COMPLEXITY REDUCTION IN MANY PARTICLES SYSTEMS WITH RANDOM INITIAL DATA.∗

LEONID BERLYAND†, PIERRE-EMMANUEL JABIN‡, AND MYKHAILO POTOMKIN§

Abstract. We consider the motion of interacting particles governed by a coupled system of ODEs with random initial conditions. Our goal is to explain the experimentally observed emergence of correlations in the collective state. We propose a new approach based on the truncation of BBGKY hierarchy that allows one to drastically reduce computational complexity. Our reduction of complexity is based on ideas similar to ideas of the classical Mean Field approach, but it also allows one to capture correlations. Our theoretical approach is validated by numerical simulations.

Key words. Mean Field, correlation, systems of a large number of particles.

1. Motivation and Settings. Systems of interacting particles described by coupled system of a large number of ODEs with random initial conditions appear in many problems of physics, cosmology, chemistry, biology, social science and economics:

\[ \dot{X}_i = \frac{\alpha}{N} \sum_{j=1}^{N} K(X_i - X_j), \quad i = 1, ..., N. \] (1.1)

Here \( X_i(t) \) denotes the position of \( i \)th particle and \( X_i \) belongs to \( D \), where \( D \) throughout this paper can stand for \( \mathbb{R}^d \), \( d \)-dimensional torus \( \mathbb{T}^d \) or a compact domain in \( \mathbb{R}^d \) in which case boundary conditions must be added.

System (1.1) is Individuum Based Model, i.e., it has an ODE for each single particle coupled with others. In various applications the role of individuum can be played by atoms, social agents, bacteria in suspensions, animals in flocks, etc. The system (1.1) has two key parameters: \( \alpha \), strength of interactions, and \( N \), number of particles. The parameter \( \alpha \) is determined by both geometry such as interparticle distance and size of a particle (note that a model particle is just a point) and physics, e.g., magnitude of self-propulsion. In this paper we restrict ourselves to the case when the right hand side of (1.1) is linear in \( \alpha \). Magnitude of \( \alpha \) plays a key role in our analysis of the system (1.1): small \( \alpha \) corresponds to almost decoupled particles; large \( \alpha \) corresponds to strong interactions which is our main focus; \( \alpha \sim 1 \) corresponds to the classical mean field regime.

Our work is motivated by experiments in bacterial suspensions \[ [1, 2, 3, 4, 5, 6, 7] \]. The experiments \[ [3, 4, 7] \] show emergence of a coarse collective scale when the concentration of bacteria exceeds a critical value. Roughly speaking, the collective scale is the correlation length of the velocity field in bacterial suspension. A striking universality property has been observed experimentally and numerically in \[ [3, 4, 8] \] the collective scale does not change when swimming speed and concentration increase (that is more energy is injected into the system).

∗The work of Leonid Berlyand and Mykhailo Potomkin was supported by DOE DE-FG-0208ER25862. The work of Pierre-Emmanuel Jabin was partially supported by NSF DMS-1312142.
†Department of Mathematics, The Pennsylvania State University, University Park, PA, USA, 16802, (berlyand@math.psu.edu).
‡Department of Mathematics, University of Maryland, College Park, MD 20742 USA, (pjabin@umd.edu)
§Department of Mathematics, The Pennsylvania State University, University Park, PA, USA, 16802, (potomkin@math.psu.edu)
The motion of bacteria can be modeled by a system of the form (1.1) where position and orientation of the $i$th bacteria are described by vector $X_i(t)$. In this case the parameter $\alpha$ equals to $NV_0(\ell/R)^2$, where $V_0$ is swimming speed of a single bacterium, $R$ is mean distance between two bacteria, $\ell$ is the characteristic size of a bacterium. The collective behavior observed in experiments \cite{3,4,7} has been also qualitatively reproduced by direct numerical simulations in \cite{8}, which validates the model of the type (1.1). However, computational cost of direct simulations of the ODE system is prohibitively high for the following reasons:

(i) number of bacteria $N$ is very large ($10^{10}$ per sm$^3$);
(ii) to draw a reliable conclusion one needs to consider many realizations which mathematically translates into random initial data.
(iii) we are mainly interested in collective state corresponding to large $\alpha$ which leads to small time steps;

Combination of the factors (i)-(iii) makes computational cost too high even for the most powerful particles methods such as Fast Multipole Method \cite{9,10,11}.

The goal of this paper is to propose computational approach that allows us to describe numerically the collective state of this system with properties (i)-(iii). More specifically, the collective state is described by the correlation length and two-point correlation function. The objective of our study is efficient computation of these two quantities.

The main idea is to replace the ODE system (1.1) by a PDE such that computational cost of its solution does not grow as $N$ goes to infinity. This idea had been used in classical mean field approach which corresponds to $\alpha$ of the order of 1 and is not valid for strong interactions, e.g., $\alpha \sim N^\gamma$ for $0 < \gamma < 1$.

This paper focuses on a PDE approach that extends beyond the mean field, so it can capture correlations in a computationally efficient way such that the computational complexity does not grow with $N$. The main idea here is to consider BBGKY hierarchy of PDEs which consists of $N$ equations (therefore it is even harder to solve than (1.1)) and obtain a closed system for 1-particle distribution and 2-particle distributions by a clever truncation of the hierarchy. Then the large parameter $N$ is present only in coefficients in an innocuous way, so it does not increase computational complexity as $N \to +\infty$. This approach computes distribution functions and therefore it avoids computing individual realizations. Thus, it allows us to overcome computational difficulties (i) and (ii). The contribution to the computational complexity from difficulty (iii) is much less than (i) and (ii), because $\alpha \sim N^\gamma$ and $\gamma < 1$.

Note that a specific feature of our method is that it is efficient for random initial conditions of system (1.1) rather than deterministic. Indeed, a seemingly simpler case of deterministic initial data leads to a solution of the truncated BBGKY hierarchy with singular initial conditions ($\delta$-functions) which is why numerical cost of solving such deterministic problem is very high. By contrast, random initial data in ODE (1.1) lead to smooth initial conditions in the truncated BBGKY hierarchy that is much easier to handle numerically.

The paper is organized as follows. We recall the mean field approach and discuss its limitations in Section 2. The truncation of BBGKY hierarchy is described in Section 3. Numerics performed to check that the truncated PDE system is reliable are described in Section 4.

