirók model, a system of N agents moving along the torus $[0, L]$. The position x_i and velocity u_i of particle i satisfy the following stochastic differential equations: for $i = 1, \ldots, N$,

$$
\begin{aligned}
dx_i &= u_i dt, \\
du_i &= \left[G(\langle u \rangle_i) - u_i \right] dt + \sigma dW_i(t),
\end{aligned}
\tag{1}
$$

where $\{W_i(t)\}_{i=1}^N$ are independent Brownian motions and $\langle u \rangle_i$ is a weighted average of the velocities $\{u_j\}_{j=1}^N$:

$$
\langle u \rangle_i = \frac{1}{N} \sum_{j=1}^N u_j \phi(\|x_j - x_i\|),
\tag{2}
$$

with the weights depending on the distance $\|\cdot\|$ on the torus between the position x_i and the positions of the other agents (i.e. $\|x_j - x_i\| = \min(|x_j - x_i|, L - |x_j - x_i|)$). Here $\phi(x)$ is a nonnegative influence function normalized so that

$$
\frac{1}{L} \int_0^L \phi(\|x\|) dx = 1,
\tag{3}
$$

and $G(u)$ is an odd and smooth function. Note that there is no loss of generality in assuming (3) since we can always rescale G to ensure that this condition is satisfied. As we will see in this paper the interesting configuration is when $u \mapsto u - G(u)$ derives from an even double-well potential, with two symmetric global minima and one saddle point at zero. For instance, we can think at

$$
G(u) = 2 \tanh(u)
\tag{4}
$$

and then $u - G(u)$ derives from the double-well potential $u^2/2 - 2\log(\cosh(u))$, or

$$
G(u) = 2u - u^3
\tag{5}
$$

and then $u - G(u)$ derives from the double-well potential $u^4/4 - u^2/2$.

We define the empirical probability measure:

$$
\mu_N(t, dx, du) = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i(t), u_i(t))}(dx, du).
\tag{6}
$$

Let us assume that $\mu_N(0, dx, du)$ converges to a deterministic measure $\bar{\rho}(x, u) dx du$ as $N \to \infty$. This happens in particular when the positions and velocities of the agents at the inital time $t = 0$ are independent and identically distributed with the distribution with density $\bar{\rho}(x, u)$. Then, as $N \to \infty$, $\mu_N(t, dx, du)$ weakly converges to the deterministic measure $\rho(t, x, u) dx du$ whose density is the solution of the nonlinear Fokker-Planck equation

$$
\frac{\partial \rho}{\partial t} = -u \frac{\partial \rho}{\partial x} - \frac{\partial}{\partial u} \left\{ \left[G \left(\iint u' \phi(\|x'\|) \rho(t, x - x', u') du' dx' \right) - u \right] \rho \right\} + \frac{1}{2} \sigma^2 \frac{\partial^2 \rho}{\partial u^2}, \tag{7}
$$

starting from $\rho(t = 0, x, u) = \bar{\rho}(x, u)$. Note that the Fokker-Planck equation (7) is degenerate because there is no diffusion in x. The convergence proof is standard and can be found in [12, 22].

3. **Stationary analysis for order and disorder states.** We look for a stationary equilibrium, that is to say, a probability density function stationary in time:

$$- u\frac{\partial \rho}{\partial x} - \frac{\partial}{\partial u}\left\{\left[G\left(\iint u'\phi(\|x'\|)\rho(x - x', u')du'dx'\right) - u\right]\rho\right\} + \frac{1}{2}\sigma^2\frac{\partial^2 \rho}{\partial u^2} = 0, \quad (8)$$

with periodic boundary conditions in x.

Proposition 1. *The probability-density-valued solutions $\rho(x, u)$ to (8) have the following form:*

$$\rho_\xi(x, u) = \frac{1}{L}F_\xi(u), \quad F_\xi(u) = \frac{1}{\sqrt{\pi\sigma^2}}\exp\left(-\frac{(u - \xi)^2}{\sigma^2}\right). \quad (9)$$

They are uniform in space, Gaussian in velocity, and their mean velocity ξ satisfies the compatibility condition:

$$\xi = G(\xi). \quad (10)$$

There are, therefore, as many stationary equilibria as there are solutions to the compatibility equation (10).

Proof. It is straightforward to check that (9) is a solution of (8). Reciprocally, let ρ be a solution of (8). After integrating (8) from $u = -\infty$ to $u = +\infty$, we obtain

$$-\frac{\partial}{\partial x}\int_{-\infty}^{\infty} u\rho(x, u)du = 0,$$

which shows that $\int_{-\infty}^{\infty} u\rho(x, u)du = \xi/L$ is a constant independent of x. Then Eq. (8) can be written as a linear equation

$$- u\frac{\partial \rho}{\partial x} - \frac{\partial}{\partial u}\{[G(\xi) - u]\rho\} + \frac{1}{2}\sigma^2\frac{\partial^2 \rho}{\partial u^2} = 0. \quad (11)$$

Let $F(u) = \int_0^L \rho(x, u)dx$. Then $F(u)$ satisfies

$$-\frac{\partial}{\partial u}\{[G(\xi) - u]F(u)\} + \frac{1}{2}\sigma^2\frac{\partial^2 F(u)}{\partial u^2} = 0, \quad (12)$$

which shows that $F(u)$ is a Gaussian density function:

$$F(u) = \frac{1}{\sqrt{\pi\sigma^2}}\exp\left(-\frac{1}{\sigma^2}(u - G(\xi))^2\right). \quad (13)$$

The condition $\xi = L\int u\rho(x, u)du$ requires that

$$\xi = \frac{1}{L}\int_0^L \xi dx = \int_0^L \int u\rho(x, u)dudx = \int uF(u)du = G(\xi). \quad (14)$$

To show that $\rho(x, u)$ is uniform in x, we first note that $\rho(x, u)$ is periodic in x. We expand it as:

$$\rho(x, u) = \sum_{k=-\infty}^{\infty} \rho_k(u)e^{-i2\pi kx/L}, \quad \rho_k(u) = \frac{1}{L}\int_0^L \rho(x, u)e^{i2\pi kx/L}dx. \quad (15)$$

From (11) and (10), ρ_k satisfies

$$\frac{i2\pi k}{L}u\rho_k - \frac{\partial}{\partial u}[(\xi - u)\rho_k] + \frac{1}{2}\sigma^2\frac{\partial^2 \rho_k}{\partial u^2} = 0. \quad (16)$$

We take a Fourier transform in u:

$$\hat{\rho}_k(\eta) = \int_{-\infty}^{\infty} \rho_k(u) e^{-i\eta u} du. \tag{17}$$

From (16), $\hat{\rho}_k$ satisfies the ordinary differential equation

$$-\left(\frac{2\pi k}{L} + \eta\right)\frac{\partial\hat{\rho}_k}{\partial\eta} = \left(i\xi\eta + \frac{1}{2}\sigma^2\eta^2\right)\hat{\rho}_k. \tag{18}$$

This equation can be solved. Let $k < 0$. For any $\eta \in (-\infty, -2\pi k/L)$ we have

$$|\hat{\rho}_k(\eta)| = |\hat{\rho}_k(0)|\left|1 + \frac{L\eta}{2\pi k}\right|^{-\frac{2\pi^2 k^2 \sigma^2}{L^2}} \exp\left(-\frac{\sigma^2\eta^2}{4} + \frac{\pi k\sigma^2\eta}{L}\right),$$

which goes to $+\infty$ as $\eta \nearrow -2\pi k/L$ if $\hat{\rho}_k(0) \neq 0$. However $|\hat{\rho}_k(\eta)| \leqslant \int |\rho_k(u)| du \leqslant (1/L)\int\int_0^L \rho(x,u)dxdu = 1/L$ is uniformly bounded, which imposes $\hat{\rho}_k(0) = 0$ and therefore $\hat{\rho}_k(\eta) = 0$ for any $\eta \in (-\infty, -2\pi k/L)$. Similarly, for any $\eta \in (-2\pi k/L, \infty)$ we have

$$|\hat{\rho}_k(\eta)| = |\hat{\rho}_k(-4\pi k/L)|\left|1 + \frac{L\eta}{2\pi k}\right|^{-\frac{2\pi^2 k^2 \sigma^2}{L^2}} \exp\left(\left(-\frac{\sigma^2\eta}{4} + \frac{2\pi k\sigma^2}{L}\right)\left(\eta + \frac{4\pi k}{L}\right)\right),$$

which goes to $+\infty$ as $\eta \searrow -2\pi k/L$ if $\hat{\rho}_k(-4\pi k/L) \neq 0$. The boundedness of $\hat{\rho}_k$ imposes $\hat{\rho}_k(-4\pi k/L) = 0$ and therefore $\hat{\rho}_k(\eta) = 0$ for any $\eta \in (-2\pi k/L, \infty)$. Moreover, $\hat{\rho}_k$ is continuous by the Riemann-Lebesgue lemma and therefore $\hat{\rho}_k(\eta) = 0$ for any $\eta \in \mathbb{R}$ and $k < 0$. Since $\hat{\rho}_{-k}(-\eta) = \overline{\hat{\rho}_k(\eta)}$, we also find that $\hat{\rho}_k(\eta) = 0$ for any $\eta \in \mathbb{R}$ and $k > 0$. Therefore $\rho_k(u) = 0$ for any $u \in \mathbb{R}$ and $k \neq 0$ and the stationary solution is equal to $\rho(x,u) = \rho_0(u)$. $\square$

When G is such that $u - G(u)$ derives from an even double-well potential, such as the two examples (4) and (5), there are three ξ satisfying the compatibility condition (10): 0 and $\pm\xi_e$, with $\xi_e > 0$ the unique positive solution of $G(\xi) = \xi$. In the experiment, the equilibrium $\rho_0(x,u)$ is the disorder state and $\rho_{\pm\xi_e}(x,u)$ are the two order states of the locusts marching on the torus, the clockwise and counterclockwise rotations. The existence of the nonzero solution $\pm\xi_e$ and hence the stationary states $\rho_{\pm\xi_e}(x,u)$ is independent of the value of the noise strength σ. This observation is unusual and in fact inconsistent with the conclusion in many models in statistical physics and opinion dynamics [12, 21]. We will see, however, that the noise strength plays a crucial role in the stability of the stationary states.

4. **Linear stability analysis.** In this section, we assume that the compatibility condition (10) has solutions and we use the linear stability analysis to study the stability of the stationary states. Let ξ be such that $G(\xi) = \xi$ and consider

$$\rho(t,x,u) = \rho_\xi(x,u) + \rho^{(1)}(t,x,u) = \frac{1}{L}F_\xi(u) + \rho^{(1)}(t,x,u), \tag{19}$$

for small perturbation $\rho^{(1)}$. By linearizing the nonlinear Fokker-Planck equation (7) we find that $\rho^{(1)}$ satisfies

$$\frac{\partial\rho^{(1)}}{\partial t} = -u\frac{\partial\rho^{(1)}}{\partial x} - \frac{\partial}{\partial u}\left[(\xi - u)\rho^{(1)}\right] \tag{20}$$

$$-\frac{1}{L}G'(\xi)\left[\iint u'\phi(\|x'\|)\rho^{(1)}(t,x-x',u')du'dx'\right]F_\xi'(u) + \frac{1}{2}\sigma^2\frac{\partial^2\rho^{(1)}}{\partial u^2}.$$

Our goal is to clarify under which circumstances the linearized system is stable.

