COMPLEXITY REDUCTION IN MANY PARTICLE SYSTEMS WITH RANDOM INITIAL DATA

Leonid Berlyand\textsuperscript{1}, Pierre-Emmanuel Jabin\textsuperscript{1}, and Mykhailo Potomkin\textsuperscript{5}

Abstract. We consider the motion of interacting particles governed by a coupled system of ODE's with random initial conditions. Direct computations for such systems are prohibitively expensive due to a very large number of particles and randomness requiring many realizations in their locations in the presence of strong interactions. While there are several approaches that address the above difficulties, none addresses all three simultaneously. Our goal is to develop such a computational approach in order to capture the experimentally observed emergence of correlations in the collective state (patterns due to strong interactions). Our approach is based on the truncation of the BBGKY hierarchy that allows one to go beyond the classical Mean Field limit and capture correlations while drastically reducing the computational complexity. Finally, we provide an example showing a numerical solution of this nonlinear and non-local system.

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

AMS subject classifications. 35L65, 35L71, 82-08

1. Motivation and Settings. Systems of interacting particles described by a coupled system of a large number of ODE's with random initial conditions appear in many problems of physics, cosmology, chemistry, biology, social science and economics; such as the first order systems we consider here

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

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 \(\mathbb{R}^d\), a \(d\)-dimensional torus \(\Pi^d\), or a compact domain in \(\mathbb{R}^d\) in which case boundary conditions must be added. The scalar function \(K\) describes the inter-particle interactions, and \(S(X_i)\) models an internal force of each particle, such as self-propulsion.

System (1.1) is an Individual Based Model, i.e., it has an ODE for each particle coupled with others. In various applications the role of an individual can be played by atoms, bacteria in suspensions (microswimmers), animals in flocks, social agents etc. The system (1.1) has two key parameters: \(\alpha\), the strength of interactions, and \(N\), the number of particles. The parameter \(\alpha\) is determined by geometry such as interparticle distance and the mass of a particle (note that a model particle is just a point) as well as physics. The magnitude of \(\alpha\) plays an important role: a small \(\alpha\) corresponds to almost decoupled interactions; large \(\alpha\)

\textsuperscript{*}The work of Leonid Berlyand and Mykhailo Potomkin was supported by DOE grant DE-FG-0208ER25862. The work of Pierre-Emmanuel Jabin was partially supported by NSF grant DMS-1312142. LB and MP wish to thank V. Rybalko for his comments and suggestions which helped to improve the manuscript.
\textsuperscript{1}Department of Mathematics, The Pennsylvania State University, University Park, Pennsylvania. ( lvb2@psu.edu)
\textsuperscript{2}Department of Mathematics, University of Maryland, College Park, MD 20742 USA, (pjabin@umd.edu)
\textsuperscript{3}Department of Mathematics, The Pennsylvania State University, University Park, Pennsylvania, (mup20@psu.edu)
corresponds to strong interactions; α ∼ 1 corresponds to the classical Mean Field (MF) regime which describes the limit N → ∞. Our main focus is on constructing the correction to the MF limit for large N that describes fluctuations. Note that the MF limit can be interpreted as a consequence of the Law of Large Numbers, whereas fluctuations would need an equivalent of a Central Limit Theorem.

Our work is motivated by experiments in bacterial suspensions [1, 2, 3, 4, 5, 6, 7]. These experiments [3, 4, 7] show the 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 a bacterial suspension (the definition of the correlation length is given in Appendix B). 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 when more energy is injected into the system (for other studies of collective state in bacterial suspensions see also [9] and references therein).

The motion of bacteria can be modeled by a system of the form (1.1) where the position and orientation of the i-th bacterium are described by the vector $X_i(t)$ (see Appendix B). In this case the parameter α equals $(ℓ/R)^2 NV_0$, where $V_0$ is the swimming speed of a single bacterium, $R$ is the mean distance between two bacteria, and ℓ is the characteristic size of a bacterium. The collective behavior observed in experiments [3, 4, 7] has also been qualitatively reproduced by direct numerical simulations in [8] for systems of $10^5$ bacteria, which validates the model of the type (1.1). However, the computational cost of direct simulations of the ODE system for a realistic number of bacteria is prohibitively high for the following reasons:

(i) the number of bacteria N is very large ($10^{10}$ per cm$^3$);
(ii) to draw a reliable conclusion one needs to consider many realizations, which mathematically translates into random initial data;
(iii) the main interest is in collective state corresponding to large α which leads to small time steps.

The combination of the factors (i)-(iii) makes the computational cost too high even for the most powerful particle methods such as Fast Multipole Method [10, 11, 12] (for FMM the computational cost of each realization is linear in N, whereas it is of the order $N^2$ for direct explicit methods).

The goal of this paper is to propose a computational approach that allows one to describe numerically the collective state of this system with properties (i)-(iii). More specifically, the collective state is described by the two-particle correlation function. The objective of our study is the efficient computation of this quantity.

The main idea is to replace the ODE system (1.1) by a PDE such that the computational cost of its solution does not grow as N goes to infinity. This idea had been used in the classical MF approach which corresponds to α of the order 1 but this approach is not valid for strong interactions.

This paper focuses on a PDE approach that extends beyond the Mean Field, so that it can capture correlations but remains efficient with a computational complexity that increases only slowly with N. We consider the BBGKY hierarchy which consists of N PDEs for k-particle probability distribution functions, $k = 1, ..., N$, and obtain a closed system for 1- and 2-particle probability distribution functions by a clever truncation of the hierarchy. Then the large parameter N is present only in coefficients in a more innocuous way than in the original
problem (1.1), and the resulting system can be handled efficiently with high order methods. This approach computes probability distribution functions and therefore it avoids computing individual realizations. Thus, it allows us to overcome the computational difficulties (i) and (ii). The contribution to the computational complexity from difficulty (iii) is still present but much less of a problem than (i) and (ii), because $\alpha \sim N^\gamma$ and $0 < \gamma < 1$ (typically, $1/3 < \gamma < 1/2$ for many applications).

The idea of the truncation is based on the following assumption: the conditional expectation of the interaction force

$$
E \left\{ K(X_3(t) - X_1(t)) \left| X_1(t) = x_1, X_2(t) = x_2 \right. \right\},
$$

which is in general a function of two variables $x_1$ and $x_2$, does not depend on $x_2$. This assumption holds exactly when particles are independent. On the time interval $[0, T]$ on which initially independent particles are dependent but still close to being independent, numerical simulations show that this assumption provides a good approximation of the two-particle correlation function. Note that the larger number of particles, the larger $T$. As we assume that all particles are identical, $X_1$, $X_2$ and $X_3$ in (1.2) can be replaced by another triple of distinct particles.