2. Random initial conditions, correlations, the Mean Field approach.

For physical reason, the initial conditions for the System (1.1) are typically random as explained below.
In the classical Mean Field theory, this leads to a drastic reduction in the computational complexity: it is possible to approximate the original solution by the solution of a PDE which does not depend on $N$.

We describe here the two classical approaches, first one is based on the so-called empirical measure and a statistical approach which is better suited for our purpose. The Mean Field limit is valid as long as the correlations between particles are vanishing. This phenomenon is known as propagation of chaos. However, our work is motivated by experimental studies of the collective state, whose key feature is the rise of correlations that corresponds to emergence of a collective scale. In this case, as we will explain below, the Mean Field approach fails.

Our approach in this paper is mostly formal. Nevertheless, we point out that the mean field theory described below can be made rigorous if some smoothness is assumed on $K$. More precisely,

$$\nabla K \in L^\infty(D), \quad K(x) \to 0 \text{ as } |x| \to \infty.$$  

On the other hand, we believe that numerical implementation of this approach will work well even for singular kernels (see remark 2.1).

### 2.1. Preliminaries.

**How to choose initial conditions: Randomness and the marginals.** By Assumption (2.1) and the standard Cauchy-Lipschitz theory, there exists a unique solution to (1.1) once each initial position $X_i(0)$ is chosen. However for most practical purposes, determining those initial positions can be a very delicate problem as the full information is not accessible from an experimental point of view. For $N \sim 10^{10}$, it is indeed completely unrealistic to measure with enough precision the position of each particle.

Instead, statistical information about the positions of the particles is accessible. Hence one usually assumes that the initial position of each particle is randomly distributed. That means that the information on the initial distribution of the particles is now encoded in the $N$-particle distribution function at time 0, $f_N(t = 0, x_1, ..., x_N)$. Given a subdomain $Q \subset D^N$, the probability of finding the initial positions $(X^0_1, ..., X^0_N) \in Q$ is given by

$$\int_Q f_N(t = 0, x_1, ..., x_N) \, dx_1 \ldots dx_N.$$  

System (1.1) is deterministic but if the initial conditions are random then the randomness will be propagated defining the $N$-particle distribution at $t > 0$. Technically $f_N(t, \cdot)$ is the push forward of $f_N(t = 0, \cdot)$ by the flow generated by (1.1).

From $f_N$ one may define the $k$-th marginal

$$f_k(t, x_1, ..., x_k) = \int_{D^{N-k}} f_N(t, x_1, ..., x_k, x_{k+1}, ..., x_N) \, dx_{k+1} \ldots dx_N.$$  

Some of marginals have natural physical interpretation. For instance, $f_1$ is the 1-particle distribution function and for $\mathcal{O} \subset D$ the average number of particles in the subset $\mathcal{O}$ is

$$\int_{\mathcal{O}} f_1(t, x) \, dx.$$  

It is still not experimentally possible to measure $f_N$ but it is possible to measure some marginals, especially $f_1$ and the 2-particle distribution function $f_2$ (see below).
In the simplest case, one assumes that the particles are initially independently and identically distributed, that is

\[ f_N(t = 0, x_1, ..., x_N) = \Pi_{i=1}^{N} f^0(x_i). \]

This independence is strongly connected to the usual mean field limit approach as explained in subsection 2.2 (see (2.13)).

**Definition of correlations.** Our main goal is to understand how correlations develop in System (1.1). Those are connected to the second marginal \( f_2 \).

We define correlation of particles’ positions by

\[
\begin{align*}
    c &= \frac{\mathbb{E}[X_1 \cdot X_2] - (\mathbb{E}[X])^2}{\mathbb{E}[X^2] - (\mathbb{E}[X])^2} \\
    &= \frac{\int x_1 \cdot x_2 f_2(x_1, x_2) dx_1 dx_2 - \left( \int x f_1(x) dx \right)^2}{\int x^2 f_1(x) dx - \left( \int x f_1(x) dx \right)^2}.
\end{align*}
\]

Observe that the correlation \( c \) can only vanish if \( f_2(x, y) = f_1(x) f_1(y) \), that is if the particles positions are independent. Therefore, roughly speaking, the correlations in the system are determined by how far \( f_2(x, y) \) is from \( f_1(x) f_1(y) \).

### 2.2. The Mean Field approach.

**Empirical measure.** Assume that the \( X_i(t) \) are solutions to (1.1), and define the empirical measure

\[
\mu_N(t, x) = \frac{1}{N} \sum_{i=1}^{N} \delta(x - X_i(t)).
\]

Note that if the particles are undistinguishable then there is just as much information in the empirical measure as in the positions vector \( (X_1, ..., X_N) \). Otherwise, it only tells that there is a particle at \( x \), but it is not clear which one.

If \( K \) is continuous, then \( \mu_N \) solves the Vlasov equation in the sense of distribution

\[
\partial_t f + \alpha \nabla_x \cdot \left( \int K(y-x) f(y) dy f(x) \right) = 0.
\]

Consider a sequence of initial positions \( X_N = \{ X_i(0) : i = 1, ..., N \} \) such that the corresponding empirical measure \( \mu_N(0) \) converges to some \( f^0 \in \Pi(D) \) as \( N \) goes to infinity. Here \( \Pi(D) \) is the space of positive measures \( \mu \) on \( D \) such that \( \mu(D) = 1 \). Then it is natural to expect that \( \mu_N \) will also converge to the corresponding solution \( f \) to (2.5) with initial data \( f^0 \). Assuming that \( f^0 \) is smooth then it is possible to compute numerically \( f \) and hence to get a good approximation to \( \mu_N \). This is the classical mean field limit theory which can be made quantitative.

Those quantitative estimates require some weak distances on the space of measures. Those are classically the so-called Monge-Kantorovich-Wasserstein (MKW) distances. For our purpose it is enough to understand that they measure some appropriate distance between probability measures. For the sake of completeness we define these distances in Appendix A.

We now give the main stability estimate behind the mean field limit. From [12], [13], and [14], it is possible to prove that if \( f \) and \( g \) are two measure-valued solutions to (2.5), then

\[
W_p(f(t, .), g(t, .)) \leq e^{t \alpha \|\nabla K\|_{L^\infty}} W_p(f(0, .), g(0, .)),
\]

where \( W_p \) is the Wasserstein distance of order \( p \).
where $W_p(\cdot, \cdot)$ is a $p$-Wasserstein or MKW distance between two measures. The inequality (2.6) is a Gronwall-type inequality. Note also that the inequality (2.6) applies for any initial conditions $f(0, \cdot)$ and $g(0, \cdot)$ which are not necessarily random.