4.1. **Expressions of the linearized modes.** The perturbation $\rho^{(1)}$ is periodic in x. We expand it as:

$$\rho^{(1)}(t,x,u) = \sum_{k=-\infty}^{\infty} \rho_k^{(1)}(t,u)e^{-i2\pi kx/L}, \quad \rho_k^{(1)}(t,u) = \frac{1}{L}\int_0^L \rho^{(1)}(t,x,u)e^{i2\pi kx/L}dx. \tag{21}$$

The mode $\rho_k^{(1)}$ satisfies

$$\frac{\partial \rho_k^{(1)}}{\partial t} = \frac{i2\pi k}{L}u\rho_k^{(1)} - \frac{\partial}{\partial u}\left[(\xi - u)\rho_k^{(1)}\right] \tag{22}$$

$$- G'(\xi)\phi_k\left[\int u'\rho_k^{(1)}(t,u')du'\right]F_\xi'(u) + \frac{1}{2}\sigma^2\frac{\partial^2\rho_k^{(1)}}{\partial u^2},$$

where the discrete Fourier coefficients

$$\phi_k = \frac{1}{L}\int_0^L \phi(\|x\|)e^{i2\pi kx/L}dx \tag{23}$$

are real-valued (because $\|L - x\| = \|x\|$) and bounded by one (because $|\phi_k| \leqslant \int_0^L \phi(\|x\|)dx/L = 1$). We note that the equations are uncoupled in k and the system is linearly stable if all modes are stable. We then take a Fourier transform in u:

$$\hat{\rho}_k^{(1)}(t,\eta) = \int_{-\infty}^{\infty}\rho_k^{(1)}(t,u)e^{-i\eta u}du. \tag{24}$$

From (22), $\hat{\rho}_k^{(1)}$ satisfies the following first-order partial differential equation:

$$\frac{\partial}{\partial t}\hat{\rho}_k^{(1)} + \left(\frac{2\pi k}{L} + \eta\right)\frac{\partial}{\partial \eta}\hat{\rho}_k^{(1)} = \left(-i\xi\eta - \frac{1}{2}\sigma^2\eta^2\right)\hat{\rho}_k^{(1)} + G'(\xi)\phi_k\left[\frac{\partial}{\partial \eta}\hat{\rho}_k^{(1)}(t,0)\right]\eta\hat{F}_\xi, \tag{25}$$

starting from an initial condition that is assumed to satisfy

$$\|\hat{\rho}_k^{(1)}(0,\cdot)\|_{L^1} + \|\hat{\rho}_k^{(1)}(0,\cdot)\|_{L^\infty} + \|\partial_\eta\hat{\rho}_k^{(1)}(0,\cdot)\|_{L^\infty} < \infty.$$

We say that the kth-order mode is stable if

$$\sup_{t\geqslant 0}\left\{\|\hat{\rho}_k^{(1)}(t,\cdot)\|_{L^1} + \|\hat{\rho}_k^{(1)}(t,\cdot)\|_{L^\infty} + \|\partial_\eta\hat{\rho}_k^{(1)}(t,\cdot)\|_{L^\infty}\right\} < \infty.$$

By the method of characteristics, we can obtain the following implicit expression for $\hat{\rho}_k^{(1)}(t,\eta)$ that is useful for the stability analysis:

Lemma 4.1. *Let* $w_k(t) = \partial_\eta\hat{\rho}_k^{(1)}(t,0)$. $\hat{\rho}_k^{(1)}(t,\eta)$ *has the following expression:*

$$\hat{\rho}_k^{(1)}(t,\eta) = e^{-g_k(t,\eta)}\hat{\rho}_k^{(1)}\left(0, e^{-t}\eta - D_k(1-e^{-t})\right) \tag{26}$$

$$+ G'(\xi)\phi_k\int_0^t e^{-g_k(t-s,\eta)}w_k(s)H_\xi\left(e^{-(t-s)}\eta - D_k(1-e^{-(t-s)})\right)ds,$$

where $D_k = 2\pi k/L$, $H_\xi(\eta) = \eta\hat{F}_\xi(\eta)$ *and*

$$g_k(t,\eta) = \frac{\sigma^2}{4}(\eta + D_k)^2(1-e^{-2t}) + (i\xi - \sigma^2 D_k)(\eta + D_k)(1-e^{-t}) \tag{27}$$

$$+ \left(-i\xi D_k + \frac{\sigma^2}{2}D_k^2\right)t.$$

Proof. The function $\hat{\rho}_k^{(1)}(t, \eta)$ satisfies (25). We consider the mapping $\eta(t) = \eta_0 e^t - 2\pi k/L$ so that $\eta'(t) = \eta(t) + 2\pi k/L$. By the method of characteristics,

$$\frac{d}{dt}\hat{\rho}_k^{(1)}(t, \eta(t)) = \left(- i\xi\eta - \frac{1}{2}\sigma^2\eta^2\right)\hat{\rho}_k^{(1)}(t, \eta) + G'(\xi)\phi_k\left[\frac{\partial}{\partial\eta}\hat{\rho}_k^{(1)}(t, 0)\right]\eta\hat{F}_\xi(\eta).$$

Then by using the method of integrating factors, we have

$$\hat{\rho}_k^{(1)}(t, \eta(t)) = e^{- \int_0^t g(s)ds}\hat{\rho}_k^{(1)}(0, \eta_0 - 2\pi k/L)$$

$$+ G'(\xi)\phi_k \int_0^t e^{- \int_s^t g(w)dw}\left[\frac{\partial}{\partial\eta}\hat{\rho}_k^{(1)}(s, 0)\right]\eta(s)\hat{F}_\xi(\eta(s))ds,$$

where $g(t) = i\xi\eta(t) + \frac{1}{2}\sigma^2\eta^2(t)$. We note that $\int_s^t g(w)dw = g_k(t - s, \eta(t))$, where g_k is defined by (27) and $D_k = 2\pi k/L$. By letting $H_\xi(\eta) = \eta\hat{F}_\xi(\eta)$ and $w_k(s) = \partial_\eta\hat{\rho}_k^{(1)}(s, 0)$, $\hat{\rho}_k^{(1)}(t, \eta)$ can be written as (26). $\square$

4.2. **Stability of the** 0**th-order mode.** In this section we find a necessary and sufficient condition for the stability of the mode $k = 0$. For $k = 0$, $\rho_0^{(1)}$ satisfies

$$\frac{\partial\rho_0^{(1)}}{\partial t} = -\frac{\partial}{\partial u}\left[(\xi - u)\rho_0^{(1)}\right] - G'(\xi)\left[\int u'\rho_0^{(1)}(t, u')du'\right]F_\xi'(u) + \frac{1}{2}\sigma^2\frac{\partial^2\rho_0^{(1)}}{\partial u^2}. \quad (28)$$

We use the method of moments. We let $m_{0,0}^{(1)}$, $m_{0,1}^{(1)}$ denote the zeroth and first moments of $\rho_0^{(1)}$, respectively:

$$m_{0,0}^{(1)}(t) = \int \rho_0^{(1)}(t, u)du, \quad m_{0,1}^{(1)}(t) = \int u\rho_0^{(1)}(t, u)du. \quad (29)$$

From (28), we get

$$\frac{\partial}{\partial t}m_{0,0}^{(1)} = 0, \quad \frac{\partial}{\partial t}m_{0,1}^{(1)} = \xi m_{0,0}^{(1)} - m_{0,1}^{(1)} + G'(\xi)m_{0,1}^{(1)}.$$

The first moment $m_{0,1}^{(1)}$ is stable if and only if $G'(\xi) < 1$. More exactly, if $G'(\xi) = 1$ then the first moment grows linearly in time. If $G'(\xi) < 1$ then the first moment is bounded. Once $m_{0,1}^{(1)}$ has been shown to bounded, then the solution $\rho_0^{(1)}$ of (28) can be seen as the solution of a linear parabolic equation which is smooth and bounded as explained in the following proposition.

Proposition 2. *The 0th-order mode is stable if and only if $G'(\xi) < 1$.*

Proof. We know that $m_{0,1}^{(1)}(t)$ is uniformly bounded in t if and only if $G'(\xi) < 1$. From (26) in Lemma 4.1 and noting that $w_0(t) = \partial_\eta\hat{\rho}_0^{(1)}(t, 0) = -im_{0,1}^{(1)}(t)$, we have the following expression for $\hat{\rho}_0^{(1)}(t, \eta)$:

$$\hat{\rho}_0^{(1)}(t, \eta) = \hat{\rho}_0^{(1)}(0, \eta e^{-t})\exp\left(- i\xi\eta(1 - e^{-t}) - \frac{1}{4}\sigma^2\eta^2(1 - e^{-2t})\right) \quad (30)$$

$$- i\int_0^t F_0(\eta e^{-s})m_{0,1}^{(1)}(t - s)\exp\left(- i\xi\eta(1 - e^{-s}) - \frac{1}{4}\sigma^2\eta^2(1 - e^{-2s})\right)ds,$$

with $F_0(\eta) = G'(\xi)\eta\hat{F}_\xi(\eta)$. We find that, for any $t \geqslant 1$,

$$|\hat{\rho}_0^{(1)}(t,\eta)| \leqslant \|\hat{\rho}_0^{(1)}(0,\cdot)\|_{L^\infty} e^{-\frac{1}{4}\sigma^2\eta^2(1-e^{-2})}$$

$$+ \|m_{0,1}^{(1)}\|_{L^\infty} \int_0^t |F_0(\eta e^{-s})| \exp\Big(-\frac{1}{4}\sigma^2\eta^2(1-e^{-2s})\Big) ds$$

$$\leqslant \|\hat{\rho}_0^{(1)}(0,\cdot)\|_{L^\infty} e^{-\frac{1}{4}\sigma^2\eta^2(1-e^{-2})} + \|m_{0,1}^{(1)}\|_{L^\infty} \int_0^1 |F_0(\eta e^{-s})| ds$$

$$+ \|m_{0,1}^{(1)}\|_{L^\infty} |G'(\xi)| \|\hat{F}_\xi\|_{L^\infty} \int_{L^1}^t |\eta| e^{-s} \exp\Big(-\frac{1}{4}\sigma^2\eta^2(1-e^{-2})\Big) ds$$

$$\leqslant \|\hat{\rho}_0^{(1)}(0,\cdot)\|_{L^\infty} e^{-\frac{1}{4}\sigma^2\eta^2(1-e^{-2})} + \|m_{0,1}^{(1)}\|_{L^\infty} \int_0^1 |F_0(\eta e^{-s})| ds$$

$$+ \|m_{0,1}^{(1)}\|_{L^\infty} |G'(\xi)| \|\hat{F}_\xi\|_{L^\infty} |\eta| e^{-\frac{1}{4}\sigma^2\eta^2(1-e^{-2})}.$$

Therefore, there exists a constant C that depends only on G, ξ and σ such that

$$\|\hat{\rho}_0^{(1)}(t,\cdot)\|_{L^1} + \|\hat{\rho}_0^{(1)}(t,\cdot)\|_{L^\infty} \leqslant C.$$

We can proceed in the same way after differentiating with respect to η the expression (30) of $\hat{\rho}_0^{(1)}$ to show that $\|\partial_\eta\hat{\rho}_0^{(1)}(t,\cdot)\|_{L^\infty}$ is bounded uniformly in t. $\qquad\square$

When $u - G(u)$ derives from an even double-well potential, such as the examples (4) and (5), then, as soon as there exist nonzero solutions $\pm\xi_e$ to the compatibility equation $\xi = G(\xi)$, we have $G'(\pm\xi_e) < 1$ because $\partial_u(u-G(u))_{u=\pm\xi_e} > 0$. Therefore, from Proposition 1 and Proposition 2, the existence of the order states $\rho_{\pm\xi_e}$ in (7) is equivalent to the stability of the 0-th mode $\rho_0^{(1)}$ in (28). In addition, the condition that the equation $\xi = G(\xi)$ has the nonzero solutions $\pm\xi_e$ implies $G'(0) > 1$ because $\partial_u(u-G(u))_{u=0} < 0$, and therefore the disorder state ρ_0 is an unstable equilibrium to the nonlinear Fokker-Planck equation (7) when the order states $\rho_{\pm\xi_e}$ exist.

4.3. **Sufficient condition for the stability of the kth-order modes.** In this section we find a sufficient condition for the stability of the kth-order modes for $k \neq 0$. We show that all the nonzero modes are stable for sufficiently large σ. From (26) in Lemma 4.1, $w_k(t) = \partial_\eta\hat{\rho}_k^{(1)}(t,0)$ satisfies

$$w_k(t) = \psi_k(t) + \int_0^t R_k(t-s)w_k(s)ds, \tag{31}$$

where

$$\beta_k(t) = D_k(1 - e^{-t}), \tag{32}$$

$$\psi_k(t) = \Big(-\partial_\eta g_k(t,0)\hat{\rho}_k^{(1)}(0,-\beta_k) + \partial_\eta\hat{\rho}_k^{(1)}(0,-\beta_k)e^{-t}\Big) e^{-g_k(t,0)}, \tag{33}$$

$$R_k(t) = G'(\xi)\phi_k\Big[-\partial_\eta g_k(t,0)H_\xi(-\beta_k(t)) + H'_\xi(-\beta_k(t))e^{-t}\Big] e^{-g_k(t,0)}, \tag{34}$$

The strategy is to show that $\psi_k \in L^\infty(0,\infty)$ and that $\int_0^\infty |R_k(t)|dt < 1$, and thus $\|w_k\|_{L^\infty} \leqslant (1 - \|R_k\|_{L^1})^{-1}\|\psi_k\|_{L^\infty}$ by (31). Once it is known that w_k is uniformly bounded, it is not difficult to show that the kth mode is stable by the inspection of the formula (26).