Note that a specific feature of our method is that it is efficient for random initial conditions of system (1.1) in contrast to deterministic approaches. 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 the numerical cost of solving such a deterministic problem is very high. Random initial data in ODE (1.1) lead instead to smooth initial conditions in the truncated BBGKY hierarchy that are much easier to handle numerically.

The truncation presented in this paper can be applied for various ODE systems of type (1.1), in dimension 1 or more. Different truncations of the BBGKY hierarchy have been made before for some specific situations, which usually rely on some perturbative arguments. For example, we refer to papers [13, 14, 15] devoted to Ostwald ripening where a truncation was motivated by expansions in (small) concentration of particles.

The paper is organized as follows. The main idea of the truncation is described in Section 2. We recall the Mean Field approach and discuss its limitations in Section 3. The properties of solutions of the truncation of BBGKY hierarchy is established in Section 4. In Section 5 we show that the truncation is the simplest truncation possible, provided that its solution preserves the key properties of the one- and two- particle probability distribution functions. Numerical simulations performed to check that the truncated PDE system is reliable are described in Section 6.

2. Main Idea of Truncation of the BBGKY Hierarchy. We consider the first order dynamics given by (1.1) with random initial conditions. For simplicity, we assume in this Section that $S(x) \equiv 0$ and $K(0) = 0$ (no self-interactions). Our goal is to find the joint probability distribution function $f_2(t, x_1, x_2)$ for positions of any pair of particles $X_i(t)$ and $X_j(t)$, $i \neq j$, and without loss of generality we choose $i = 1$ and $j = 2$.

The following lemma provides an equation for $f_2$. Its proof is relegated to Appendix.

**Lemma 2.1.** The marginal $f_2(t, x_1, x_2)$ solves the following Liouville equation

$$
\partial_t f_2 + \partial_{x_1} (F_1 f_2) + \partial_{x_2} (F_2 f_2) = 0,
$$
where $F_1$ and $F_2$ are given by the conditional expectation:

$$
F_i(x_1, x_2, t) = \mathbb{E} \left\{ \frac{\alpha}{N} \sum_{j \neq i} K(X_j(t) - X_i(t)) \bigg| X_1(t) = x_1, X_2(t) = x_2 \right\}, \ i = 1, 2.
$$

If $F_i$, $i = 1, 2$, are given, then in order to find $f_2$, equation (2.1) must be supplied with initial conditions:

$$
f_2(0, x_1, x_2) = \int f_N(t = 0, x_1, x_2, x_3, ..., x_N) dx_3 .. dx_N,
$$

where $f_N(t = 0, x_1, ..., x_N)$ is the joint probability distribution function of positions of $N$ particles at $t = 0$.

We can rewrite (2.2) by substituting the joint condition $X_i(t) = x_i$, $i = 1, 2$, into the expression for the force $\frac{\alpha}{N} K$:

$$
F_1(x_1, x_2, t) = \mathbb{E} \left\{ \frac{\alpha}{N} \sum_{j = 2}^{N} K(X_j(t) - x_1) \bigg| X_1(t) = x_1, X_2(t) = x_2 \right\}.
$$

Note that the first term in the sum in (2.2) is deterministic, since it is completely determined by the conditions $X_1(t) = x_1$ and $X_2(t) = x_2$. Since particles are identical, the expectation of all terms in the sum for $j \geq 3$ are equal to each other. Thus, we can simplify (2.3) as follows

$$
F_1(x_1, x_2, t) = \frac{\alpha}{N} K(x_2 - x_1) + \frac{\alpha(N - 2)}{N} \mathbb{E} \left\{ K(X_3(t) - x_1) \big| X_1(t) = x_1, X_2(t) = x_2 \right\}.
$$

An analogous formula holds for $F_2$.

It follows from (2.4) that $F_i$ are determined by the joint probability distribution function $f_3$ for triples $(X_1(t), X_2(t), X_3(t))$. Thus, (2.1) is not a closed form equation for $f_2$, since coefficients $F_i$ depend on $f_3$. Analogous to Lemma 1, one can derive the Liouville equation for $f_3(t, x_1, x_2, x_3)$, whose solutions will be determined by the joint probability distribution function $f_4$ of four particles $(X_1(t), X_2(t), X_3(t), X_4(t))$, and so on. This is well known in statistical physics as the BBGKY hierarchy. Thus, for practical computations a truncation of this system is necessary.

The “first order” truncation is obtained by dropping both conditions in (2.3) and substituting these forces $F_i$ into (2.1). Integration of the resulting equation in $x_2$ yields the following closed form equation for $f_1$:

$$
\partial_t f_1 + \frac{\alpha}{N} \frac{N - 1}{N} \partial_x \left( \int K(y - x) f_1(y) dy f_1(x) \right) = 0.
$$

Taking the limit $N \to \infty$ in (2.5) leads to the well known Mean Field (MF) limit: The Vlasov equation.

Since the main objective of our work is the description of collective states that are characterized by the correlation length and two-particle correlation functions, the MF approximation is not sufficient and we have to find the next order truncation (equation for $f_2$).
To this end, rewrite (2.2) in the form

$$F_1 = \frac{\alpha}{N} K(x_2 - x_1) + \mathbb{E} \left\{ \frac{\alpha}{N} \sum_{j=3}^{N} K(X_j(t) - X_1(t)) \mid X_1(t) = x_1, X_2(t) = x_2 \right\}.$$  

and assume that for all \( j \geq 3 \)

$$\mathbb{E} \{ K(X_j(t) - X_1(t)) \mid X_1(t) = x_1, X_2(t) = x_2 \} \text{ does not depend on } x_2,$$

and, analogously,

$$\mathbb{E} \{ K(X_j(t) - X_2(t)) \mid X_1(t) = x_1, X_2(t) = x_2 \} \text{ does not depend on } x_1.$$

To explain the assumption (2.7), we need to understand how the conditions \( X_1(t) = x_1 \) and \( X_2(t) = x_2 \) affect the term \( K(X_j(t) - X_1(t)) \) in (2.6). The expectation in (2.7) can depend on each of the conditions \( X_1(t) = x_1 \) and \( X_2(t) = x_2 \) explicitly and implicitly. Explicit dependence means that either \( X_1(t) \) or \( X_2(t) \) is present in the expression of the term (e.g., \( X_1(t) \) is present in \( K(X_3(t) - X_1(t)) \), but \( X_2(t) \) is not). Thus, the term \( K(X_j(t) - X_1(t)) \) for \( j > 2 \) explicitly depends on the condition \( X_1(t) = x_1 \), but not on \( X_2(t) = x_2 \). Implicit dependence is due to the fact that positions of all particles are dependent (e.g., the position of the second particle \( X_2(t) \) may affect the position of the third particle and, thus, the term \( K(X_3(t) - X_1(t)) \), in which \( X_2(t) \) is not even present). Therefore, for all \( j > 2 \), \( K(X_j(t) - X_1(t)) \) implicitly depends on both \( X_1(t) = x_1 \) and \( X_2(t) = x_2 \). When particles are independent or close to be independent, implicit dependence is zero or negligible, respectively, and only explicit dependence affects \( K(X_j(t) - X_1(t)) \). We want to study the regime when correlations are not zero, but not too large, so that the positions of particles are dependent, but still close to independent state (a correction to MF). Therefore, we are interested in regimes when implicit dependence is less important than explicit. Since \( K(X_j(t) - X_1(t)) \) for \( j > 2 \) depends on \( X_2(t) = x_2 \) only implicitly, we thus assume that it does not depend on \( X_2(t) = x_2 \), so that its expectation does not depend on \( x_2 \) (which is the assumption (2.7)).