**The Mean field limit.** In our context, the initial conditions are random as it was explained before. In particular, the empirical measure at time $t = 0$ is itself random.

If the initial law is chosen according to (2.2), then a large deviation for law of large numbers applies ($\mathbb{E} \mu_N = f^0$) and ensures that, in fact, the initial measure $\mu_N(t = 0)$ is very close to $f^0$. More precisely, it is proved for example in Boissard [15, Appendix A, Proposition 1.2], that if $f^0$ is a nonnegative measure with compact support of diameter $R$, then for some constant $C$ and positive coefficients $\gamma_1$ and $\gamma_2$, depending only on the dimension $d$ of $D$ and $R$

\[
\mathbb{P}\left( W_1(\mu_N(t = 0), f^0) \geq \frac{CR}{N^{\gamma_1}} \right) \leq e^{-CN^{\gamma_2}}.
\]

This says that with exponentially large probability, $\mu_N(t = 0)$ and $f^0$ are polynomially close in $N$. Denote by $f$ the solution to (2.5) with $f^0$ as an initial data. By combining the deterministic stability (2.6) with a law of large numbers in the form (2.7) we obtain that with probability larger than $(1 - e^{-CN^{\gamma_2}})$

\[
W_1(\mu_N(t, \cdot), f(t, \cdot)) \leq \frac{CR}{N^{\gamma_1}} e^{t \alpha \|\nabla K\|_{L^\infty}}.
\]

**Why random initial conditions make computations much easier in the mean field framework?** Looking for a solution of the Vlasov equation (2.5) in a form of a sum of $N$ Dirac masses like $\mu_N$ is just as complex and computationally costly as solving the original ODE system (1.1).

However looking for smooth solutions to the Vlasov equation is comparatively much faster and obviously independent of $N$ (provided the solution is independent of $N$). Since the initial distribution $f^0$ is usually assumed to be smooth, the corresponding solution $f$ is smooth as well and computing $f$ numerically is thus far easier than solving (1.1) and the cost is independent of $N$.

The key to the reduction in the computational complexity in this mean field limit approach is that one does not solve the original ODE system (1.1) but instead one solves the Vlasov PDE for $f$. The previous inequality (2.8) implies that this $f$ will be a good approximation of the original $\mu_N$ up to a time $t$ of order

\[
\log N \leq \alpha \|\nabla K\|_{L^\infty}.
\]

Note that in certain circumstances, this time can be considerably extended to become polynomial in $N$. This usually requires to have a stable equilibrium to Eq. (2.5), see [16] for instance.

**2.3. Propagation of chaos.** It is possible to interpret the mean field limit in terms of propagation of chaos on the marginals. For this, we introduce the hierarchy of equations on the marginals.

First, it is well-known that $f_N$ solves the Liouville equation:

\[
\partial_t f_N + \frac{\alpha}{N} \sum_{i=1}^N \partial x_i \left( \sum_{j=1}^N K(x_j - x_i) f_N \right) = 0.
\]
By integrating the equation on $f_N$, one obtains an equation satisfied by each $f_k$
\begin{equation}
\partial_t f_k + \frac{\alpha}{N} \sum_{i=1}^{k} \sum_{j=1}^{k} \partial_i (K(x_j - x_i) f_k) + \frac{\alpha(N - k)}{N} \sum_{i=1}^{k} \partial_i \int K(y-x_i) f_{k+1}(t, x_1, \ldots, x_k, y) \, dy = 0.
\end{equation}

For example, the PDE for $f_1$
\begin{equation}
\partial_t f_1(t,x_1) + \frac{\alpha K(0)}{N} \partial_{x_1} f_1(t,x_1) + \frac{\alpha}{N} \partial_{x_1} \cdot \left\{ \int K(y-x_1) f_2(t,x_1,y) \, dy \right\} = 0.
\end{equation}

By taking the formal limit $N \to +\infty$ in the equation (2.12) and assuming the independence condition $f_2(t,x_1,x_2) = f_1(t,x_1) f_1(t,x_2)$, we get equation (2.5).

This leads us to the crucial concept of propagation of chaos. Under some mild conditions on smoothness of $K$, for the initial positions that are close to being independent (that is (2.2) is assumed) as $N \to \infty$ we have
\begin{equation}
f_k(t,x_1,\ldots,x_k) \to \Pi_{i=1}^{k} f(t,x_i),
\end{equation}
where $f(t,x)$ solves the mean field equation (2.5).

Note that for a finite $N$, one cannot have equality in (2.13) and, in particular, $\Pi_{i=1}^{N} f(t,x_i)$ cannot be a solution to (2.10). Hence, for a finite but large $N$ and for initial conditions of the form $f_N|_{t=0} = \Pi_{i=1}^{N} f_0(x_i)$, the particles’ positions are not independent but their correlation is very small, at least on the time interval when the mean field limit holds, i.e. up to a time of order (2.9).

**Beyond Mean Field.** The mean field limit predicts that correlations are vanishing. In our context however this would imply a correlation as defined in (2.3) is close to 0 which is in contradiction with the experimentally observed phenomenon which we wish to explain.

Let us recall here that System (1.1) is scaled with a $1/N$ factor. In a given experimental setting, there is no particular reason why $\alpha$ should be of order 1. Instead because of the scaling, the coefficient $\alpha$ is typically polynomial in $N$. Hence the time $T$ defined by (2.9) is also very small. After this time during which the mean field theory guarantees that correlations are small, those correlations may start developing and creating the collective behavior we wish to explain.

In general computation of those correlations would require to solve the full equation (2.10) on $f_N$. Unfortunately, $f_N$ is a function of $N+1$ variables and computational cost of the numerical solution of (2.10) is much too large.

We would like to compute directly some of the marginals $f_k$, in particular $f_1$ and $f_2$, since $f_2$ essentially capture correlations. However, the equation (2.11) on $f_k$ is never closed unless $k = N$ as it involves $f_{k+1}$. This is the problem we focus on in this paper; we propose a solution consisting in truncating the hierarchy by choosing an ansatz for $f_3$ in terms of $f_1$ and $f_2$.