Lemma 4.2. *1. ψ_k defined by (33) belongs to $L^\infty[0,\infty)$.*
2. Let R_k be defined by (34). If $|G'(\xi)\phi_k| < 1$ for all nonzero k, then $\int_0^\infty |R_k(t)|dt < 1$ for all nonzero k if σ is large enough.

Proof. See Appendix A.   □

Proposition 3. *If $|G'(\xi)\phi_k| < 1$ for all nonzero k and σ is large enough, then all the kth-order modes for $k \neq 0$ are stable.*

Proof. First we write $\hat{\rho}_k^{(1)}(t, \eta)$ as (26). By Lemma 4.2, $w_k \in L^\infty[0, \infty)$ and thus we can show that $\|\hat{\rho}_k^{(1)}(t, \cdot)\|_{L^1} + \|\hat{\rho}_k^{(1)}(t, \cdot)\|_{L^\infty} + \|\partial_\eta \hat{\rho}_k^{(1)}(t, \cdot)\|_{L^\infty}$ is uniformly bounded by the estimate similar to the proof of Proposition 2.   □

4.4. **Necessary and sufficient condition for the instability of the kth-order modes.** The linear stability analysis that we have presented reveals that the stability of the k-th mode ($k \neq 0$) is determined by the behavior of the function w_k solution of (31). The k-th mode is stable if and only if the function w_k is bounded. We have shown that the condition $\int_0^\infty |R_k(t)| dt < 1$ is a sufficient condition for the stability of the k-th mode, which is not, however, necessary. It is possible to give a necessary and sufficient condition for the instability of the k-th mode using the Fourier-Laplace transform of (31). This condition is that

$$C_k = \left\{ \gamma \in \mathbb{C} \text{ s.t. } \mathrm{Re}(\gamma) > 0 \text{ and } \int_0^\infty R_k(t) e^{-\gamma t} dt = 1 \right\} \tag{35}$$

is non-empty. If this happens, then the Fourier-Laplace transform of w_k:

$$\mathcal{W}_k(\gamma) = \int_0^\infty w_k(t) e^{-\gamma t} dt$$

blows up when the complex frequency γ goes to an element in C_k, so we can conclude that $|w_k(t)|$ grows exponentially with the growth rate $\gamma_r(k) = \sup\{\mathrm{Re}(\gamma), \gamma \in C_k\}$. If there exists a unique $\gamma(k) \in C_k$ with $\mathrm{Re}(\gamma(k)) = \gamma_r(k)$, then we can predict that the k-th mode grows like $\exp(\gamma(k)t)$. The real part $\gamma_r(k)$ is the growth rate of the k-th mode, the imaginary part $\gamma_i(k)$ describes the dynamical behavior of the unstable mode as follows.

The mode with the largest growth rate is the one that drives the instability. If we denote $k_{\max} = \mathrm{argmax}_k(\gamma_r(k))$, then the instability has the form (up to a multiplicative constant):

$$\exp[-i(2\pi k_{\max} x / L - \gamma_i(k_{\max})t)] \exp[\gamma_r(k_{\max})t]. \tag{36}$$

For a fixed time, its spatial profile has the form of a periodic modulation with the spatial period $L/k_{\max}$. This modulation moves in time at the velocity $\gamma_i(k_{\max})L/(2\pi k_{\max})$ and grows exponentially in time with the rate $\gamma_r(k_{\max})$. This shows that the initial uniform distribution of the positions is not stable and that spatial clustering appears.

In Figure 1 we plot the growth rate of the most unstable mode in the situation addressed in the numerical section 7. The double-well potential G defined by (47) is parameterized by h which determines the depths of the wells and one can see that the linear stability is optimal when h is nor too small, neither too large. This will be confirmed by the numerical simulations presented below.

4.5. **Conclusion.** If $\xi = G(\xi)$ and $G'(\xi) < 1$, then the 0th-order mode of the stationary state ρ_ξ is stable. If, additionally, $|G'(\xi)\phi_k| < 1$ for all $k \neq 0$, then all the kth-order modes for $k \neq 0$ are stable for a sufficiently large σ. Under these conditions the stationary state ρ_ξ is stable. When G is such that $u - G(u)$ derives
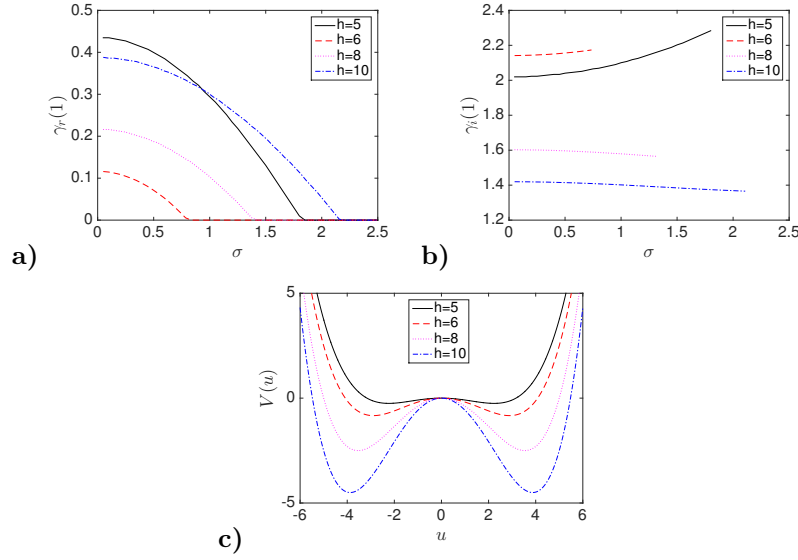
FIGURE 1. The real (a) and imaginary (b) parts of the complex growth rate of the first mode (the most unstable mode) as a function of σ for $h = 5$ (solid black), $h = 6$ (dashed red), $h = 8$ (magenta dotted), and $h = 10$ (blue dot-dashed). Here ϕ is given by (46) and $L = 10$. G is given by (47) and it derives from the double-well potential plotted in picture c. We address the linear stability of the order state ρ_{ξ_e}. One can see that the threshold value for the noise level σ to ensure linear stability is 1.8 for $h = 5$, 0.85 for $h = 6$, 1.4 for $h = 8$, and 2.2 for $h = 10$. The most stable situation is the one corresponding to $h = 6$.

from an even double-well potential, such as the two examples (4) and (5), this means that the order states

$$\rho_{\pm\xi_e}(x, u) = \frac{1}{L} \frac{1}{\sqrt{\pi\sigma^2}} \exp\left(- \frac{(u \mp \xi_e)^2}{\sigma^2} \right), \quad \xi_e = G(\xi_e) > 0,$$

are stable equilibria to the nonlinear Fokker-Planck equation (7), while the disorder state

$$\rho_0(x, u) = \frac{1}{L} \frac{1}{\sqrt{\pi\sigma^2}} \exp\left(- \frac{u^2}{\sigma^2} \right) \tag{37}$$

is an unstable equilibrium. This shows that the noise strength σ improves the stability of the order states $\rho_{\pm\xi_e}(x, u)$. Such a phenomenon that the randomness of the movements establishes the alignment of collective motion is also found in [41] both numerically and experimentally.

Note, however, an interesting phenomenon: if $G'(\xi) < 1$, then the 0th-order mode is stable, but for some $k \neq 0$ the kth-order mode may be unstable. This instability is very different from the instability that results from violating the condition $G'(\xi) < 1$, because it manifests itself in an instability in the distribution of the positions of the particles which cannot stay uniform. This situation happens when $u - G(u)$ derives from a double-well potential and the wells are very deep, as we will see later in the numerical section. The prediction that the order states become unstable when

the depths of the wells are too deep seems counterintuitive and indeed this never happens when the interaction is uniform $\phi \equiv 1$, but it happens when the interaction is local.

5. **Fluctuations analysis.** In this section, we consider the fluctuation analysis for the interacting system of agents around the stationary state. Recall that μ_N is the empirical probability measure of the locations and velocities (6). Here we consider a solution ξ to the compatibility equation $\xi = G(\xi)$. We assume that, at time 0, the initial positions and velocities of the agents are independent and identically distributed with the distribution with density $\rho_\xi(x, u)$. Therefore $\mu_N(0, dx, du)$ converges to the stationary state $\rho_\xi(x, u)dxdu$ in (9) as $N \to \infty$. Moreover, $\mu_N(t, dx, du)$ converges to $\rho_\xi(x, u)dxdu$ for all t as $N \to \infty$ by (7).

We define, for any test function $f(x, u)$ and any measure $X(dx, du)$ (in the space $\mathcal{S}'$ of tempered Schwartz distributions on $[0, L] \times \mathbb{R}$):

$$\langle X, f \rangle = \int_0^L \int_{-\infty}^{\infty} f(x, u) X(dx, du).$$

We have

$$\langle \mu_N(t), f \rangle = \frac{1}{N} \sum_{i=1}^{N} f(x_i(t), u_i(t)).$$

We define the fluctuation process around $\rho_\xi(x, u)dxdu$:

$$X_N(t, dx, du) = \sqrt{N}[\mu_N(t, dx, du) - \rho_\xi(x, u)dxdu], \tag{38}$$

and study its dynamics as $N \to \infty$.

First, by central limit theorem, at time 0,

$$\sqrt{N}\Big(\langle \mu_N(0), f \rangle - \langle \rho_\xi dxdu, f \rangle\Big) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \big(f(x_i(0), u_i(0)) - \langle \rho_\xi dxdu, f \rangle\big)$$

converges to a Gaussian random variable with mean zero and variance $\langle \rho_\xi dxdu, f^2 \rangle - \langle \rho_\xi dxdu, f \rangle^2$. This shows that $X_N(t = 0, dx, du)$ converges weakly as $N \to \infty$ to a Gaussian measure $X(t = 0, dx, du)$ with mean zero and covariance

$$\mathbb{E}\big[\langle X(t = 0), f_1 \rangle \langle X(t = 0), f_2 \rangle\big] = \int_0^L \int_{-\infty}^{\infty} f_1(x, u) f_2(x, u) \rho_\xi(x, u)dxdu$$

$$- \int_0^L \int_{-\infty}^{\infty} f_1(x, u) \rho_\xi(x, u)dxdu \int_0^L \int_{-\infty}^{\infty} f_2(x, u) \rho_\xi(x, u)dxdu, \tag{39}$$

for any test functions f_1 and f_2.