Applying the assumption (2.7) to (2.6), we keep the first conditioning \( X_1(t) = x_1 \) and drop the second one \( X_2(t) = x_2 \) in the second term of (2.6) to obtain the approximation

$$F_1 \approx \frac{\alpha}{N} K(x_2 - x_1) + \mathbb{E} \left\{ \frac{\alpha}{N} \sum_{j \neq 1, 2} K(X_j(t) - X_1(t)) \mid X_1(t) = x_1 \right\}.$$

Similarly, for \( i = 2 \) we obtain

$$F_2 \approx \frac{\alpha}{N} K(x_1 - x_2) + \mathbb{E} \left\{ \frac{\alpha}{N} \sum_{j \neq 1, 2} K(X_j(t) - X_2(t)) \mid X_2(t) = x_2 \right\}.$$

Substitute (2.9) and (2.10) into (2.1). Then we get a closed form PDE for \( f_2(t, x_1, x_2) \):

$$\partial_t f_2 + \frac{\alpha}{N} \partial_{x_1} (K(x_2 - x_1) f_2) + \frac{\alpha}{N} \partial_{x_2} (K(x_1 - x_2) f_2)$$

$$+ \frac{\alpha}{N} \sum_{j \neq 1, 2} \left( \mathbb{E} \left\{ \frac{\alpha}{N} K(X_3(t) - X_1(t)) \mid X_1(t) = x_1 \right\} f_2 \right)$$

$$+ \frac{\alpha}{N} \sum_{j \neq 1, 2} \left( \mathbb{E} \left\{ \frac{\alpha}{N} K(X_3(t) - X_2(t)) \mid X_2(t) = x_2 \right\} f_2 \right) = 0,$$

(2.11)
where the conditional expectations presented in the equation (2.12) are determined by $f_2$:

$$
(2.12) \quad \mathbb{E} \left\{ \frac{\alpha}{N} K(X_3(t) - X_1(t)) \mid X_1(t) = x_1 \right\} = \frac{\alpha}{\int f_2(t, x_1, y) dy} \int K(y - x_1) f_2(t, x_1, y) dy.
$$

3. Random initial conditions, correlations, the Mean Field approach. For physical reasons, 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 to derive the Mean Field limit. The first one is based on the so-called empirical measure. The second one is a statistical approach which is better suited for our purpose.

The Mean Field limit is valid as long as the correlations between particles are negligible. 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 corresponding to the 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,

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

On the other hand, we believe that the numerical implementation of this approach will work well even for singular kernels (see Remark 3.1).

3.1. Preliminaries. How to choose initial conditions: Randomness and marginals. By assumption (3.1) 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 the position of each particle with enough precision.

Instead, what is accessible is some statistical information about the positions of the particles. 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 $(X_1^0, ..., X_N^0) \in Q$ is given by

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

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

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

$$
(3.2) \quad f_k(t, x_1, ..., x_k) = \int_{D^{N-k}} f_N(t, x_1, ..., x_k, x_{k+1}, ..., x_N) \, dx_{k+1}...dx_N.
$$
Some of marginals have a 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 \( f_2 \).

In the simplest case, one assumes that the particles are initially independently and identically distributed, that is
\[
(3.3) \quad f_N(t = 0, x_1, ..., x_N) = \prod_{i=1}^{N} f^0(x_i).
\]
This independence is strongly connected to the usual Mean Field limit approach as explained in subsection 2.2 (see (3.15)).

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

The correlation of particles’ positions is classically defined by
\[
(3.4) \quad c = \frac{\mathbb{E}[X_1 \cdot X_2] - (\mathbb{E}[X])^2}{\mathbb{E}[X^2] - (\mathbb{E}[X])^2} = \frac{\int x_1 \cdot x_2 f_2(x_1, x_2) \, dx_1 \, dx_2 - \left( \int x f_1(x) \, dx \right)^2}{\int x^2 f_1(x) \, dx - \left( \int x f_1(x) \, dx \right)^2}.
\]

The numerator in (3.4) can be written as follows:
\[
\int x_1 x_2 (f_2(x_1, x_2) - f_1(x_1) f_1(x_2)) \, dx_1 \, dx_2.
\]
Observe that \( c \) vanishes if
\[
f_2(x, y) = f_1(x) f_1(y),
\]
that is, if the particles positions are independent. However, the particles positions can be dependent even if \( c = 0 \), for example: \( f_2(x_1, x_2) = 2\chi_{\{|x_1|+|x_2|<1/2\}} \). Thus, we will use instead the following definition of \( c \):
\[
(3.5) \quad c = \int |f_2(x_1, x_2) - f_1(x_1) f_1(x_2)| \, dx_1 \, dx_2,
\]
which vanishes if and only if the positions are independent.

### 3.2. The Mean Field approach.

**Empirical measure.** Assume that the \( X_i(t) \) are solutions to (1.1), and define the empirical measure
\[
(3.6) \quad \mu_N(t, x) := \frac{1}{N} \sum_{i=1}^{N} \delta(x - X_i(t)).
\]
Note that if the particles are indistinguishable then there is just as much information in the empirical measure as in the position 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 distributions

$$
\frac{\partial f(t,x)}{\partial t} + \nabla_x \cdot (S(x)f(t,x)) + \alpha \nabla_x \cdot \left( \int K(y-x)f(t,y)\,dy \right) = 0.
$$

Consider a sequence of initial positions $X_N = \{X_i(0) : i = 1, \ldots, N\}$ 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 probability 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 (3.7) 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. These are classically the so-called Monge-Kantorovich-Wasserstein (MKW) distances. For our purpose it is enough to understand that they correspond to some appropriate distance between probability measures.