**Remark 2.1** (On singular kernels) As mentioned before, the rigorous justification of classical mean field theory requires some smoothness on the interaction kernel, $K$ Lipschitz. Many physical kernels are more singular, in particular in the context we are interested in, i.e., the context of bacteria interacting through a fluid.

It is widely conjectured that the mean field theory can be extended to more singular kernels and some results are already available, see for example [17], [18], [19] or [20] in the phase space framework.
In this work, we are not concerned with rigorous justification of our results under proper assumptions on smoothness of \(K\), however, just as in the Mean Field approach, we believe that the numerical implementation of our approach will work well for a wide class of kernels \(K\) (including singular ones).

3. Truncation of hierarchy. In this section we first discuss possibility of truncation ansatz \(f_3 = F[f_1, f_2]\) such that the full BBGKY hierarchy becomes a system of two PDEs for marginals \(f_1\) and \(f_2\) only, and the following properties hold for its solution:

1. \(f_2(x_1, x_2) = f_2(x_2, x_1)\) (particles are identical);
2. \(\int f_2 \equiv \text{const, } f_1, f_2 \geq 0\) provided that initial conditions for \(f_1\) and \(f_2\) are positive (mass preserving and positivity);
3. \(f_1 = \int f_2 \) (consistency);
4. If \(f_2 = f_1 \otimes f_1\), then \(f_3(x_1, x_2, x_3) = F(f_1, f_1 \otimes f_1) = f_1(x_1)f_1(x_2)f_1(x_3)\).

By \(f_2 = f_1 \otimes f_1\) we mean that for all \(x, y\) the equality holds \(f_2(x, y) = f_1(x)f_1(y)\).

We reformulate some of these properties as requirements on functional \(F\) and then prove that such an ansatz does not exist. Next, we present a truncation which is not based on 'a unique ansatz', and its solution satisfies four properties above.

Consider a representation for \(f_3\):

\[
(3.1) \quad f_3(x_1, x_2, x_3) = F[f_1, f_2](x_1, x_2, x_3),
\]

where \(F\) is a functional of \(f_1\) and \(f_2\).

Now we reformulate key properties as requirements on \(f_3\).

1. Symmetry with respect to arguments \(x_1\) and \(x_2\), i.e., \(f_2(x_1, x_2) = f_2(x_2, x_1)\) is equivalent to

\[
(3.2) \quad f_3(x_1, x_2, x_3) = f_3(x_2, x_1, x_3) \text{ for all } x_1, x_2, x_3.
\]

2. Positivity of \(f_1\) and \(f_2\) will follow, if

\[
(3.3) \quad \text{for all } x_1, x_2 : (f_2(x_1, x_2) = 0 \Rightarrow f_3(x_1, x_2, x_3) = 0 \text{ for all } x_3)
\]

The requirement (3.3) implies that there exists a function \(h(x_1, x_2, x_3)\) such that \(f_3(x_1, x_2, x_3) = h(x_1, x_2, x_3)f_2(x_1, x_2)\). Indeed, if \(f(x_1, x_2) \neq 0\), then

\[
h(x_1, x_2, x_3) = \frac{f_3(x_1, x_2, x_3)}{f_2(x_1, x_2)} \text{ for all } x_3.
\]

If \(f_2(x_1, x_2) = 0\), then \(h\) can be defined arbitrarily. By method of characteristics, this property implies positivity of solutions to truncated system provided that initial data is positive.

3. In order to have consistency property \(f_1(x) = \int f_2(x, y)dy\) we impose

\[
(3.4) \quad f_2(x_1, x_2) = \int f_3(x_1, x_3, x_2)dx_3
\]

The equality (3.4) is equivalent to the statement that if we integrate equation for \(k = 2\) from the BBGKY hierarchy with respect to one of the spatial variables, say, \(x_2\), we get equation for \(k = 1\).

4. In addition to key properties mentioned in Introduction we add one more natural requirement:

\[
(3.5) \quad (f_2(x_1, x_2) \geq 0 \text{ for all } x_1, x_2) \Rightarrow (f_3(x_1, x_2, x_3) \geq 0 \text{ for all } x_1, x_2, x_3).
\]
Proposition 3.1. There is no such representation (3.1) that all requirements (3.2), (3.3), (3.4) and (3.5) hold true.

Proof. The proof is by contradiction. The idea is to combine requirements 20 and 30.

\[(3.6)\quad f_2(x_1, x_2) = \int h(x_1, x_3, x_2) f_2(x_1, x_3) dx_3\]

and to find such \(f_2\) that the LHS of (3.6) is zero, but the RHS is not zero.

Take

\[\Omega = \{(x_1, x_2) : |x_1 - 1/2| + |x_2 - 1/2| < 1/2\} \setminus \{|x_1 - 1/2| < 1/4, |x_2 - 1/2| < 1/4\}\]

and \(f_2(x_1, x_2) = \frac{1}{|I|} \chi_{\Omega}(x_1, x_2)\). Note \(f_1(x) > 0\) for all \(x \in (0, 1) \setminus \{1/4, 3/4\}\).

The property (3.3) implies existence of such function \(h(x_1, x_2, x_3)\) that \(f_3(x_1, x_2, x_3) = h(x_1, x_2, x_3) f_2(x_1, x_2)\). Thus, from (3.5) we obtain

\[(3.7)\quad f_2(x_1, x_2) = \int h(x_1, x_3, x_2) f_2(x_1, x_3) dx_3.\]

Let \((x_1, x_2) \notin \Omega\), then (3.7) implies that

\[(3.8)\quad 0 = \int h(x_1, x_3, x_2) f_2(x_1, x_3) dx_3 = 4 \int_{x_3(x_1, x_3) \in \Omega} h(x_1, x_3, x_2) dx_3.\]

Thus,

\[(3.9)\quad h(x_1, x_3, x_2) = 0, \text{ if } (x_1, x_2) \notin \Omega \text{ and } (x_1, x_3) \in \Omega.\]

Using symmetry of \(h\) w.r.t. first two arguments we get \(h(x_1, x_3, x_2) = h(x_3, x_1, x_2)\) and

\[(3.10)\quad h(x_1, x_3, x_2) \equiv 0 \text{ if } (x_2, x_3) \notin \Omega \text{ and } (x_1, x_3) \in \Omega.\]