Second, by Itô's formula,

$$d\langle \mu_N, f \rangle = \frac{1}{N} \sum_{i=1}^{N} df(x_i(t), u_i(t))$$

$$= \langle \mu_N, u\frac{\partial f}{\partial x} \rangle dt + \langle \mu_N, \big[G\big(\langle \mu_N(t, dx', du'), u'\phi(\|x' - x\|)\rangle\big) - u\big]\frac{\partial f}{\partial u} \rangle dt$$

$$+ \langle \mu_N, \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial u^2} \rangle dt + \frac{1}{N} \sum_{i=1}^{N} \langle \delta_{(x_i, u_i)}(dx, du), \sigma\frac{\partial f}{\partial u} \rangle dW_i(t).$$

By integration by parts, we can write

$$d\mu_N = -u\frac{\partial \mu_N}{\partial x}dt - \frac{\partial}{\partial u}\left\{[G(\langle\mu_N(t,dx',du'),u'\phi(\|x'-x\|)\rangle) - u]\mu_N\right\}dt$$
$$+ \frac{1}{2}\sigma^2\frac{\partial^2\mu_N}{\partial u^2}dt - \sigma\frac{\partial}{\partial u}\Big[\frac{1}{N}\sum_{i=1}^N \delta_{(x_i,u_i)}(dx,du)dW_i(t)\Big].$$

Therefore X_N satisfies

$$dX_N = \sqrt{N}d[\mu_N(t,dx,du) - \rho_\xi(x,u)dxdu]$$
$$= -u\frac{\partial X_N}{\partial x}dt - \frac{\partial}{\partial u}\left\{[G(\langle\mu_N(t,dx',du'),u'\phi(\|x'-x\|)\rangle) - u]\sqrt{N}\mu_N\right\}dt$$
$$+ \frac{\partial}{\partial u}[(G(\xi) - u)\sqrt{N}\rho_\xi]dt + \frac{1}{2}\sigma^2\frac{\partial^2 X_N}{\partial u^2}dt$$
$$- \frac{\sigma}{\sqrt{N}}\sum_{i=1}^N \frac{\partial}{\partial u}\delta_{(x_i,u_i)}(dx,du)dW_i(t).$$

Note that, because

$$\xi = G(\xi) = G\big(\langle\rho_\xi(x',u')dx'du', u'\phi(\|x'-x\|)\rangle\big),$$

we have

$$G(\langle\mu_N(t,dx',du'),u'\phi(\|x'-x\|)\rangle)$$
$$= G\big(\langle\rho_\xi(x',u')dx'du' + N^{-1/2}X_N(t,dx',du'), u'\phi(\|x'-x\|)\rangle\big)$$
$$= G\big(\xi + N^{-1/2}\langle X_N(t,dx',du'), u'\phi(\|x'-x\|)\rangle\big),$$

so that we can also write that X_N satisfies

$$dX_N = -u\frac{\partial X_N}{\partial x}dt$$
$$- \frac{\partial}{\partial u}\left\{[G\big(\xi + N^{-1/2}\langle X_N(t,dx',du'),u'\phi(\|x'-x\|)\rangle\big) - u]X_N\right\}dt$$
$$+ \frac{\partial}{\partial u}\left\{\sqrt{N}[G(\xi) - G\big(\xi + N^{-1/2}\langle X_N(t,dx',du'),u'\phi(\|x'-x\|)\rangle\big)]\rho_\xi\right\}dt$$
$$+ \frac{1}{2}\sigma^2\frac{\partial^2 X_N}{\partial u^2}dt + \sigma dW^N(t), \tag{40}$$

where $W^N(t)$ is the measure-valued stochastic process

$$W^N(t) = -\int_0^t \frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{\partial}{\partial u}\delta_{(x_i(s),u_i(s))}(dx,du)dW_i(s).$$

The process $W^N(t)$ in $\mathcal{C}([0,\infty),\mathcal{S}')$ is such that, for any test function $f(x,u)$,

$$\langle W^N(t),f\rangle = -\int_0^t\int_0^L\int_{-\infty}^\infty f(x,u)\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{\partial}{\partial u}\delta_{(x_i(s),u_i(s))}(dx,du)dW_i(s)$$
$$= \int_0^t \frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{\partial f}{\partial u}(x_i(s),u_i(s))dW_i(s).$$

It is a zero-mean martingale whose quadratic variation is, for any test functions f_1 and f_2,

$$\left[\langle W^N, f_1 \rangle, \langle W^N, f_2 \rangle\right]_t = \int_0^t \frac{1}{N} \sum_{i=1}^N \frac{\partial f_1}{\partial u}(x_i(s), u_i(s)) \frac{\partial f_2}{\partial u}(x_i(s), u_i(s)) ds$$

$$= \int_0^t \langle \mu_N(s, dx, du), \frac{\partial f_1}{\partial u} \frac{\partial f_2}{\partial u} \rangle ds,$$

which converges to the deterministic process

$$\left[\langle W^N, f_1 \rangle, \langle W^N, f_2 \rangle\right]_t \overset{N \to \infty}{\longrightarrow} t \langle \rho_\xi(x, u) dx du, \frac{\partial f_1}{\partial u} \frac{\partial f_2}{\partial u} \rangle.$$

In other words, as $N \to \infty$, W^N converges to a Brownian field, and this implies that X_N converges to a generalized Ornstein-Uhlenbeck process, as explained in the following proposition.

Proposition 4. *Let $\mu_N(t, dx, du) = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i(t), u_i(t))}(dx, du)$ and ξ be a solution to $\xi = G(\xi)$. We assume that the initial positions and velocities of the agents are independent and identically distributed with the distribution with density $\rho_\xi(x, u)$.*
1) As $N \to \infty$, $\mu_N(0, dx, du)$ converges to the stationary state $\rho_\xi(x, u) dx du$ and $X_N(0, dx, du)$ converges to the Gaussian measure-valued process $X(0, dx, du)$ with mean zero and covariance (39).
2) As $N \to \infty$, $X_N(t, dx, du)$ converges to the Gaussian measure-valued process $X(t, dx, du)$ satisfying

$$dX = -u \frac{\partial X}{\partial x} dt - \frac{\partial}{\partial u}[(\xi - u)X] dt \tag{41}$$

$$- G'(\xi) \frac{\partial \rho_\xi}{\partial u} \left[\int_0^L \int_{-\infty}^\infty u' \phi(\|x' - x\|) X(t, dx', du') \right] dt + \frac{1}{2} \sigma^2 \frac{\partial^2 X}{\partial u^2} dt + \sigma dW_\xi,$$

where $W_\xi(t, x, u)$ is a Gaussian process independent of $X(0, dx, du)$ with mean zero and covariance

$$\mathbf{Cov}\left(\int_0^L \int_{-\infty}^\infty W_\xi(t, x, u) f_1(x, u) dx du, \int_0^L \int_{-\infty}^\infty W_\xi(t', x, u) f_2(x, u) dx du \right)$$

$$= \min(t, t') \int_0^L \int_{-\infty}^\infty \frac{\partial f_1}{\partial u}(x, u) \frac{\partial f_2}{\partial u}(x, u) \rho_\xi(x, u) dx du, \tag{42}$$

for any test functions f_1 and f_2.

The convergences hold in the sense of weak convergence on $\mathcal{C}([0, \infty), \mathcal{S}')$ where $\mathcal{S}'$ denotes the space of tempered Schwartz distributions on $[0, L] \times \mathbb{R}$.

Proof. We only provide the main steps of the derivation of Proposition 4, which follows the one of [12, Theorem 4.1.1]. We first note that the solution X_N of (40) is the solution of a martingale problem. To prove the existence of a weak limit, we need to prove that the set $\{X_N\}_{N=1}^\infty$ solution of (40) is weakly compact and thus the sequence $\{X_N, N = 1, 2, 3, \ldots\}$ has limits. To prove the uniqueness of the weak limit, we need to prove that any weak limit X is solution of a well-posed martingale problem. Note that the second and third terms in the right-hand side of (40) satisfy

$$\left[G\left(\xi + N^{-1/2} \langle X_N(t, dx', du'), u' \phi(\|x' - x\|) \rangle\right) - u \right] X_N \overset{N \to \infty}{\longrightarrow} (\xi - u) X$$

and

$$\sqrt{N}[G(\xi) - G(\xi + N^{-1/2}\langle X_N(t, dx', du'), u'\phi(\|x' - x\|)\rangle)]$$
$$\xrightarrow{N\to\infty} -G'(\xi)\langle X(t, dx'du'), u'\phi(\|x' - x\|)\rangle,$$

when X_N weakly converges to X. This shows that X is solution of the martingale problem associated to the Markov diffusion process (41). Then the uniqueness of the limit is ensured by the uniqueness of the solution of the limit martingale problem. $\square$

The equation (41) for the fluctuation process X involves the same linearized operator as in (20). So the linear stability analysis carried out in the previous section can be invoked to claim that the system (1) is linearly stable when the initial empirical distribution is initially close to ρ_ξ, where ξ is such that $G(\xi) = \xi$, $G'(\xi) < 1$, $|\phi_k G'(\xi)| < 1$ for all $k \neq 0$, and σ is large enough. When G is such that $u - G(u)$ derives from an even double-well potential, such as the two examples (4) and (5), this means that the order states $\rho_{\pm\xi_e}$ are stable equilibria, in the sense that if the initial empirical distribution is close to one of them, then the empirical distribution remains close to it, with small normal fluctuations described by (41) in the limit $N \to \infty$, while the disorder state ρ_0 is an unstable equilibrium, in the sense that if the initial empirical distribution is close to it, then the empirical distribution quickly moves away from it.

6. **Large deviations analysis.** Here we assume that two order states $\rho_{\pm\xi_e}$ exist and are stable and that the initial empirical measure $\mu_N(t = 0, dx, du)$ converges to ρ_{ξ_e}. By the mean-field theory, the empirical measure (6) converges to $\rho_{\xi_e}(x, u)dxdu$ as $N \to \infty$ and the stability analysis ensures that $\mu_N(t, dx, du)$ is approximately ρ_{ξ_e} over the time interval $[0, T]$. However, when N is large but finite, the empirical measure μ_N is still stochastic and there is a very small probability that, eventhough $\mu_N \approx \rho_{\xi_e}$ at time $t = 0$, $\mu_N \approx \rho_{-\xi_e}$ at time $t = T$. Here we use the large deviation principle (LDP) to write down the asymptotic probability of a transition from one stable order state to the other one. The empirical measure μ_N satisfies the large deviation principle in the space of continuous probability-measure-valued processes as the following: for a set A of paths of probability measures $\mu = (\mu(t, dx, du))_{t\in[0,T]}$, we have

$$-\inf_{\mu\in\mathring{A}} I(\mu) \leqslant \liminf_{N\to\infty} \frac{1}{N} \log \mathbf{P}(\mu_N \in A)$$

$$\leqslant \limsup_{N\to\infty} \frac{1}{N} \log \mathbf{P}(\mu_N \in A) \leqslant -\inf_{\mu\in\bar{A}} I(\mu),$$

where $\mathring{A}$ and $\bar{A}$ are the interior and closure of A, respectively, with an appropriate topology [13]. The function I is called the rate function:

$$I(\mu) = \frac{1}{2\sigma^2} \int_0^T \sup_{f(x,u):\langle\mu(t,\cdot,\cdot),(\frac{\partial f}{\partial u})^2\rangle\neq 0} \frac{\langle\frac{\partial\mu}{\partial t}(t,\cdot,\cdot) - \mathcal{L}^*_{\mu(t,\cdot,\cdot)}\mu(t,\cdot,\cdot), f\rangle^2}{\langle\mu(t,\cdot,\cdot),(\frac{\partial f}{\partial u})^2\rangle} dt, \quad (43)$$

where $\mathcal{L}^*_\nu$ is the differential operator associated to the Fokker-Planck equation:

$$\mathcal{L}^*_\nu\mu = -u\frac{\partial\mu}{\partial x} - \frac{\partial}{\partial u}\left[\left(G\left(\iint u'\phi(\|x'\|)\nu(x - x', u')du'dx'\right) - u\right)\mu\right] + \frac{1}{2}\sigma^2\frac{\partial^2\mu}{\partial u^2}.$$

We are interested in the rare event A which corresponds to the transition from one order state $\rho_{\xi_e}(x, u)dxdu$ to the other one $\rho_{-\xi_e}(x, u)dxdu$. We therefore assume

that the initial conditions correspond to independent agents with the distribution $\rho_{\xi_e}(x, u)dxdu$ and we consider the event

$$A = \left\{ (\mu(t, dx, du))_{t\in[0,T]} \ : \ \||\mu(T, dx, du) - \rho_{-\xi_e}(x, u)dxdu\|| \leqslant \delta \right\}, \qquad (44)$$

for some small δ, where $\||\cdot\||$ stands for the metric of the space of probability measures.

Roughly speaking, for large but finite N, the probability of transitions from one stable state to the other is an exponential function of N whose exponential decay rate is governed by the rate function I:

$$\mathbf{P}(\mu_N \in A) \overset{N\gg 1}{\approx} \exp\left(-N \inf_{\mu\in A} I(\mu)\right). \qquad (45)$$

The transition probability (45) of the switching of the alignments of locusts is experimentally and numerically observed in [7].

Remark 1. In this section we have only provided a formal large deviation description, because the classical large deviation result [13] requires that the noise must be non-degenerate. Therefore, the result of [13] cannot be directly applied to our model and we are not able to well define the space for our case until the rigorous large deviation is proven. However, a rigorous large deviation principle could be possibly constructed by the technique in [6].

7. **Numerical simulations.** We use the numerical simulations to illustrate and verify our theoretical analysis. The spatial domain is the torus $[0, L] = [0, 10]$, the influence function is

$$\phi(\|x\|) = 5 \times \mathbf{1}_{[0,1]}(\|x\|), \qquad (46)$$

which is such that $\int_0^L \phi(\|x\|)dx/L = 1$ and $\phi_k = \mathrm{sinc}(\pi k/5)$, and we let

$$G(u) = \frac{h+1}{5}u - \frac{h}{125}u^3, \qquad (47)$$

which is such that $u - G(u)$ derives from the potential $V(u) = \frac{4-h}{10}u^2 + \frac{h}{500}u^4$, that is a double-well potential as soon as $h > h_c := 4$. The parameter h allows us to quantify the depths of the two wells of the double-well potential when $h > h_c$ (see Figure 1c).