Now we give the main stability estimate behind the Mean Field limit. From [16], [17], and [18], it is possible to prove that if $f$ and $g$ are two measure-valued solutions to (3.7), then

$$
W_p(f(t,\cdot), g(t,\cdot)) \leq e^{t\alpha \|
abla K\|_{L^\infty}} W_p(f(0,\cdot), g(0,\cdot))
$$

where $W_p(\cdot, \cdot)$ is a $p$-Wasserstein or MKW distance between two measures. The inequality (3.8) is a Gronwall-type inequality. Note also that the inequality (3.8) 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 (3.3), then a large deviation for the law of large numbers applies ($E\mu_N = f^0$) 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 [19, Appendix A, Proposition 1.2] (see also [20]), that if $f^0$ is a non-negative 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 of $D$ and $R$

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

This says that with a probability exponentially close to 1, $\mu_N(t = 0)$ and $f^0$ are polynomially close in $N$. Denote by $f(x,t)$ the solution to (3.7) with $f^0(x)$ as an initial data. By combining the deterministic stability (3.8) with a law of large numbers in the form of (3.9) we obtain that

$$
P \left( W_1(\mu_N(t, \cdot), f(t, \cdot)) \leq \frac{CR}{N^{\gamma_1}} e^{t\alpha \|
abla K\|_{L^\infty}} \right) \geq 1 - e^{-CN^{\gamma_2}}.
$$

Why random initial conditions make computations much easier in the Mean Field framework? Looking for a solution of the Vlasov equation (3.7) in the 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 (3.7) is comparatively much faster and obviously independent of $N$ (provided the solution of (3.7) is independent of $N$). Since the initial distribution $f^0$ is usually assumed to be smooth, the corresponding solution $f$ is smooth as well. Computing $f$ numerically is thus far easier than solving (1.1), because computational cost does not depend on $N$.

The key to the reduction in the computational complexity in this Mean Field 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 (3.10) implies that this $f$ will be a good approximation of the original $N$ up to a time $t$ of order

\[
\log N \leq \frac{1}{\alpha \|K\|_{L^\infty}}.
\]

Note that in certain circumstances, this time can be considerably extended to become polynomial in $N$. This usually requires a stable equilibrium to equation (3.7), see [21] for instance.

### 3.3. The BBGKY Hierarchy

Here we introduce the hierarchy of equations on the marginals $f_k$, $k = 1, ..., N$.

First, it is well-known that $f_N$ solves the Liouville equation (see, e.g., [18, 22]):

\[
\partial_t f_N + \sum_{i=1}^N \partial_{x_i} (S(x_i)f_N) + \frac{\alpha}{N} \sum_{i=1}^N \partial_{x_i} \left( \sum_{j=1}^N K(x_j - x_i)f_N \right) = 0.
\]

Next, by integrating the equation for $f_N$, one obtains an equation satisfied by each marginal $f_k$

\[
\partial_t f_k + \sum_{i=1}^k \partial_{x_i} (S(x_i)f_k) + \frac{\alpha}{N} \sum_{i=1}^k \sum_{j=1}^k \partial_{x_i} (K(x_j - x_i)f_k)
\]

\[
+ \frac{\alpha(N-k)}{N} \sum_{i=1}^k \partial_{x_i} \left( \int K(y - x_i)f_{k+1}(t,x_1,\ldots,x_k,y) \, dy \right) = 0.
\]

The system of $N$ PDE’s consisting of (3.12) and (3.13) for $k = 1, ..., N - 1$ is the BBGKY hierarchy for marginals. Its main features are

- The equation (3.12) (or, equivalently, (3.13) for $k = N$) is a closed PDE with respect to unknown $f_N$. Thus, theoretically, one can find $f_N$ by solving (3.12) only, and then all other $f_k$’s are found through (3.2).
- On the other hand, $f_N$ is a function of $N + 1$ variables which makes the numerical solution of (3.12) impossible (unless the number of particles $N$ is small, $N \leq 3$). Moreover, if one needs, e.g., $f_2$ only, then finding the whole $f_N$ is not reasonable.
- The BBGKY hierarchy is a coupled system of $N$ PDE’s: the PDE (3.13) for the marginal $f_k$, $k = 1, ..., N - 1$, requires the higher order marginal $f_{k+1}$. Therefore it is not a priori possible to solve independently any given equation from the hierarchy without solving the whole hierarchy. However, if $f_{k+1}$ is known, then it is straightforward to solve (3.13) and it is less computationally expensive than (3.12).
**Propagation of chaos.** The PDE for $f_1$ is

$$
\partial_t f_1(t, x_1) + \partial_{x_1} \left( S(x_1) f_1(t, x_1) \right) + \frac{\alpha K(0)}{N} \partial_{x_1} f_1(t, x_1)
+ \alpha \frac{N - 1}{N} \partial_{x_1} \left\{ \int K(y - x_1) f_2(t, x_1, y) dy \right\} = 0.
$$

(3.14)

By taking the formal limit $N \to \infty$ in the equation (3.14) 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 (3.7).

This leads us to the crucial concept of *propagation of chaos* (see [18, 23] and more recent papers [24, 25] and the review [22]). Under some mild conditions on the smoothness of $K$, for initial positions that are close to being independent (that is (3.3) is assumed) as $N \to \infty$ we have

$$
f_k(t, x_1, ..., x_k) \to \Pi_{i=1}^k f(t, x_i),
$$

(3.15)

where $f(t, x)$ solves the Mean Field equation (3.7).

Note that for a finite $N$, one cannot have equality in (3.15) and, in particular, $\Pi_{i=1}^N f(t, x_i)$ cannot be a solution to (3.12). 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 (3.11).

**Beyond Mean Field: the truncation of the BBGKY hierarchy.** The Mean Field limit leads to a closed equation on $f_1$ but it only offers a rough estimate of $f_2$. In particular, it cannot give any estimate on correlations since it relies on the premises that they are vanishing. In our context, however, this means that we cannot use the Mean Field framework to evaluate correlations as defined by (3.4) which are small but non 0 either, in line with the experimentally observed phenomenon that we wish to explain.

In general the exact computation of those correlations would require one to exactly solve equation on $f_2$. As it was pointed out above, this equation would in turn require to solve the equation on $f_3$ and so on.

Any exact solution would require solving the full equation (3.12) on $f_N$. Unfortunately, $f_N$ is a function of $N + 1$ variables and the computational cost of the numerical solution of (3.12) is much too large to be even remotely realistic.

We would like instead to compute directly the marginals up to $f_k$ for some $k$, approximately if it is not possible to do it exactly. This leads us to the key question of possible **truncations for the BBGKY hierarchy**. A truncation at level $k$ is an ansatz which expresses the terms involving $f_{k+1}$ in terms of $f_k$ and lower order marginals. Using this ansatz makes the first $k$ equations of the hierarchy closed thus letting us solve them.

In that sense, the Mean Field limit can be seen as a particular case of truncation at order $k = 1$. In this paper, we focus and propose a possible truncation at order $k = 2$. We are able to show through numerical experiments that it is valid as long as correlations are not too large.

**Remark 3.1.** (On singular kernels) As mentioned before, the rigorous justification of the classical Mean Field theory requires some smoothness on the interaction kernel, $K(x)$ is Lip
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 [26], [27], [28] or [29] 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).