Finally, calculate \(f_1(1/8)\). On the one hand, \(f_1(1/8) = \int f_2(1/8, y) dy > 0\). On the other hand,

\[(3.11)\quad f_1(x_2) = 4 \int \int_{(x_1, x_3) \in \Omega} h(x_1, x_3, x_2) \chi_{\Omega}(x_1, x_3) dx_3 dx_1.\]
where $\mathcal{O} = \{(x_1, x_3) : h(x_1, x_2, x_3)\chi_{\Omega}(x_1, x_3) \neq 0\}$. The domain $\mathcal{O}$ depends on $x_2$ and for $x_2 = 1/8$

$$\mathcal{O} = \{h(x_1, x_3, 1/8) \neq 0 \text{ and } \chi_{\Omega}(x_1, x_3) \neq 0\} = \text{def of } \chi_{\Omega}$$
$$= \{(x_1, x_3) \in \Omega, h(x_1, x_3, 1/8) \neq 0\} \subset \{(3/8, 5/8), x_3 \in (3/8, 5/8)\} = \emptyset.$$

Integral in (3.11) is taken over empty set. Thus, $f_1(1/8) = 0$. Contradiction. \(\square\)

Instead of using a unique representation ansatz for $f_3$ we use two different but similar representation ansatzes for $f_3$, $f_3 = f_3^{(I)}(x_1, x_2, x_3)$ and $f_3 = f_3^{(II)}(x_1, x_2, x_3)$, for two different places where $f_3$ appears in the equation $k = 2$ such that key properties are preserved. Namely, the equation $k = 2$ is rewritten as follows

$$\partial_t f_2 + \frac{1}{N} \nabla_x \cdot (K(x_2 - x_1)f_2) + \frac{1}{N} \nabla_{x_2}(K(x_1 - x_2)f_2)$$
$$+ \frac{N - 2}{N} \nabla_{x_1} \int K(x_3 - x_1)f_3^{(I)}(x_1, x_2, x_3)dx_3$$
$$+ \frac{N - 2}{N} \nabla_{x_2} \int K(x_3 - x_2)f_3^{(II)}(x_1, x_2, x_3)dx_3 = 0,$$

where

$$f_3^{(I)}(x_1, x_2, x_3) = \begin{cases} f_2(x_1, x_2) f_2(x_1, x_3) & \int f_2(x_1, y)dy > 0, \\ f_2(x_1, y)dy = 0, \end{cases}$$

and

$$f_3^{(II)}(x_1, x_2, x_3) = \begin{cases} f_2(x_1, x_2) f_2(x_2, x_3) & \int f_2(y, x_2)dy > 0, \\ f_2(y, x_2)dy = 0. \end{cases}$$

One can see that four key properties listed in Introduction of $f_1$ and $f_2$ after such truncation preserve.

Remark 3.1. Such a truncation can be explained as follows. The equation for the second marginal from original (not truncated) BBGKY hierarchy can be written as

$$\partial_t f_2 + \nabla_{x_1} \cdot \left\{ \frac{1}{N} \sum_j K(x_j - x_1)|x_1, x_2)f_2 \right\}$$
$$+ \nabla_{x_2} \cdot \left\{ \frac{1}{N} \sum_j K(x_j - x_2)|x_1, x_2)f_2 \right\} = 0.$$

Here conditional expectation $\langle G(x_1, ..., x_N)|x_1, x_2 \rangle$ is calculated by the formula:

$$< G|x_1, x_2 > f_2(x_1, x_2) = \int G(x_1, ..., x_N)f_N(x_1, ..., x_N)dx_3...dx_N.$$

Thus, the truncation is a simplification of the conditional expectation in the $\nabla_{x_1}$-term in the equation for $f_2$:

$$\langle \frac{1}{N} \sum_j K(x_j - x_1)|x_1, x_2 \rangle \approx \frac{\int \frac{1}{N} \sum_j K(x_j - x_1)f_N(x_1, ..., x_N)dx_3...dx_N}{f_1(x_1)}.$$
and the similar simplification of the $\nabla x_2$-term. If one compares the definition of conditional expectation and the ansatz, he will notice that in the $\nabla x_1$-term we suppressed variable $X_2$ and in the $\nabla x_2$-term we suppressed variable $x_1$.

**Remark 3.2.** The truncation presented above can be applied for various original ODE systems (1.1), where each position can have any integer dimension. Note that different truncations of the Boltzmann hierarchy have been made before for some specific situations. For example, we refer to papers [21, 22, 23] devoted to Ostwald ripening where a truncation was motivated by expansions in concentration of particles.

### 4. Numerics.

In this section we provide numerical comparison between marginals $f_1$ and $f_2$ which solve the truncated system (3.12) and histograms built on solutions to system (1.1).

**General description of Numerics.** The numerics are performed for $x \in [0, 1]$ and periodic boundary conditions with period 1. Types of interaction kernels $K(\cdot)$ are described below (see (4.1), (4.2), (4.3), (4.4)). For all the types, $K$ is such that for $x \in [0, 1]$ the function $K(x - \cdot)$ has values separated from zero by the number larger than $10^{-8}$ in $[0, 1]$, and in $[-1, 0] \cup [1, 2]$, i.e., two neighbor periods. By direct simulation of the ODE system (1.1) we understand numerical solution using the Euler method of the system with initial positions $X_i(0)$ distributed by the law $f_0$, given by (4.5). The truncated system (3.12) was solved numerically using finite difference and upwind scheme of the 1st order, the Euler scheme is explicit. Spatial step $dx$ and time step $dt$ are specified below each figure, in caption.

**Type of interactions.** Kernel $K$ has a support in $[-1, 1]$, and in this interval the kernel is given by formulas (4.1), (4.2), (4.3), and (4.4), $S = 36$. The parameter $S$ stands for the size of the support of $K$, i.e., the size of the set $\{x : |K(x)| > 10^{-8}\}$.

1. **Translating kernel** $K_{\text{trans}}$:

(4.1) $K_{\text{trans}}(x) := e^{-Sx^2}, \ x \in [-1, 1].$

$K \geq 0$. The $j$th particle pushes $i$th particle to the right regardless whether $i$th particle is situated to the right from $j$th or otherwise. Note that if one considers motion of one particle ($N = 1$), then the particle will move to the right with speed $K(0)$.