We discretize the time domain: $t_n = n\Delta t$, $n = 0, 1, 2, 3, \ldots$ and use the Euler method to simulate (1):

$$x_i^{n+1} - x_i^n = u_i^n \Delta t,$$
$$u_i^{n+1} - u_i^n = \left[G(\langle u\rangle_i^n) - u_i^n\right]\Delta t + \sigma\Delta W_i^n,$$

where $\{\Delta W_i^n\}_{i,n}$ are independent Gaussian random variables with mean 0 and variance Δt, and $\langle u\rangle_i^n$ is a weighted average of the velocities with respect to u_i^n:

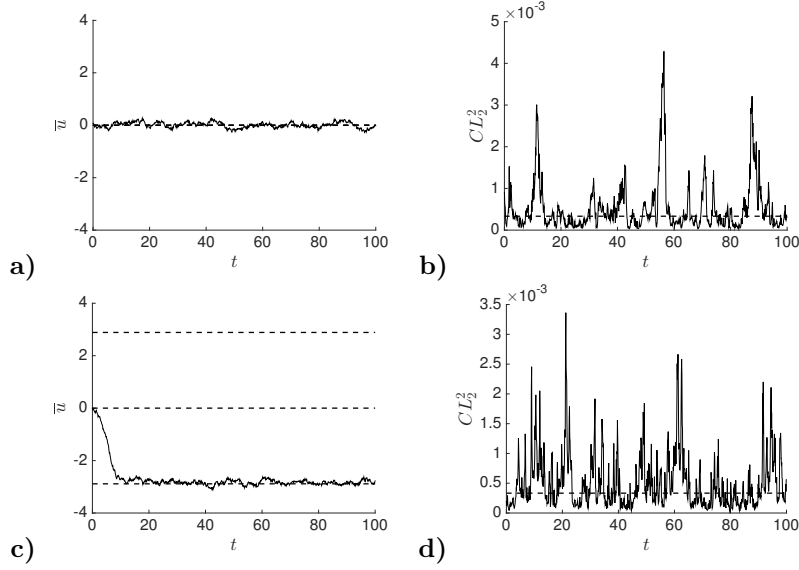$$\langle u\rangle_i^n = \frac{1}{N}\sum_{j=1}^N u_j^n \phi(\|x_j^n - x_i^n\|). \qquad (48)$$

FIGURE 2. The empirical average velocity $\bar{u}^n$ and the square centered L^2-discrepancy $CL_2^2(n)$ at each time step t_n for $h = 2$ (a-b) and $h = 6$ (c-d). The dashed lines in the velocity plots are the solutions of $G(\xi) = \xi$. The dashed lines in the discrepancy plots stand for the value (53) corresponding to a uniform sampling. Here $\Delta t = 0.1$, $N = 500$, and $\sigma = 2$. One can see that the spatial distribution is uniform and the velocity average is 0 when $h = 2$ and $-\xi_e$ when $h = 6$.

7.1. **Existence of the order states.** From Proposition 1, the number of stationary states (7) is equal to the number of solutions ξ for the compatibility condition (10). When $h \leqslant h_c = 4$, the function $u - G(u)$ is increasing and 0 is the unique solution of the compatibility equation (10). When $h > h_c$, the function $u - G(u)$ derives from a double well-potential and there are three solutions $0, \pm\xi_e$ to the compatibility equation (10), with

$$\xi_e = 5\sqrt{\frac{h-4}{h}}. \tag{49}$$

In Figure 2 we test the cases of $h = 2$ and $h = 6$ and plot the empirical average velocity

$$\bar{u}^n = \frac{1}{N}\sum_{i=1}^{N} u_i^n \tag{50}$$

and the square centered L^2-discrepancy:

$$CL_2^2(n) = \int_0^1 |F_n(x) - x|^2 dx, \qquad F_n(x) = \frac{\text{Card}(i = 1, \ldots, N\,,\, x_i^n/L \leqslant x)}{N}. \tag{51}$$

It measures the discrepancy between the empirical distribution of the positions and the uniform distribution over $[0, L]$. By [24] it is given by

$$CL_2^2(n) \;=\; \frac{13}{12} - \frac{1}{N}\sum_{i=1}^{N}\left(2 + \left|\frac{x_i^n}{L} - \frac{1}{2}\right| - \left|\frac{x_i^n}{L} - \frac{1}{2}\right|^2\right)$$

$$+ \frac{1}{2N^2}\sum_{i,j=1}^{N}\left(2 + \left|\frac{x_i^n}{L} - \frac{1}{2}\right| + \left|\frac{x_j^n}{L} - \frac{1}{2}\right| - \left|\frac{x_i^n}{L} - \frac{x_j^n}{L}\right|\right). \qquad (52)$$

It is known that, if the positions are independent and identically distributed according to the uniform distribution over $[0, L]$, then the square centered L^2-discrepancy has mean [20]:

$$CL_{2,u}^2 = \frac{1}{N}\left(\frac{5}{4} - \frac{13}{12}\right), \qquad (53)$$

while its variance is of order N^{-2}. Therefore, as long as the empirical distribution in positions stays uniform, the square centered L^2-discrepancy stays close to the value (53).

In Figure 2 the initial locations and velocities $\{(x_i^0, u_i^0)\}_{i=1}^N$ are sampled from the distribution $\rho_0(x, u)dxdu$, the stationary state in (9) with $\xi = 0$. That is, $\{x_i^0\}_{i=1}^N$ are uniformly sampled over $[0, L]$ and $\{u_i^0\}_{i=1}^N$ are sampled from the Gaussian distribution with mean 0 and variance $\sigma^2/2$.

- When $h = 2 < h_c = 4$, the only solution of $\xi = G(\xi)$ is $\xi = 0$, which means that ρ_0 is the unique stationary state. Moreover, $G'(0) = (h+1)/5 < 1$ and the noise level is strong enough to make it linearly stable. Indeed, we can see in Figure 2a that the empirical average velocity $\bar{u}^n$ oscillates around zero and the square centered L^2-discrepancy oscillates around the value $CL_{2,u}^2$, which indicates that the empirical distribution of locations and velocities stays close to ρ_0.

- When $h = 6 > h_c = 4$, the order states $\rho_{\pm\xi_e}$ exist. Moreover, $G'(\xi_e) = (13 - 2h)/5 < (13 - 2h_c) = 1$ and the noise level is strong enough to make the order states $\rho_{\pm\xi_e}$ linearly stable. In addition, $G'(0) = (h+1)/5 > (h_c+1)/5 = 1$, therefore, the disorder state ρ_0 is an unstable state. Under these circumstances, the initial empirical distribution that is close to ρ_0 quickly changes as can be seen in Figure 2c-d: $\bar{u}^n$ oscillates around $-\xi_e$ (it could have been ξ_e), while the square centered L^2-discrepancy oscillates around the value $CL_{2,u}^2$. This indicates that the empirical distribution of locations and velocities becomes close to $\rho_{-\xi_e}$.

7.2. **Linear stability.** Here we assume that $h > h_c = 4$ so that there are three stationary states ρ_0, $\rho_{\pm\xi_e}$. We have $G'(0) = (h+1)/5 > 1$, $G'(\pm\xi_e) = (13-2h)/5 < 1$, and thus the 0th order mode of ρ_0 is unstable while the 0th-order modes of $\rho_{\pm\xi_e}$ are stable. An immediate conclusion is that ρ_0 is an unstable state for the nonlinear Fokker-Planck equation (7). The zeroth-order mode of $\rho_{\pm\xi_e}$ is linearly stable. However, the stability of $\rho_{\pm\xi_e}$ requires that all the nonzeroth-order modes of $\rho_{\pm\xi_e}$ are linearly stable. From Proposition 3, we know that for a sufficiently large σ, all the nonzeroth-order modes are stable. Moreover the critical value of σ ensuring the stability can be determined as explained in subsection 4.4.

If $h = 6$, then the theory in subsection 4.4 predicts that the condition for the stability of the order states $\rho_{\pm\xi_e}$ is that σ should be larger than 0.85. The simulations shown in Figure 3 confirm these predictions: $\rho_{\pm\xi_e}$ is stable when σ exceeds the threshold value 0.85. They also allow us to illustrate the instability mechanism exhibited above: when σ is below the threshold value, the average velocity of the particles is linearly stable, but the uniform distribution of the positions of the particles is not stable and a spatial modulation grows that gives rise to one moving
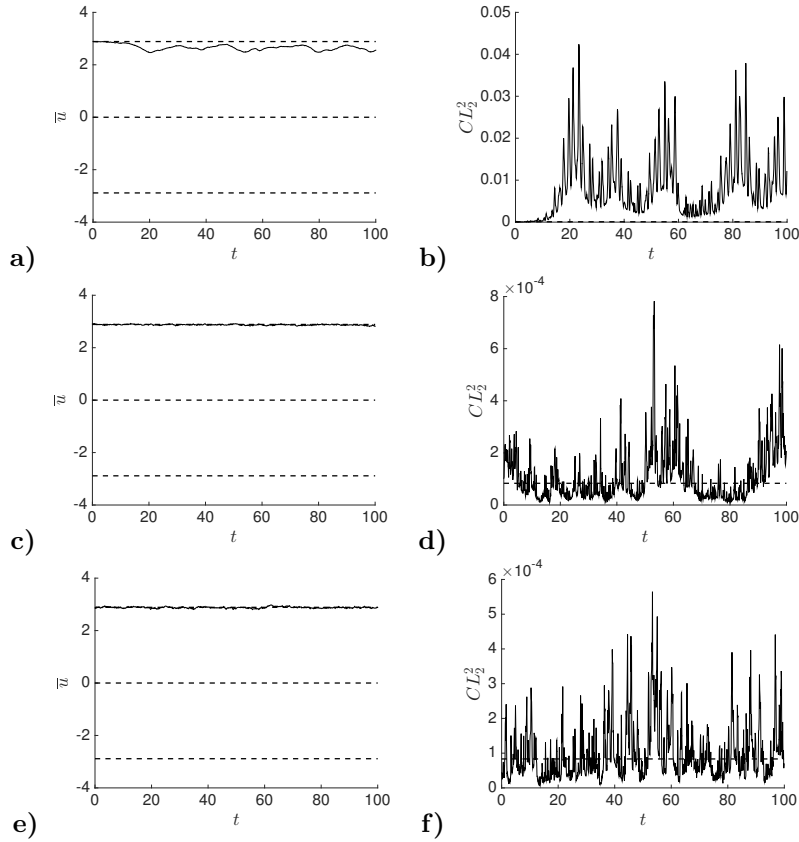
FIGURE 3. The empirical average velocity $\bar{u}^n$ and the square centered L^2-discrepancy $CL_2^2(n)$ for $\sigma = 0.5$ (a-b), $\sigma = 1$ (c-d), $\sigma = 1.5$ (e-f). The initial positions $\{x_i^0\}_{i=1}^N$ are uniformly sampled over $[0, L]$ and the initial velocities $\{u_i^0\}_{i=1}^N$ are sampled from the Gaussian distribution with mean ξ_e and variance $\sigma^2/2$. Here $\Delta t = 0.1$, $N = 2000$, and $h = 6$. One can see that the average velocity is not ξ_e and the spatial distribution is not uniform when $\sigma = 0.5$ while the average velocity is ξ_e and the spatial distribution is uniform when $\sigma = 1$ or 1.5.