4. Truncation for the BBGKY Hierarchy at order $k = 2$. In this section we discuss the proposed truncation for the BBGKY hierarchy at order $k = 2$ introduced in Section 2. First, we formulate the truncation in terms of an ansatz which expresses $f_3$ in terms of $f_1$ and $f_2$ and we show that the solutions of the truncated system preserve key properties of one- and two- particle distribution functions. Second, in Proposition 4.1 we describe the general form of ansatz based on the assumption (2.7) formulated in Section 2 and the key properties of one- and two- particle probability distribution functions.

In Lemma 2.1 it was shown that $f_2$ which solves the full (not truncated) BBGKY hierarchy satisfies the Liouville equation (2.1) which is another form of (3.13) for $k = 2$. Indeed, the forces $F_i$ can be rewritten in terms of marginals as follows:

\begin{equation}
F_i = \frac{\alpha}{N} K(x_{3-i} - x_i) + \frac{\alpha N - 2}{N} \int K(x_3 - x_i) \frac{f_3(t, x_1, x_2, x_3)}{f_2(t, x_1, x_2)} dx_3.
\end{equation}

If one substitutes (4.1) into (2.1), he gets the equation (3.13) for $k = 2$.

Equation (2.11) which was written in Section 2 applying approximations (2.9) and (2.10) can be obtained by introducing two different, but symmetric, truncation ansatz for $f_3$ in the equation (3.13) for $k = 2$:

\begin{equation}
\begin{aligned}
f_3^{(I)}(t, x, x_2, x_3) &= \frac{f_2(t, x, x_2)f_2(t, x_1, x_3)}{\int f_2(t, x, y)dy} \\
\text{and } f_3^{(II)}(t, x, x_2, x_3) &= \frac{f_2(t, x_1, x_2)f_2(t, x_3, x_2)}{\int f_2(t, y, x_2)dy}.
\end{aligned}
\end{equation}

The two ansatzes for $f_3$ are used in two different places where $f_3$ appears in the equation (3.13) for $k = 2$:

\begin{equation}
\begin{aligned}
\partial_t f_2 + \frac{\alpha}{N} \partial_{x_1} (K(x_2 - x_1)f_2) + \frac{\alpha}{N} \partial_{x_2} (K(x_1 - x_2)f_2) \\
+ \frac{\alpha N - 2}{N} \partial_{x_1} \left\{ \int K(x_3 - x_1) f_3^{(I)}(t, x_1, x_2, x_3) dx_3 \right\} \\
+ \frac{\alpha N - 2}{N} \partial_{x_2} \left\{ \int K(x_3 - x_2) f_3^{(II)}(t, x_1, x_2, x_3) dx_3 \right\} &= 0.
\end{aligned}
\end{equation}

We list below four key properties of one- and two- particle distribution functions which are preserved by solutions (4.3), and for each property we give arguments why it holds:

1. **Symmetry of $f_2$:** $f_2(t, x_1, x_2) = f_2(t, x_2, x_1)$ (particles are identical).

   This property follows from symmetry of the equation (4.3) with respect to $x_1$ and $x_2$. 

...
2. **Mass preserving and positivity:** \( \int f_2 \, dx_1 \, dx_2 = \) const, \( f_1, f_2 \geq 0 \) provided that initial conditions for \( f_1 \) and \( f_2 \) are positive.

This property follows from the fact that equation (4.3) can be rewritten as a standard conservation law.

3. **Consistency of \( f_1 \) and \( f_2 \):** \( f_1 = \int f_2 \).

Integrating (4.3) in \( x_2 \), one obtains an equation for \( \int f_2(t, x_1, x_2) \, dx_2 \) which coincides with the equation for \( f_1 \). Hence, this property holds.

4. **Propagation of chaos:** if \( f_1 \) and \( f_2 \) solve (3.14) and (4.3) for \( N = \infty \), and \( f_2(0, x_1, x_2) = f_1(0, x_1)f_1(0, x_2) \), then \( f_2(t, x_1, x_2) = f_1(t, x_1)f_1(t, x_2) \) for all \( t > 0 \).

This follows from the fact that if \( f_2(x_1, x_2) = f_1(x_1)f_1(x_2) \), then \( f_3^{(1)} = f_3^{(2)} \).

The main assumption (2.7) implies that the second term in (4.1) (which is the expectation of \( K(X_i(t) - X_i(t)) \) with conditions \( X_1(t) = x_1 \) and \( X_2(t) = x_2 \) does not depend on \( x_2 \) for \( i = 1 \) and does not depend on \( x_1 \) for \( i = 2 \). Does this assumption imply (2.9)-(2.10) (or, equivalently, (4.2))? To make this question rigorous, consider \( \tilde{f}_1 \) and \( \tilde{f}_2 \), an approximation for \( f_1 \) and \( f_2 \), respectively, and equations (3.13) for \( k = 1 \) and \( k = 2 \) where we replaced the second terms in (4.1) in the equation for \( k = 2 \) by functions which do not depend on \( x_2 \) and \( x_1 \), respectively (\( G_1(t, x_1) \) and \( G_2(t, x_2) \)). In other words, we consider the following system for \( \tilde{f}_1 \) and \( \tilde{f}_2 \):

\[
\begin{align*}
(4.4) \quad & \partial_t \tilde{f}_1 + \frac{N-1}{N} \partial_{x_1} \left( \int K(x_2 - x_1) \tilde{f}_2 \, dx_2 \right) = 0, \\
(4.5) \quad & \partial_t \tilde{f}_2 + \partial_{x_1} \left( \left( \frac{\alpha}{N} K(x_2 - x_1) + G_1(t, x_1) \right) \tilde{f}_2 \right) + \partial_{x_2} \left( \left( \frac{\alpha}{N} K(x_1 - x_2) + G_2(t, x_2) \right) \tilde{f}_2 \right) = 0.
\end{align*}
\]

We want \( \tilde{f}_1 \) and \( \tilde{f}_2 \) to preserve basic properties of \( f_1 \) and \( f_2 \) such as symmetry of \( f_2 \) with respect to \( x_1 \) and \( x_2 \) and \( f_1 = \int f_2 \). The following proposition implies the specific structure for \( G_i \), provided that the properties are preserved.

**Proposition 4.1.** Assume that \( \tilde{f}_1(t, x_1), \tilde{f}_2(t, x_1, x_2) \) solve equations (4.4)-(4.5). Then

\( \tilde{f}_2(t, x_1, x_2) = \tilde{f}_2(t, x_2, x_1) \) and \( \tilde{f}_1(t, x_1) = \int \tilde{f}_2(t, x_1, x_2) \, dx_2 \)

if and only if

\[
(4.6) \quad G_1(t, x) = G_2(t, x) = \alpha \frac{N-2}{N} \int K(x_3 - x) \frac{\tilde{f}_2(t, x, x_3)}{f_1(t, x)} \, dx_3 + \frac{\beta}{f_1(t, x)},
\]

where \( \beta \) is independent of \( x_1, x_2 \), but may depend on \( t, \tilde{f}_1, \tilde{f}_2 \) and \( N \).