2. **Repelling kernel** $K_{\text{rep}}$:

(4.2) $K_{\text{rep}}(x) := -Sxe^{-Sx^2}, \ x \in [-1, 1].$

Particles repel each other. The distribution will tend to uniform.

3. **Attracting kernel** $K_{\text{attr}}$:

(4.3) $K_{\text{attr}}(x) := Sxe^{-Sx^2}, \ x \in [-1, 1].$

Particles attract each other. The distribution will tend to a $\delta$-function per fundamental period.

4. **Attraction/Repulsion** $K_{\text{attr/rep}}$:

(4.4) $K_{\text{attr/rep}}(x) := x(r^2 - x^2)e^{-Sx^2}, \ x \in [-1, 1].$

Particles form clusters.
**Boundary and initial conditions.** Boundary conditions are [0, 1]-periodic. Initial conditions:

\[(4.5)\]

\[f_1(x_1) = 0.2 \sin 2\pi x_1 + 1, \quad f_2(x_1, x_2) = f_1(x_1) f_1(x_2).\]

We separate two cases \(\alpha = 1\) and \(\alpha = 30\).

**4.1.** \(\alpha = 1\). Numerics show that marginal \(f_1\) has the similar behavior to histogram \(\tilde{f}_1\) defined by

\[\tilde{f}_1 = \frac{1}{N} \# \{X_i(t) \in [jh, (j+1)h]\}, \quad x \in [jh, (j+1)h),\]

where \(h \sim 1/\sqrt{N}\). See figures 4.1, 4.2, 4.3.
4.2. $\alpha = 30$. Numerics are performed for positive kernel $K$. In direct simulations we test $R = 30$ initial realizations, i.e., $X_i^{(r)}$: position of $i$th particle in $r$th realization. Marginal $f_1$ is compared with histogram:

$$\hat{f}_1 = \frac{1}{NR} \sum_{r=1}^{R} \# \left\{ X_i^{(r)}(t) \in [jh, (j+1)h) \right\}, \quad x \in [jh, (j+1)h).$$

See figure 4.4. Marginal $f_2$ is compared with statistical joint distribution summed over

$$B = \{ (x_1, x_2) : 0 < x_1, x_2 < 1/2 \},$$

i.e.,

$$B_{f_2} := \frac{1}{R} \frac{N_{1/2}(N_{1/2} - 1)}{N(N - 1)}, \quad \text{where} \quad N_{1/2} = \# \{ X_i : 0 < X_i < 1/2 \}.$$  

We compare $\int_B f_2$ and $B_{f_2}$. On small time steps $\int_B f_2$ oscillates, while $B_{f_2}$ does not. On larger time steps both quantities coincide. See figure 4.5 and 4.6.
$\text{Fig. 4.5. Positive kernel, } \alpha = 30. \text{ Left: } f_2 \text{ solution to (3.12) integrated over } B; \text{ Right: } B_{f_2}; dt = 1/600. \text{ Horizontal axis stands for time } t$

$\text{Fig. 4.6. Positive kernel, } \alpha = 30. \text{ Left: } f_2 \text{ solution to (3.12) integrated over } B; \text{ Right: } B_{f_2}; dt = 1/60. \text{ Horizontal axis stands for time } t$

### 4.3. Saturation of correlations.

We are interested in how correlations change in time with initially independent distributions. We introduce the function of time $t$:

$$c(t) := \int_0^1 \int_0^1 |f_2(t,x_1,x_2) - f_1(t,x_1)f_1(t,x_1)| dx_1 dx_2.$$ 

The function $c(t)$ first increases for certain time and then oscillates (see Fig. 4.7). We define saturation time $T_s$ as the first local maximum of the function $c(t)$ (see Tabular below; numerics are performed for positive $K$, $dt = 0.05$ and $\alpha = 1$).

| $N$    | $T_s$  | $c(T_s)$ |
|--------|--------|----------|
| 100    | $\approx 70dt$ | 0.0054   |
| 200    | $\approx 80dt$ | 0.0032   |
| 300    | $\approx 90dt$ | 0.0028   |
| 400    | $\approx 95dt$ | 0.0025   |
| 500    | $\approx 100dt$ | 0.0024   |
| 30 000 | $\approx 115dt$ | 0.0023   |

Saturation time $T_s$ increases and correlations decrease as $N$ grows. This corresponds to the mean field limit: on fixed time-interval correlations vanish as $N$ goes to $\infty$. 
5. Conclusions and future work. In this work we proposed a computational approach to study correlations in many particles system. It is efficient in comparison with direct simulations of particles motion. The method is based on a clever truncation of the BBGKY hierarchy for many particles system with interactions. It preserves key properties of marginals (positivity, conservation of mass, the relation between the first and the second marginal, etc.). Numerics show the solution of the truncated system describes the motion of particles appropriately.

One-dimensional model has poor large-time behavior and, up to our best knowledge, does not allow for analysis the dependence of collective state on the interaction strength parameter. In the problem \( \text{RHS} \) depends linearly in \( \alpha \), so \( \alpha \) stands simply for time scales. We intend to study the more realistic dependence on interaction parameter \( \alpha \) in multidimensional model.

REFERENCES

[1] X.-L. Wu and A. Libchaber, “Particle diffusion in a quasi-two-dimensional bacteria bath,” Physical Review Letters, vol. 84, p. 3017, 2000.
[2] C. Dombrowski, L. Cisneros, S. Chatkaew, R. Goldstein, and J. Kessler, “Self-concentration and large-scale coherence in bacterial dynamics,” Physical Review Letters, vol. 93, p. 98103, 2004.
[3] A. Sokolov, I. Aranson, J. Kessler, and R. Goldstein, “Concentration dependence of the collective dynamics of swimming bacteria,” Physical Review Letters, vol. 98, no. 15, p. 158102, 2007.
[4] A. Sokolov and I. Aranson, “Reduction of viscosity in suspension of swimming bacteria,” Phys. Rev. Lett., vol. 103, p. 148101, Sep 2009.
[5] A. Sokolov, M. Apodaca, B. Grzybowski, and I. Aranson, “Swimming bacteria power microscopic gears,” PNAS, vol. 107, pp. 969–974, 2010.
[6] K. Leptos, J. Guasto, J. Collub, A. Pesci, and R. Goldstein, “Dynamics of enhanced tracer diffusion in suspensions of swimming eukaryotic microorganisms,” Physical Review Letters, vol. 103, p. 198103, 2009.
[7] A. Sokolov and I. Aranson, “Physical properties of collective motion in suspensions of bacteria,” Physical Review Letters, vol. 109, p. 248109, 2012.
[8] S. Ryan, A. Sokolov, L. Berlyand, and I. Aranson, “Correlation properties of collective motion in bacterial suspension,” New Journal of Physics, 2013.
[9] L. Greengard and V. Rokhlin, “A fast algorithm for particle simulation,” Journal of Computational Physics, vol. 73, pp. 325–348, 1987.
Appendix A. Wasserstein distances. Wasserstein distance or Monge-Kantorovich-Wasserstein (MKW) quantities the difference between two given measures. Roughly speaking, a measure can be viewed as a pile of sand. The MKW distance between two such piles is an optimal work of transferring one pile into another.