cluster. Note that Figure 3a indeed shows that the uniform distribution of the positions is not stable when $\sigma = 0.5$ since the square L^2-discrepancy takes values much larger than (53). The theory in subsection 4.4 predicts that the most unstable mode is the first one $k_{\max} = 1$, which means that there should be one moving cluster. Figure 1 plots the predicted growth rate $\gamma_r(1)$ of the first mode and the coefficient $\gamma_i(1)$. For $\sigma = 0.5$, we have $\gamma_i(1) \simeq 2.15$ and the apparent velocity for the cluster is predicted to be approximately equal to $\gamma_i(1) \times L/(2\pi) \simeq 3.4$ by (36), which is close to, although slightly larger than, the average velocity of the particles in the stationary distribution $\xi_e = 2.9$. There is no contradiction as the apparent velocity of the cluster can be interpreted as a "phase velocity" while the average velocity of the particles can be interpreted as a "group velocity". The growth of the first mode giving rise to the moving cluster can be observed in the simulations (see Figure 4).
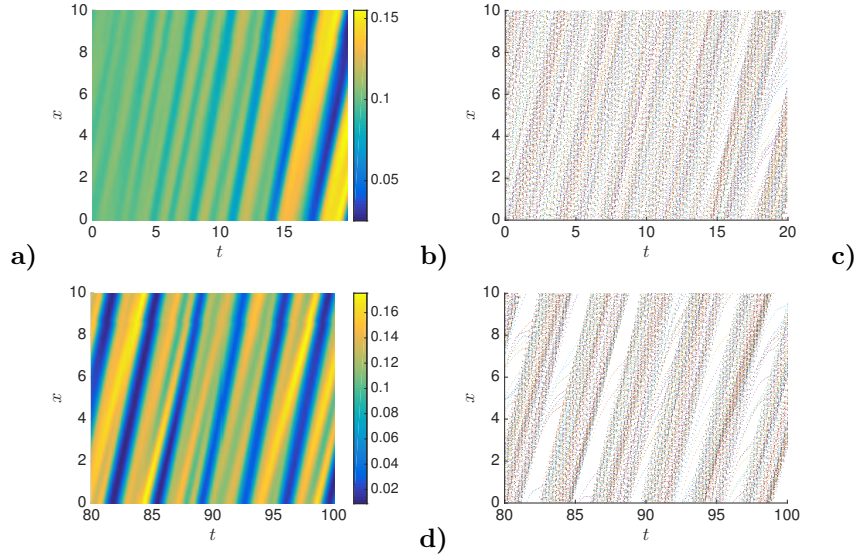
FIGURE 4. The empirical position distribution smoothed by kernel density estimation (a and c) and the first 200 trajectories (b and d). The initial positions $\{x_i^0\}_{i=1}^N$ are uniformly sampled over $[0, L]$ and the initial velocities $\{u_i^0\}_{i=1}^N$ are sampled from the Gaussian distribution with mean ξ_e and variance $\sigma^2/2$. Here $\Delta t = 0.1$, $N = 2000$, $\sigma = 0.5$, and $h = 6$. One can see that the spatial distribution is not uniform and a cluster is moving with an apparent velocity of 3.6.

Moreover, we can see in the right pictures of Figure 4 that the particles in the tail of the moving cluster are transfered to the front of the "next" cluster (which is the same one, by periodicity), which indeed allows the cluster to apparently move faster than the average velocity of the particles: in the simulations, the average velocity of the particles is about 2.8, while the velocity of the cluster is about 3.6, very close to the predicted value 3.4.

The previous results illustrate the fact that the order states $\rho_{\pm\xi_e}$ are linearly stable when σ exceeds the threshold value. This means that, if the initial conditions are close to $\rho_{\pm\xi_e}$, then the empirical distribution remains close to it. This is what we see in Figure 3c-f. When the initial conditions are far from $\rho_{\pm\xi_e}$ and when the noise level is significantly larger than the threshold value, then the empirical distribution quickly converges to one of the two stationary distributions $\rho_{\pm\xi_e}$, as we can see in Figure 5e-f: after a transient period, the empirical distribution in positions becomes uniform and the mean velocity becomes equal to $\pm\xi_e$. When the initial conditions are far from the two stable stationary states and when the noise level is larger than but close to the threshold value, then the empirical distribution reaches a state that is not stationary but looks like a moving cluster, as shown in Figure 5a-d.

As revealed by the linear stability analysis carried out in subsection 4.4, the results are not monotoneous as a function of h. Figure 6 and 7 show that the stationary states $\rho_{\pm\xi_e}$ are unstable when $\sigma = 1.0$ for $h = 5$ and for $h = 10$ while they are stable for $h = 6$ (see Figure 3c-d), as predicted by the theory in subsection 4.4.
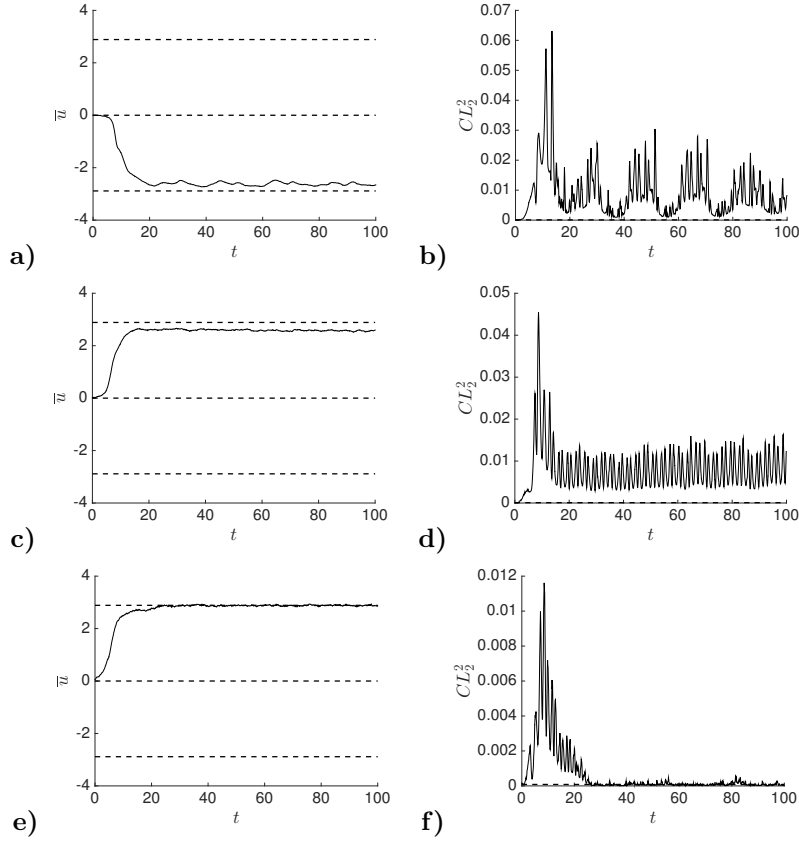
FIGURE 5. The empirical average velocity $\bar{u}^n$ and the square centered L^2-discrepancy $CL_2^2(n)$ for $\sigma = 0.5$ (a-b), $\sigma = 1$ (c-d), $\sigma = 1.5$ (e-f). The initial positions $\{x_i^0\}_{i=1}^N$ are uniformly sampled over $[0, L]$ and the initial velocities $\{u_i^0\}_{i=1}^N$ are sampled from the Gaussian distribution with mean 0 and variance $\sigma^2/2$. Here $\Delta t = 0.1$, $N = 2000$, and $h = 6$. One can see that the average velocity is not $\pm\xi_e$ and the spatial distribution is not uniform when $\sigma = 0.5$ or 1 while the average velocity is ξ_e and the spatial distribution is uniform when $\sigma = 1.5$.

This means that the two order states are stable when the depths of the wells of the double-well potential are neither too large nor too small.

7.3. **Large deviations.** In this subsection we assume that the conditions are fulfilled so that ρ_{ξ_e} and $\rho_{-\xi_e}$ exist and are stable. We consider the transition probability from one stable state to the other:

$$\mathbf{P}(\mu_N \in A) \overset{N \gg 1}{\approx} \exp\left(-N \inf_{\mu \in A} I(\mu)\right),$$

where the rare event A is the set (44) of all possible transitions from ρ_{ξ_e} at $t = 0$ to $\rho_{-\xi_e}$ at $t = T$, and the rate function I is defined in (43).
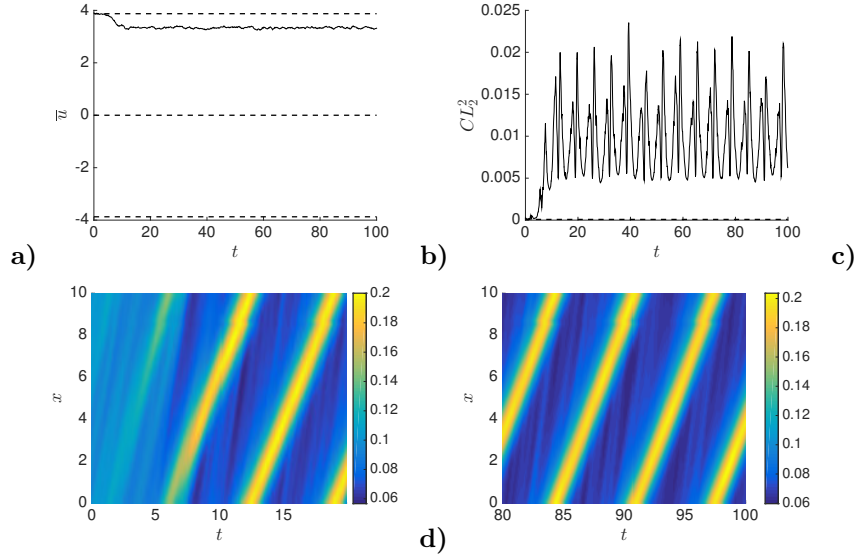
FIGURE 6. The empirical average velocity $\bar{u}^n$ (a), the square centered L^2-discrepancy $CL_2^2(n)$ (b), and the empirical position distribution smoothed by kernel density estimation (c and d). The initial positions $\{x_i^0\}_{i=1}^N$ are uniformly sampled over $[0, L]$ and the initial velocities $\{u_i^0\}_{i=1}^N$ are sampled from the Gaussian distribution with mean ξ_e and variance $\sigma^2/2$. Here $\Delta t = 0.1$, $N = 2000$, $\sigma = 1$, and $h = 10$. One can see that the average velocity is not ξ_e and the spatial distribution is not uniform.

In Figures 8, 9, and 10, we qualitatively examine the effects of N, σ, and h to the transition probability $\mathbf{P}(\mu_N \in A)$ by checking the frequencies of transitions of the empirical average velocity $\bar{u}^n$ between the two stable order states $\rho_{\pm\xi_e}$. In Figure 8, we can observe that the system has less transitions as N becomes larger. In other words, the probability of transition $\mathbf{P}(\mu_N \in A)$ becomes smaller as N becomes larger; this is consistent with the fact that the large deviation principle predicts that $\mathbf{P}(\mu_N \in A)$ is an exponential decay function of N. In Figure 9, we can observe that the system experiences more transitions as σ becomes larger. Therefore, $\mathbf{P}(\mu_N \in A)$ becomes larger as σ becomes larger and this is confirmed by the fact that the rate function is approximately proportional to $1/\sigma^2$ (note, however, that σ is also in the operator $\mathcal{L}^*$). In Figure 10, we find that the probability of transition decreases as h increases. This is qualitatively consistent with the fact that h determines the height of the potential barrier between the two potential wells, in analogy with the classical problem of diffusion exit from a domain [17, Section 5.7].

8. **The nonsymmetric case.** Let us briefly revisit the previous analysis when the average velocity has the form [27]

$$\langle u \rangle_i = \begin{cases} \frac{1}{N_i} \sum_{j=1}^N u_j \phi(\|x_j - x_i\|), & \text{if } N_i > 0, \\ 0, & \text{if } N_i = 0, \end{cases} \tag{54}$$
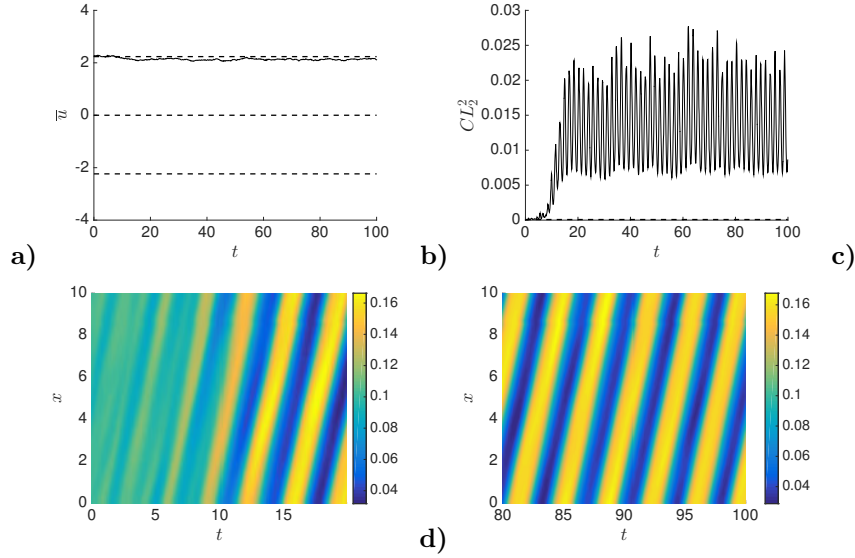
FIGURE 7. The empirical average velocity $\bar{u}^n$ (a), the square centered L^2-discrepancy $CL_2^2(n)$ (b), and the empirical position distribution smoothed by kernel density estimation (c and d). The initial positions $\{x_i^0\}_{i=1}^N$ are uniformly sampled over $[0, L]$ and the initial velocities $\{u_i^0\}_{i=1}^N$ are sampled from the Gaussian distribution with mean ξ_e and variance $\sigma^2/2$. Here $\Delta t = 0.1$, $N = 2000$, $\sigma = 1$, and $h = 5$. One can see that the average velocity is not ξ_e and the spatial distribution is not uniform.

where N_i is the weighted number of agents in the neighborhood of the ith agent:

$$N_i = \sum_{j=1}^N \phi(\|x_j - x_i\|). \tag{55}$$

If $\mu_N(0, dx, du)$ converges to a deterministic measure $\bar{\rho}(x, u)dxdu$, then $\mu_N(t, dx, du)$ converges to the deterministic measure $\rho(t, x, u)dxdu$ whose density is the solution of the nonlinear Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = -u\frac{\partial \rho}{\partial x} - \frac{\partial}{\partial u}\left[\left(G\left(\frac{\iint u'\phi(\|x'\|)\rho(t, x - x', u')du'dx'}{\iint \phi(\|x'\|)\rho(t, x - x', u')du'dx'}\right) - u\right)\rho\right] + \frac{1}{2}\sigma^2\frac{\partial^2 \rho}{\partial u^2}, \tag{56}$$

starting from $\rho(t = 0, x, u) = \bar{\rho}(x, u)$.