**Proof.** Here we will show (4.6). Integrate equation (4.5) in \( x_2 \):

\[
(4.7) \quad \partial_t \tilde{f}_1 + \frac{\alpha}{N} \partial_{x_1} \left( \int K(x_2 - x_1) \tilde{f}_2 \, dx_2 \right) + \partial_{x_1} \left\{ G_1(t, x_1) \tilde{f}_1 \right\} = 0.
\]

Subtracting (4.4) from (4.7) we obtain

\[
(4.8) \quad \partial_{x_1} \left( G_1(t, x_1) \tilde{f}_1 \right) = \alpha \frac{N-2}{N} \partial_{x_1} \left( \int K(x_3 - x_1) \frac{\tilde{f}_2(t, x_1, x_3)}{f_1(t, x_1)} \, dx_3 \tilde{f}_1 \right).
\]
Integration of (4.8) in $x_1$ implies (4.6).

Remark 4.2. The property of mass preserving and positivity holds for the system (4.4)-(4.5) for arbitrary $G_i$'s. The property of propagation of chaos holds, iff $\beta \to 0$ as $N \to \infty$.

If $\beta \equiv 0$ and $\tilde{f}_1$ and $\tilde{f}_2$ are true one- and two- particle distribution functions (the marginals solving the full BBGKY hierarchy), then using (2.12) we obtain

$$G_i(t, x_i) = \mathbb{E}\{K(X_3(t) - X_i(t))| X_i(t) = x_i\}.$$  

Thus, by imposing $\beta \equiv 0$ in (4.6) we get equation (2.11). On the other hand, presence of non-zero $\beta$ does not contradict to the key properties and, maybe, it contains a 'higher order' correction to (2.9)-(2.10) which depends on specific structure of $K$.

5. Why not simpler. To truncate the BBGKY hierarchy at level $k = 2$ we introduced two different truncation ansatzes (4.2) for $f_3$ in two different places where $f_3$ appears in the equation (3.13) for $k = 2$. In this section we discuss the possibility of a single truncation ansatz of the form $f_3 = \mathcal{F}[f_1, f_2]$. It changes the equation for $f_2$ as follows:

$$\partial_t f_2 + \frac{\alpha}{N} \partial_{x_1} (K(x_2 - x_1)f_2) + \frac{\alpha}{N} \partial_{x_2} (K(x_1 - x_2)f_2)$$

$$+ \frac{N-2}{N} \partial_{x_1} \left\{ \int K(x_3 - x_1) \mathcal{F}[f_1, f_2](t, x_1, x_2, x_3) \, dx_3 \right\}$$

$$+ \frac{N-2}{N} \partial_{x_2} \left\{ \int K(x_3 - x_2) \mathcal{F}[f_1, f_2](t, x_1, x_2, x_3) \, dx_3 \right\} = 0.$$  

(5.1)

We want the solution of system (3.14)-(5.1) to satisfy the key properties (see Section 4): symmetry of $f_2$, mass preserving, positivity and consistency of $f_1$ and $f_2$. We reformulate these properties as requirements on the function $\mathcal{F}$ and then prove that such an ansatz does not exist.

Consider a representation for $f_3$:

$$f_3(x_1, x_2, x_3) = \mathcal{F}[f_1, f_2](x_1, x_2, x_3),$$

where $\mathcal{F}$ is a function (in general, a nonlinear operator) of $f_1$ and $f_2$. We reformulate the key properties as requirements on $f_3$ calculated by (5.2) for given $f_1$ and $f_2$.

First, the symmetry of $f_2$ with respect to arguments $x_1$ and $x_2$ is equivalent to:

$$f_2(x_1, x_2) = f_2(x_2, x_1) \text{ for all } x_1, x_2$$

$$\Rightarrow f_3(x_1, x_2, x_3) = f_3(x_2, x_1, x_3) \text{ for all } x_1, x_2, x_3.$$  

(5.3)

Next, in order to preserve positivity of $f_1$ and $f_2$, we need to impose

$$f_2(x_1, x_2) \geq 0 \text{ for all } x_1, x_2$$

$$\Rightarrow f_3(x_1, x_2, x_3) \geq 0 \text{ for all } x_1, x_2, x_3.$$  

(5.5)
The proof is by contradiction. The idea is to combine requirements (5.7) which holds due to (5.8). The equality (5.9) holds if (5.10) provided that the initial data is positive.

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

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

The equality (5.6) is equivalent to the statement that if we integrate the equation for \( k = 2 \) from the BBGKY hierarchy with respect to one of the spatial variables, say, \( x_2 \), we get the equation for \( k = 1 \).

**Proposition 5.1.** There is no such representation (5.2) that all requirements (5.3),(5.4),(5.5) and (5.6) hold true.

**Proof.** The proof is by contradiction. The idea is to combine requirements (5.4) and (5.5):

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

and to find such \( f_2 \) that the LHS of (5.7) is zero, but the RHS is not zero.

Assume that (5.3),(5.4),(5.5) and (5.6) hold true. Take

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

and \( f_2(x_1, x_2) = \frac{1}{12} \chi_\Omega(x_1, x_2) = 4 \chi_\Omega(x_1, x_2) \). Here \( \chi_\Omega \) is a characteristic function of domain \( \Omega \). Note \( f_1(x) > 0 \) for all \( x \in (0, 1) \setminus \left\{ \frac{1}{4}, \frac{3}{4} \right\} \) because of the equality \( f_1(x) = \int f_2(x, y) \, dy \) which holds due to (5.6).

The property (5.4) implies the existence of such a 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 (5.6) we obtain

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

Let \( (x_1, x_2) \not\in \Omega \), then (5.8) implies that

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

Thus

\[
(5.10) \quad h(x_1, x_3, x_2) = 0, \text{ if } (x_1, x_2) \not\in \Omega \text{ and } (x_1, x_3) \in \Omega.
\]
By using the symmetry of \( h \) with respect to first two arguments we get \( h(x_1, x_3, x_2) = h(x_3, x_1, x_2) \) and

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

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,

\[
(5.12) \quad f_1(x_2) = 4 \int \int_{(x_1, x_3) \in O} h(x_1, x_3, x_2) \chi_\Omega(x_1, x_3) \, dx_3 \, dx_1.
\]

where \( O = \{ (x_1, x_3) : h(x_1, x_2, x_3) \chi_\Omega(x_1, x_3) \neq 0 \} \). The domain \( O \) depends on \( x_2 \). We claim that \( O \) is empty for \( x_2 = 1/8 \). Indeed,

\[
O = \{ h(x_1, x_3, 1/8) \neq 0 \text{ and } \chi_\Omega(x_1, x_3) \neq 0 \} = \text{[definition of } \chi_\Omega]\]

\[
\subset \{ (x_1, x_3) \in \Omega, h(x_1, 1/8) \neq 0, (x_1, x_3, 1/8) \in \Omega \}
\subset \{ (x_1, x_3) \in \Omega, x_1 \in (3/8, 5/8), x_3 \in (3/8, 5/8) \} = \emptyset.
\]

Therefore, integral in (5.12) is taken over empty set. Thus, \( f_1(1/8) = 0 \) and we reached a contradiction. ■

6. Numerical example. The goal of this section is (i) to test the truncation (4.3) on a simple 1D example and (ii) to describe one way of handling nonlocality and nonlinearity in the numerical resolution. Here we use numerical methods which are explicit, allows for comparison with direct simulations, and are not necessarily the most efficient. The development of more advanced numerical methods capturing, e.g., 2D, non-smooth kernels or large times, are left for a subsequent work.