Given two measures \( \mu_1 \) and \( \mu_2 \) in \( M^1(D) \), one may define the set of transference plans between \( \mu_1 \) and \( \mu_2 \) as the set \( \Pi(\mu_1, \mu_2) \) of measures \( \pi \in M^1(D \times D) \) s.t.

\[
\mu_1(x) = \int_D \pi(x, dy), \quad \mu_2(y) = \int_D \pi(dx, y).
\]

The \( p \) MKW distance \( W_p(\mu_1, \mu_2) \) between \( \mu_1 \) and \( \mu_2 \) is given by

\[
W_p(\mu_1, \mu_2) = \inf_{\pi \in \Pi(\mu_1, \mu_2)} \int_{D^2} |x - y|^p \pi(dx, dy).
\]

If \( D \) is the torus, then \( |x - y| \) is replaced by the corresponding distance (in general in a manifold, it would be the geodesic distance).

For measures with bounded moments, the MKW distances metrize the weak-* topology. Moreover, on bounded domain the \( W_1 \) distance is essentially equivalent to the negative Sobolev norm \( W^{-1.1} \). The \( p \)-MKW distances play an important role for particle systems as the \( p \)-MKW distance between two empirical measures is typically comparable with the \( p \) distance between the two vectors of positions, that is

\[
W_p \left( \frac{1}{N} \sum_i \delta_{x_i}, \frac{1}{N} \sum_i \delta_{y_i} \right) \sim \frac{1}{N} \sum_i |x_i - y_i|^p,
\]

up to a permutation of indices on the \( y_i \).

Appendix B. Deterministic is a degenerate case of randomness. Is our
method efficient in the deterministic case?. Consider the following initial $N$-particle distribution:

$$f_N(0, x_1, ..., x_N) = \frac{1}{N!} \sum_{i_1, ..., i_N, i_k \neq i_l} \prod_{k=1}^{N} \delta(x_k - X_{i_k,0}),$$

where $X_{i,0}$ are given numbers. Such initial conditions mean that with probability 1 $N$ particles occupy positions $X_{1,0}, ..., X_{N,0}$. Hence initial conditions are deterministic.

What will happen if one starts solving the truncated system for initial conditions \([B.1]\)? Numerical solution with sufficient accuracy will require very small size of mesh due to singular initial conditions and, what can be claimed \textit{a priori}, singular exact solution. The problem can be reduced to solving the original system \([1.1]\) with initial deterministic positions $X_{1,0}, ..., X_{N,0}$.

**Conclusion:** method of using the truncated system is efficient for smooth (non-singular) initial distributions. Smoothness in this case means that the distribution is essentially random.

**Appendix C. Collective scale and correlation length, dependence on marginals $f_1$ and $f_2$.**

In this subsection we describe the notion of correlation length which was used in experimental and numerical works on active bacterial suspensions. Then we show that the correlation length can be evaluated if marginals $f_1$ and $f_2$ are known.

First, we explain how the collective scale or correlation length is computed in works \([7, 8]\). For each realization the autocorrelation function is defined by

$$C(r) = \frac{\int_{D \times S^{d-1}} V(x + r s) \cdot V(x) \, dx \, ds - \left( \int_D V(x) \, dx \right)^2}{\int_D V^2(x) \, dx - \left( \int_D V(x) \, dx \right)^2}.$$  

Here $r$ is the shift and the velocity field $V(x)$ is the velocity field created by the bacteria in the fluid. In particular, $V(x)$ depends intrinsically on each particles’ coordinates and hence on the realization. $S^{d-1}$ is unit $d$-dimensional sphere. Therefore $C(r)$ should be considered as a random function. For each realization, a correlation length $L_{\text{corr}}$ is extracted by fitting autocorrelation function $C(r)$ to an exponential function of the form $e^{-r/L_{\text{corr}}}$. This correlation length depends as well on the realization. The final correlation length $\bar{L}_{\text{corr}}$ is then obtained by taking the average of all $L_{\text{corr}}$. The fact that autocorrelation function is computed for each realization can be explained by ergodicity assumption.

In this article, we will use a slightly different definition of the correlation length. We take averages already inside the autocorrelation function. Thus we define

$$\tau(r) = \frac{\mathbb{E} \int_{D \times S^{d-1}} V(x + r s) \cdot V(x) \, dx \, ds - \left( \mathbb{E} \int_D V(x) \, dx \right)^2}{\mathbb{E} \int_D V^2(x) \, dx - \left( \mathbb{E} \int_D V(x) \, dx \right)^2}.$$  

In our context, the velocity field $V$ is defined by

$$V(x) = \frac{\alpha}{N} \sum_{j=1}^{N} K(x - X_j(t)).$$
Note that \( V \) depends on the dynamics itself through the \( X_j(t) \) term and it is therefore random as well.

The goal of this subsection is to prove that the autocorrelation function \( c(r) \) depends only on \( f_1 \) and \( f_2 \).