The nonlinear Fokker-Planck equation (56) has stationary states that we describe in the following proposition.

Proposition 5. *Let Ξ be the set of solutions of the compatibility condition equation*

$$\xi = G(\xi). \tag{57}$$

For any $\xi \in \Xi$, the state

$$\rho_\xi(x, u) = \frac{1}{L}F_\xi(u), \quad F_\xi(u) = \frac{1}{\sqrt{\pi\sigma^2}}\exp\left(-\frac{(u - \xi)^2}{\sigma^2}\right) \tag{58}$$
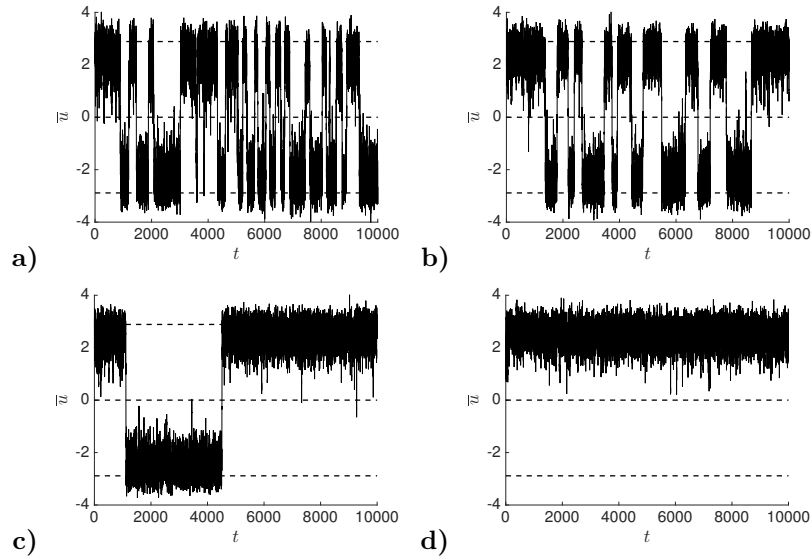
is a stationary solution of (56).

FIGURE 8. The empirical average velocity $\bar{u}^n$ at each time step t_n for $N = 80$ (a), $N = 100$ (b), $N = 120$ (c), and $N = 140$ (d). Here $\Delta t = 0.1$, $h = 6$ and $\sigma = 5$. The frequencies of the transitions between the two stable order states decay when N increases. The system has less transitions with a higher number of agents.
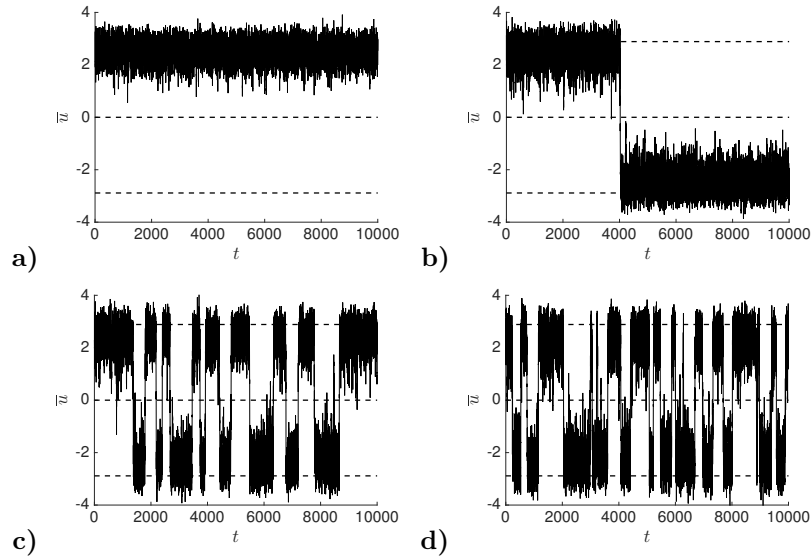

FIGURE 9. The empirical average velocity $\bar{u}^n$ at each time step t_n for $\sigma = 4$ (a), $\sigma = 4.5$ (b), $\sigma = 5$ (c), and $\sigma = 5.5$ (d). Here $\Delta t = 0.1$, $N = 100$, and $h = 6$. The frequencies of the transitions between the two stable order states increase when σ increases. The system has more transitions with a higher σ.
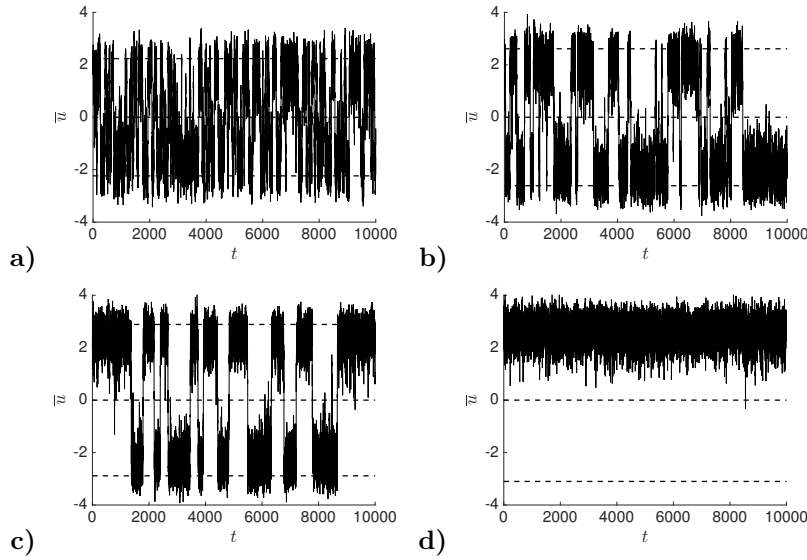
FIGURE 10. The empirical average velocity $\bar{u}^n$ at each time step t_n for $h = 5$ (a), $h = 5.5$ (b), $h = 6$ (c), and $h = 6.5$ (d). Here $\Delta t = 0.1$, $N = 100$, and $\sigma = 5$. The frequencies of the transitions between the two stable order states decay with h.

Note, however, that we could not prove that any stationary state is of the form (58). It is possible to prove that any stationary state that is uniform in space is of the form (58), but we could not prove that a stationary state must be uniform in space.

The linear stability analysis of the stationary states listed in Proposition 5 follows the same lines as in the symmetric case (2). Let $\xi \in \Xi$ and consider

$$\rho(t, x, u) = \rho_\xi(x, u) + \rho^{(1)}(t, x, u) = \frac{1}{L} F_\xi(u) + \rho^{(1)}(t, x, u), \tag{59}$$

for small perturbation $\rho^{(1)}$. We linearize the nonlinear Fokker-Planck equation (56)

$$\frac{\partial \rho^{(1)}}{\partial t} = -u \frac{\partial \rho^{(1)}}{\partial x} - \frac{\partial}{\partial u} \left[(\xi - u) \rho^{(1)} \right] \tag{60}$$
$$+ \frac{1}{L} G'(\xi) \left[\iiint (\xi - u') \phi(\|x'\|) \rho^{(1)}(t, x - x', u') du' dx' \right] F_\xi'(u) + \frac{1}{2} \sigma^2 \frac{\partial^2 \rho^{(1)}}{\partial u^2}.$$

The perturbation $\rho^{(1)}$ is periodic in x and can be expanded as (21). The Fourier coefficients $\hat{\rho}_k^{(1)}$ satisfy uncoupled equations

$$\frac{\partial \rho_k^{(1)}}{\partial t} = \frac{i 2 \pi k}{L} u \rho_k^{(1)} - \frac{\partial}{\partial u} \left[(\xi - u) \rho_k^{(1)} \right]$$
$$+ G'(\xi) \phi_k \left[\int (\xi - u') \rho_k^{(1)}(t, u') du' \right] F_\xi'(u) + \frac{1}{2} \sigma^2 \frac{\partial^2 \rho_k^{(1)}}{\partial u^2}. \tag{61}$$

The necessary and sufficient condition for the linear stability of the 0th order mode is

$$G'(\xi) < 1. \tag{62}$$

The sufficient condition for the linear stability of the other modes is that $|G'(\xi)\phi_k| <$ 1 for all $k \neq 0$ and the noise level σ should be larger than a threshold value σ_c, which is, however different from the threshold value of the symmetric case (1-2).

9. **Conclusion.** We have analyzed the Czirók model for the collective motion of locusts. The mean-field theory is used to obtain a nonlinear Fokker-Planck equation as the number of agents tends to infinity. We analyze the phase transition between the disorder and order states by the existence condition for the order states. We then perform the linear stability analysis on the stationary states, and prove that the order states are stable for a sufficiently large noise level and when the interaction between the velocities of the particles is neither too small nor too strong. We provide the fluctuation analysis and the large deviations principle. For a large but finite system we calculate the asymptotic, exponentially small transition probability from one order state to the other. Our analytical findings are verified by the extensive numerically simulations and are found in agreement with the experimental observations.

Appendix A. **Proof of Lemma 4.2.** We have

$$g_k(t,0) = \frac{\sigma^2}{4}D_k^2(1 - e^{-2t}) + (i\xi - \sigma^2 D_k)D_k(1 - e^{-t}) + (-i\xi D_k + \frac{\sigma^2}{2}D_k^2)t$$

$$\simeq (-i\xi D_k + \frac{\sigma^2}{2}D_k^2)t, \quad \text{as } t \to \infty,$$

$$\partial_\eta g_k(t,0) = \frac{\sigma^2}{2}D_k(1 - e^{-2t}) + (i\xi - \sigma^2 D_k)(1 - e^{-t}) \simeq (i\xi - \frac{\sigma^2}{2}D_k), \quad \text{as } t \to \infty.$$

Since $\hat{\rho}_k^{(1)}(0, \cdot)$ and $\partial_\eta \hat{\rho}_k^{(1)}(0, \cdot)$ are bounded we find that $\psi_k(t)$ decays exponential in time as $\exp(-\sigma^2 D_k^2 t/2)$, which gives the first item of the Lemma.