To test the truncation we compare probability distribution functions (marginals) \( f_1 \) and \( f_2 \) obtained by numerical solution of the truncated system (4.3) with histograms of particles satisfying the original ODE system (1.1). The histograms are built on many realizations of initial particle positions.
Next we compare the numerical complexity of the marginal approach vs. direct simulations.

In direct simulations the size of the time step $dt$ has to be chosen less than $1/N$ (average inter-particle distance); we denote the number of realizations by $R$. Every time step for each realization has a complexity of $N^2$ (which can be improved to $O(N \log N)$ or even $O(N)$, if one uses more advanced method such as the Fast Multipole Method). Thus, the overall computational complexity of the simple direct simulations is $O(N^3 R)$.

In the marginal approach, we use a second order scheme. In order to capture terms of order $1/N$, we need both time step $dt$ and spatial step $dx$ of order smaller than $1/\sqrt{N}$. Thus, the overall complexity of the marginal approach is $1/(dt \, dx^2) = N^{3/2}$. Therefore, as $N$ grows, the numerical complexity of the direct approach becomes much larger than the complexity of the marginal approach.

Specific setting. Numerical simulations are performed for the one-dimensional problem, $x \in \mathbb{R}$, and periodic boundary conditions with period 1. The interaction kernel $K$ is periodic with period 1 and for $|X_j - X_i| < 1/2$ it is given by $K(X_j - X_i) = e^{-12(X_j - X_i)^2}$. Initially particles are independent:

$$f_2(0, x_1, x_2) = f_1(0, x_1)f_1(0, x_2), \quad \text{where } f_1(0, x_1) = 0.4 \sin 2\pi x_1 + 1.$$ 

Number of particles is $N = 100$ per one periodic cell $x \in [0, 1]$, $\alpha = 3$.

Description of numerical methods. In order to solve the PDE (4.3) we face difficulties that come from the fact that the equation is a non-local non-linear 2D conservation law. For a detailed discussion of difficulties in numerical solution of non-linear conservation laws and the way to resolve them we refer to [30]. In this example we want to simulate accurately terms of order $1/N$, since they are the source of correlations. In other words, if one erases these terms in (4.3), then the solution of equation (4.3) with initial conditions (6.1) will be of the form $f_2(t, x_1, x_2) = f_1(t, x_1)f_1(t, x_2)$, i.e., with no correlations. This motivates us to use a second order scheme, for example, the one with flux limiters for which we have converging numerical solutions with reasonable spatial and time steps. This scheme is described below.

The PDE (4.3) can be rewritten as follows

$$\partial_t f_2 + \partial_{x_1}(A_1 f_2) + \partial_{x_2}(A_2 f_2) = 0,$$

where for $f_2(x_1, x_2) > 0$ functions $A_1$ and $A_2$ are given by

$$A_k(t, x_1, x_2) = \frac{\alpha}{N} \sum_{i=1,2} K(x_i - x_k) + \frac{\alpha(N - 2)}{N} \int K(y - x_k) f_2(t, x_k, y) \, dy.$$ 

Denote by $f_{i,j}^m$ the approximation for $f_2(t, x_1, x_2)$ with $t = mdt$, $x_1 = idx$, $x_2 = jdx$, where $dt$ and $dx$ are time and spatial steps, respectively. For given $m$, $i$ and $j$ introduce the following finite difference approximations for $\partial_k \{A_k f_2\}$, $k = 1, 2$: 

$$r_{11} := A_{i,j}^m f_{i,j}^m - A_{i-1,j}^m f_{i-1,j}^m, \quad r_{12} := A_{i+1,j}^m f_{i+1,j}^m - A_{i,j}^m f_{i,j}^m, \quad r_{21} := A_{i,j}^m f_{i,j}^m - A_{i-1,j}^m f_{i-1,j}^m, \quad r_{22} := A_{i,j}^m f_{i,j}^m - A_{i,j+1}^m f_{i,j+1}^m.$$
Introduce also an auxiliary function (flux limiter) \( \phi(r) = \max [0, 0.5 \min(r, 1.5)] \).

The following finite difference scheme is used in the numerical solution of PDE (4.3):

\[
\begin{aligned}
  f^{n+1}_{i,j} &= f^n_{i,j} - \frac{dt}{dx} \left[ \sum_{k=1,2} \left\{ r_{k1} + \phi \left( \frac{r_{k2}}{r_{k1}} \right) (r_{k2} - r_{k1}) \right\} \right].
\end{aligned}
\]

In order to compute \( f_2 \) for \( t > 0 \) directly from the system of ODE’s (1.1) we consider \( R = 5 \cdot 10^5 \) realizations of \( N = 100 \) particles initially identically distributed with probability distribution function \( f(x) = 0.4 \sin 2\pi x + 1 \). Denote by \( X_i^{(r)}(t) \) the position of the \( i \)-th particle, \( i = 1, \ldots, N \) in the \( r \)-th realization, \( r = 1, \ldots, R \) at time \( t \). For each \( r = 1, \ldots, R \) the positions \( \{ X_i^{(r)}(t) \}_{i=1}^N \), \( t > 0 \), are found by solving the ODE system (1.1) using the explicit Euler method of the first order with the time step \( \Delta t = 0.01 \).

We compute the following histogram which converges as \( R \to \infty \) to the probability that the first particle is in the interval \( \Delta_j = [jh, (j+1)h) \) at time \( t \):

\[
\hat{f}_1(t, \Delta_j) = \frac{1}{Rh} \# \left\{ X_1^{(r)}(t) \in \Delta_j, \ r = 1, \ldots, R \right\}.
\]

Here \( h = 0.05 \) is the size of a histogram bin.

Histogram \( \hat{f}_2 \) can be computed as follows

\[
\hat{f}_2(t, \Delta_i, \Delta_j) = \frac{1}{R h^2} \# \left\{ (X_1^{(r)}(t), X_2^{(r)}(t)) \in \Delta_i \times \Delta_j, \ r = 1, \ldots, R \right\}.
\]

We compare \( f_1 = \int f_2 \, dx \) and \( f_2 \) calculated by (6.3) with histograms \( \hat{f}_1 \) and \( \hat{f}_2 \) calculated by (6.4) and (6.5).