Averages, for convenience, are taken outside mathematical expectations.

\[
(C.2) \quad c(r) = \int \frac{\int V(x + rs)V(x)f_N(x_1, ..., x_N)dx_1...dx_N - |\int V(x)f_N(x_1, ..., x_N)dx_1...dx_N|^2}{\int |V(x)|^2f_N(x_1, ..., x_N)dx_1...dx_N - |\int V(x)f_N(x_1, ..., x_N)dx_1...dx_N|^2} \, ds \, dx
\]

Simplify this definition by using that particles are identical (i.e., \( f_N \) is symmetrical in spatial variables) and definition of \( V(x) \) (C.1) as a random variable:

\[
\int V(x + rs)V(x)f_N(x_1, ..., x_N)dx_1...dx_N = \frac{\alpha^2}{N^2} \sum_{i,j=1}^{N} K(x + rs - x_1)K(x - x_j)f_N(x_1, ..., x_N)dx_1...dx_N
\]

\[
= \frac{\alpha^2(N-1)}{N} \sum_{i,j=1}^{N} K(x + rs - x_1)K(x - x_j)f_N(x_1, ..., x_N)dx_1...dx_N
\]

\[
= \frac{\alpha^2}{N} \sum_{i=1}^{N} K(x - x_i)f_N(x_1, ..., x_N)dx_1...dx_N
\]

\[
(C.3) \quad = \alpha \int K(x - x_1)f_1(x_1)dx_1
\]

\[
\int |V(x)|^2f_N(x_1, ..., x_N)dx_1...dx_N = \alpha^2 \frac{N-1}{N} \sum_{i,j=1}^{N} K(x - x_1)K(x - x_2)f_2(x_1, x_2)dx_1dx_2 + \frac{\alpha^2}{N} \sum_{i,j=1}^{N} K(x - x_1)f_1(x_1)dx_1...dx_N
\]

So the autocorrelation is

\[
c(r) = \int \left[ \frac{1}{(\sigma(x)/\alpha)^2} \sum_{i,j=1}^{N} K(x + rs - x_1)K(x - x_j)f_2(x_1, x_2)dx_1dx_2 + \frac{\alpha^2}{N} \sum_{i,j=1}^{N} K(x - x_1)f_1(x_1)dx_1...dx_N \right] dxds
\]

where \( \sigma > 0 \) is a deviation of both \( V(x) \) and \( V(x + r\omega) \).

In the case of independence \( f_2(x_1, x_2) = f_1(x_1)f_1(x_2) \) function \( c(r) \) is not necessarily zero:

\[
c(r) = \int \frac{\int K(x + rs - x_1)K(x - x_j)f_2(x_1, x_2)dx_1dx_2 - \int K(x - x_1)f_1(x_1)dx_1|^2}{\int |K(x - x_1)|^2f_1(x_1)dx_1 - \int K(x - x_1)f_1(x_1)dx_1|^2} \, dxds
\]

**Conclusion:** correlation length does depend on \( f_2 \) and \( f_1 \) only. The reason for dependence of \( c(r) \) on \( f_1 \) and \( f_2 \) only (and not on \( f_3 \), for example) is that correlation function \( c(r) \) involves at most second power of random variables \( X_1, ..., X_N \) in its definition.

**Appendix D. How key properties of solutions become requirements on functional \( F[f_1, f_2] \).** Consider the original system (for the sake of simplicity we
Fig. E.1. Positive kernel, $\alpha = 30$, $f_1$ histogram with $R = 30$. Left: for $N = 3000$; Right: for $N = 5000$. at time $t = 0$ (Green), $t = 0.33$ (Pink), $t = 0.66$ (Blue); $dt = 1/600$.

Integrate the second equation with respect to $X_2$. We get the equation for $f_1(X_1)$:

\[
\partial_t f_1 + \frac{N-1}{N} \partial_1 \left\{ \int K(X_2 - X_1) f_2(X_1, X_2) dX_2 \right\} = 0,
\]

\[
\partial_t f_2 + \frac{1}{N} \partial_1 (K(X_2 - X_1) f_2) + \frac{1}{N} \partial_2 (K(X_1 - X_2) f_2)
\]

\[
+ \frac{N-2}{N} \partial_1 \left\{ \int K(X_3 - X_1) f_3(X_1, X_2, X_3) dX_3 \right\}
\]

\[
+ \frac{N-2}{N} \partial_2 \left\{ \int K(X_3 - X_2) f_3(X_1, X_2, X_3) dX_3 \right\} = 0,
\]

Integrate the second equation with respect to $X_2$. We get the equation for $f_1(X_1)$:

\[
\partial_t f_1 + \frac{1}{N} \partial_1 \left( \int K(X_2 - X_1) f_2 dX_2 \right) + \frac{N-2}{N} \partial_1 \left\{ \int K(X_3 - X_1) \int f_3(X_1, X_2, X_3) dX_2 dX_3 \right\} = 0,
\]

Now swap letters $X_2$ and $X_3$ in the last term

\[
\text{(D.1)} \quad + \frac{N-2}{N} \partial_1 \left\{ \int K(X_2 - X_1) \int f_3(X_1, X_2, X_3) dX_2 dX_3 \right\}
\]

Thus, in order to have equation for $f_1$ we need the following

\[
\text{(D.2)} \quad f_2(X_1, X_2) = \int f_3(X_1, X_2, X_3) dX_3.
\]

This is the condition $\text{(3.4)}$.

**Appendix E.** Additional numerics. Right figure 4.4 for $N = 3000$ and $N = 5000$:

**Appendix F.** Stability results with respect to number of particles $N$ when $\alpha = 1$ in direct simulations. Numerics show that $N = 1500$ is enough for attracting, repelling and mixing kernel. See figures F.1, F.2 and F.3.

**Appendix G.** Correlations. The definition of correlation $c(t)$ from Section 2 was used in Numerics. See Figures G.1 and G.2.
Fig. F.1. Repulsion kernel, $\alpha = 1$ for $N = 1500, 2500, 3500$. Horizontal axis stands for time $t$, vertical axis stands for particles distribution.
Fig. F.2. Attraction kernel, $\alpha = 1$ for $N = 1500, 2500, 3500$. Horizontal axis stands for time $t$, Vertical axis stands for particles distribution.
Fig. F.3. Attraction kernel, $\alpha = 1$ for $N = 5500, 6500, 10000$. Horizontal axis stands for time $t$, Vertical axis stands for particles distribution.

Fig. G.1. Repulsion kernel, Horizontal axis stands for time $t$, Vertical axis stands for Left: correlations $c(t)$; Right: measure of nonindependency.
Fig. G.2. Attraction kernel, Horizontal axis stands for time $t$, Vertical axis stands for correlations $c(t)$. 