We can show that $R_k(t) \in L^1[0, \infty)$ because it is the product of a bounded function with $e^{-g_k(t,0)}$ which decays exponentially as $t \to \infty$. Note that it is important that $k \neq 0$ and $\sigma > 0$ to ensure the decay. Recall that $D_k = 2\pi k/L$ and $\beta_k = D_k(1 - e^{-t})$. Then we can write $g_k(t,0)$, $\partial_\eta g_k(t,0)$, $H_\xi(-\beta_k)$, and $H_\xi'(-\beta_k)$ in terms of D_k and β_k:

$$g_k(t,0) = -\frac{1}{4}\sigma^2\beta_k^2 - \frac{1}{2}\sigma^2 D_k\beta_k + i\xi\beta_k - i\xi D_k t + \frac{1}{2}\sigma^2 D_k^2 t,$$

$$\partial_\eta g_k(t,0) = \frac{1}{2}\sigma^2(e^{-t} - 1)\beta_k + i\xi(1 - e^{-t}),$$

$$H_\xi(-\beta_k) = -\beta_k \hat{F}_\xi(-\beta_k) = -\beta_k \exp\left(i\xi\beta_k - \frac{1}{4}\sigma^2\beta_k^2\right),$$

$$H_\xi'(-\beta_k) = \left(1 + i\xi\beta_k - \frac{1}{2}\sigma^2\beta_k^2\right)\exp\left(i\xi\beta_k - \frac{1}{4}\sigma^2\beta_k^2\right).$$

Therefore,

$$-\partial_\eta g_k(t,0)H_\xi(-\beta_k) + e^{-t}H_\xi'(-\beta_k) = \left(-\frac{1}{2}\sigma^2\beta_k^2 + i\xi\beta_k + e^{-t}\right)\exp\left(i\xi\beta_k - \frac{1}{4}\sigma^2\beta_k^2\right),$$

and we can also write $R_k(t)$ in terms of D_k and β_k:

$$R_k(t) = G'(\xi)\phi_k\left[-\partial_\eta g_k(t,0)H_\xi(-\beta_k) + e^{-t}H_\xi'(-\beta_k)\right]e^{-g_k(t,0)}$$

$$= G'(\xi)\phi_k\left(-\frac{1}{2}\sigma^2\beta_k^2 + i\xi\beta_k + e^{-t}\right)\exp\left[\frac{\sigma^2}{2}D_k\beta_k + i\xi D_k t - \frac{\sigma^2}{2}D_k^2 t\right].$$

The L^1-norm of R_k can be bounded by

$$\int_0^\infty |R_k(t)|dt \leqslant |G'(\xi)\phi_k| \int_0^\infty \left|-\frac{1}{2}\sigma^2\beta_k^2+i\xi\beta_k+e^{-t}\right| \exp\left[\frac{\sigma^2}{2}D_k\beta_k-\frac{\sigma^2}{2}D_k^2t\right]dt. \quad (63)$$

Because $\beta_k = D_k(1 - e^{-t})$, we use the following bounds for (63):

$$\left|-\frac{1}{2}\sigma^2\beta_k^2+i\xi\beta_k+e^{-t}\right| \leqslant \frac{\sigma^2}{2}D_k^2(1-e^{-t})^2 + |\xi||D_k|(1-e^{-t}) + e^{-t}$$

$$\leqslant \begin{cases} \frac{\sigma^2}{2}D_k^2t^2 + |\xi||D_k|t+1, & 0 \leqslant t \leqslant 1, \\ \frac{\sigma^2}{2}D_k^2 + |\xi||D_k| + e^{-t}, & 1 < t, \end{cases}$$

$$\exp\left[\frac{\sigma^2}{2}D_k\beta_k - \frac{\sigma^2}{2}D_k^2t\right] = \exp\left[\frac{\sigma^2}{2}D_k^2(1-e^{-t}-t)\right] \leqslant \begin{cases} e^{\frac{\sigma^2}{2}D_k^2(-\frac{1}{4}t^2)}, & 0 \leqslant t \leqslant 1, \\ e^{\frac{\sigma^2}{2}D_k^2(1-t)}, & 1 < t. \end{cases}$$

We have the following estimates:

$$\int_0^1 \left(\frac{\sigma^2}{2}D_k^2t^2 + |\xi||D_k|t+1\right)e^{\frac{\sigma^2}{2}D_k^2(-\frac{1}{4}t^2)}dt \leqslant \frac{\sqrt{2\pi}}{\frac{\sigma}{2}|D_k|}\left(1+\sqrt{\frac{2}{\pi}}\frac{|\xi|}{\sigma}+\frac{1}{2}\right),$$

$$\int_1^\infty \left(\frac{\sigma^2}{2}D_k^2 + |\xi||D_k| + e^{-t}\right)e^{\frac{\sigma^2}{2}D_k^2(1-t)}dt = 1 + \frac{|\xi|}{\frac{\sigma^2}{2}|D_k|} + \frac{e^{-1}}{1+\frac{\sigma^2}{2}D_k^2}.$$

Therefore a sufficient condition to ensure that $\int_0^\infty |R_k(t)|dt < 1$ is

$$|G'(\xi)\phi_k| < \left(1 + \frac{3\sqrt{2\pi}}{\sigma|D_k|} + \frac{3|\xi|}{\frac{\sigma^2}{2}|D_k|} + \frac{e^{-1}}{1+\frac{\sigma^2}{2}D_k^2}\right)^{-1}. \quad (64)$$

By the fact that $|D_k| = 2\pi|k|/L$ is increasing in $|k|$, we can have a simplified sufficient condition from (64):

$$|G'(\xi)\phi_k| < \left(1 + \frac{3\sqrt{2\pi}}{\sigma D_1} + \frac{3|\xi|}{\frac{\sigma^2}{2}D_1} + \frac{e^{-1}}{1+\frac{\sigma^2}{2}D_1^2}\right)^{-1}. \quad (65)$$

If $|G'(\xi)\phi_k| < 1$ for all nonzero k, then this condition is satisfied for all nonzero k if σ is large enough. This completes the proof of the second item of the Lemma.

REFERENCES

[1] G. Ariel and A. Ayali, Locust collective motion and its modeling, *PLoS Comput Biol*, **11** (2015), e1004522.

[2] G. Ariel, Y. Ophir, S. Levi, E. Ben-Jacob and A. Ayali, Individual pause-and-go motion is instrumental to the formation and maintenance of swarms of marching locust nymphs, *PLoS ONE*, **9** (2014), e101636.

[3] E. Bertin, M. Droz and G. Grégoire, Boltzmann and hydrodynamic description for self-propelled particles, *Phys. Rev. E*, **74** (2006), 022101.

[4] E. Bertin, M. Droz and G. Grégoire, Hydrodynamic equations for self-propelled particles: Microscopic derivation and stability analysis, *Journal of Physics A: Mathematical and Theoretical*, **42** (2009), 445001.

[5] E. Bertin, H. Chaté, F. Ginelli, S. Mishra, A. Peshkov and S. Ramaswamy, Mesoscopic theory for fluctuating active nematics, *New Journal of Physics*, **15** (2013), 085032.

[6] A. Budhiraja, P. Dupuis and M. Fischer, Large deviation properties of weakly interacting processes via weak convergence methods, *Ann. Probab.*, **40** (2012), 74–102.

[7] J. Buhl, D. J. T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller and S. J. Simpson, From disorder to order in marching locusts, *Science*, **312** (2006), 1402–1406.

[8] J.-B. Caussin, A. Solon, A. Peshkov, H. Chaté, T. Dauxois, J. Tailleur, V. Vitelli, and D. Bartolo, Emergent spatial structures in flocking models: A dynamical system insight, *Phys. Rev. Lett.*, **112** (2014), 148102.

[9] H. Chaté, F. Ginelli, G. Grégoire, F. Peruani and F. Raynaud, Modeling collective motion: variations on the vicsek model, *The European Physical Journal B*, **64** (2008), 451–456.

[10] I. D. Couzin, J. Krause, N. R. Franks and S. A. Levin, Effective leadership and decision-making in animal groups on the move, *Nature*, **433** (2005), 513–516.

[11] A. Czirók, A.L. Barabási and T. Vicsek, Collective motion of self-propelled particles: Kinetic phase transition in one dimension, *Phys. Rev. Lett.*, **82** (1999), 209–212.

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

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

[14] P. Degond and S. Motsch, Macroscopic limit of self-driven particles with orientation interaction, *C.R. Math. Acad. Sci. Paris*, **345** (2007), 555–560.

[15] P. Degond and H. Yu, Self-Organized Hydrodynamics in an annular domain: modal analysis and nonlinear effects, *Mathematical Models and Methods in Applied Sciences*, **25** (2015), 495–519.

[16] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, Addison-Wesley, Boston, 2012.

[17] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, Berlin, 1998.

[18] L. Edelstein-Keshet, J. Watmough and D. Grunbaum, Do travelling band solutions describe cohesive swarms? an investigation for migratory locusts, *Journal of Mathematical Biology*, **36** (1998), 515–549.

[19] C. Escudero, C. A. Yates, J. Buhl, I. D. Couzin, R. Erban, I. G. Kevrekidis and P. K. Maini, Ergodic directional switching in mobile insect groups, *Phys. Rev. E*, **82** (2010), 011926.

[20] K.-T. Fang, C.-X. Ma and P. Winker, Centered L2-discrepancy of random sampling and Latin hypercube design, and construction of uniform designs, *Math. Comp.*, **71** (2002), 275–296.

[21] J. Garnier, G. Papanicolaou and T.-W. Yang, Consensus convergence with stochastic effects, *Vietnam Journal of Mathematics*, **45** (2017), 51–75.

[22] J. Gärtner, On the McKean-Vlasov limit for interacting diffusions, *Math. Nachr.*, **137** (1988), 197–248.

[23] S.-Y. Ha and E. Tadmor, From particle to kinetic and hydrodynamic descriptions of flocking, *Kinetic and Related Models*, **1** (2008), 415–435.

[24] F. J. Hickernell, A generalized discrepancy and quadrature error bound, *Math. Comp.*, **67** (1998), 299–322.

[25] T. Ihle, Kinetic theory of flocking: Derivation of hydrodynamic equations, *Phys. Rev. E*, **83** (2011), 030901.

[26] S. Mishra, A. Baskaran and M. C. Marchetti, Fluctuations and pattern formation in self-propelled particles, *Phys. Rev. E*, **81** (2010), 061916.

[27] S. Motsch and E. Tadmor, A new model for self-organized dynamics and its flocking behavior, *J. Stat. Phys.*, **144** (2011), 923–947.

[28] S. Motsch and E. Tadmor, Heterophilious dynamics enhances consensus, *SIAM Review*, **56** (2014), 577–621.

[29] O. J. O'Loan and M. R. Evans, Alternating steady state in one-dimensional flocking, *Journal of Physics A: Mathematical and General*, **32** (1999), L99.

[30] A. M. Reynolds, G. A. Sword, S. J. Simpson and D. R. Reynolds, Predator percolation, insect outbreaks, and phase polyphenism, *Current Biology*, **19** (2009), 20–24.

[31] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner and L. Schimansky-Geier, Active brownian particles, *The European Physical Journal Special Topics*, **202** (2012), 1–162.

[32] A. V. Savkin, Coordinated collective motion of groups of autonomous mobile robots: Analysis of Vicsek's model, *IEEE Trans. Automat. Control,* **49** (2004), 981–983.

[33] A. P. Solon and J. Tailleur, Revisiting the flocking transition using active spins, *Phys. Rev. Lett.*, **111** (2013), 078101.

[34] A. P. Solon and J. Tailleur, Flocking with discrete symmetry: The 2d active Ising model, *Phys. Rev. E*, **92** (2015), 042119.

[35] A. P. Solon, J.-B. Caussin, D. Bartolo, H. Chaté and J. Tailleur, Pattern formation in flocking models: A hydrodynamic description, *Phys. Rev. E*, **92** (2015), 062111.

[36] J. Toner and Y. Tu, Long-range order in a two-dimensional dynamical XY model: How birds fly together, *Phys. Rev. Lett.*, **75** (1995), 4326–4329.

[37] J. Toner and Y. Tu, Flocks, herds, and schools: A quantitative theory of flocking, *Phys. Rev. E*, **58** (1998), 4828–4858.

[38] C. M. Topaz, M. R. D'Orsogna, L. Edelstein-Keshet, and A. J. Bernoff, Locust dynamics: Behavioral phase change and swarming, *PLoS Comput Biol*, **8** (2012), e1002642, 11 pp.

[39] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen and O. Shochet, Novel type of phase transition in a system of self-driven particles, *Phys. Rev. Lett.*, **75** (1995), 1226–1229.

[40] T. Vicsek and A. Zafeiris, Collective motion, *Phys. Rep.*, **517** (2012), 71–140.

[41] C. A. Yates, R. Erban, C. Escudero, I. D. Couzin, J. Buhl, I. G. Kevrekidis, P. K. Maini and D. J. T. Sumpter, Inherent noise can facilitate coherence in collective swarm motion, *Proceedings of the National Academy of Sciences*, **106** (2009), 5464–5469.

E-mail address: josselin.garnier@polytechnique.edu
E-mail address: papanicolaou@stanford.edu
E-mail address: yangx953@umn.edu