Simulations were performed on a machine with 3.06 Ghz Intel core CPU 8 GB of RAM. Numerical solution of (4.3) for \( t = 1 \) for \( dx = 0.0025 \) and \( dx/dt = 200 \) takes approximately 32 hours. Numerical solution of (1.1) on \( R = 10^5 \) realization, \( t = 1 \) and time step \( \Delta t = 1/50 \) takes approximately 83 hours. Besides the long time of computations direct simulations face another difficulty which is the large amount of data. This creates technical difficulties in data movement, its analysis and visualization. Also note that the cost of direct simulations would increase much faster with \( N \) than the cost of our approach.

Results of numerical simulations. Plots in Fig. 2 show that marginal \( f_1 \) is close to histogram \( \hat{f}_1 \).

In order to visualize comparisons between \( f_2 \) and \( \hat{f}_2 \) we plot these functions integrated over domain \( B = \{(x_1, x_2) : 0 \leq x_1, x_2 \leq 1/2 \} \):

\[
Q_{\text{marg}} := \int_B f_2 \, dx_1 \, dx_2, \quad Q_{\text{hist}} := \int_B \hat{f}_2 \, dx_1 \, dx_2 = \frac{h^2}{2} \sum_{i,j \leq 1/(2h)} \hat{f}_2(t, \Delta_i, \Delta_j).
\]

Plot in Fig. 3 (left) shows that quantities \( Q_{\text{marg}} \) defined by marginal \( f_2 \) and \( Q_{\text{hist}} \) defined by histogram \( \hat{f}_2 \) seem to be in very good agreement.
Figure 2. Left: Marginal $f_1$ with $dx = 0.0025$ and $dt = dx/200$; Right: Histogram $\tilde{f}_1$ for $h = 0.05$ and $dt = 0.001$.

Figure 3. Left: Comparison between $Q_{\text{marg}}$ and $Q_{\text{hist}}$; Right: propagation of chaos -- correlation vanishes as $N \to \infty$, clearly observed on the time interval $[0, 1.5]$.

Notice that the quarter cube $B$ was chosen arbitrarily (and in particular there is no conservation of mass on $B$, unlike conservation of mass for the entire cell $[0, 1]^2$). The agreement between $Q_{\text{marg}}$ and $Q_{\text{hist}}$ suggests that the integrals of $f_2$ and $\tilde{f}_2$ over any subdomain of the cell $[0, 1]^2$ would similarly be close. The apparent periodicity in time is due to the choice of periodic boundary conditions.

We compute correlations defined by (3.5). For the marginal approach (i.e., solution of (4.3) and $f_1 = \int f_2 \, dx_2$) correlations are computed as follows

\begin{equation}
(6.6) 
\quad c(t) := \int \int |f_2(t, x_1, x_2) - f_1(t, x_1)f_1(t, x_2)| \, dx_1 \, dx_2.
\end{equation}

Numerical simulations show that as $N \to \infty$ the correlation $c(t)$ computed by (6.6) converges to zero uniformly on the bounded time interval, see Fig. 3 (right). Thus, we have the property of propagation of chaos in a wide sense: if initially $f_2$ is ‘independent’ ($f_2(0, x_1, x_2) =$
many particle systems with randomness 19

Figure 4. Comparisons of correlations computed by marginal approach (4.3) and direct simulations of the system (1.1)

\[ f_1(0, x_1)f_1(0, x_2), \text{ then the solution } f_2 \text{ of (4.3) is not only 'independent' } (f_2(t, x_1, x_2) = f_1(t, x_1)f_1(t, x_2)) \text{ for all } t > 0 \text{ and } N = \infty, \text{ but is also close to be 'independent' for } 1 \ll N < \infty. \]

In direct simulations correlations are computed in a similar way to the formula (6.6) with histograms in place of probability distribution functions:

\[
\tilde{c}(t) = \int \int \left| \tilde{f}_2(t, x_1, x_2) - \tilde{f}_1(t, x_1)\tilde{f}_1(t, x_2) \right| dx_1 dx_2 = h^2 \sum_{i,j} \left| \tilde{f}_2(t, \Delta_i, \Delta_j) - \tilde{f}_1(t, \Delta_i)\tilde{f}_1(t, \Delta_j) \right|.
\]

As it is seen on Fig. 4, plots for correlations computed on marginals and in direct simulations for \( R = 5 \times 10^5 \) have similar qualitative behavior and order of magnitude. The value of correlations is a small number and thus its computation requires high accuracy to reduce the error to an order less than that of the correlations. In direct simulations, this requires a large number of realizations which makes the computations unreasonably long, in contrast to the marginal approach where the computation of correlations is much faster.

Note that correlations observed in this numerical example are not large (in comparison with the maximal possible value of correlations \( c_{\text{max}} = 2 \)). In order to observe large correlations (e.g., \( \sim 0.1 \)) we need to solve (4.3) for large times which is very costly. Moreover, it is delicate to predict the time when correlations will reach a given value. This question is left for subsequent works. Nevertheless, relatively small correlations for times of order 1 may be enough for the solution of the original BBGKY hierarchy to be essentially different from the one obtained by the Mean Field approach. In that case our approach with (4.3) would still capture the correct solution in contrast to Mean Field.

Finally, in this example we numerically justify our ansatz (2.7) (or, equivalently, (4.2)) by direct simulations of the system (1.1). To this end, for given \( i \) and \( j \) we introduce the set
$g_{ij}$ at $t = 3$

![Figure 5](image.png)

**Figure 5.** Left: Plot of $g_{ij}$ at $t = 3$ illustrates that $g_{ij}$ is independent from $x_2$; Right: the average of $g_{ij}$ with respect to $x_1$ – independent of $x_2$

\[
\mathcal{R}_{ij}^t := \left\{ r : X_1^{(r)}(t) \in \Delta_i, X_2^{(r)}(t) \in \Delta_j \right\}
\]

and compute the following quantity:

\[
g_{ij} = \frac{1}{\#\mathcal{R}_{ij}^t} \sum_{r \in \mathcal{R}_{ij}^t} K(X_3^{(r)}(t) - X_1^{(r)}(t)), \tag{6.7}
\]

which approximates the expectation:

\[
\mathbb{E} \left\{ K(X_3(t) - X_1(t)) | X_1^{(r)}(t) \in \Delta_i, X_2^{(r)}(t) \in \Delta_j \right\}. \tag{6.8}
\]

Numerical simulations for $N = 100$, $R = 5 \times 10^4$ and $t \in [0, 60]$ show that $g_{ij}$ does not depend on $j$ for $t < 10$, see Fig. 5. This implies that our main assumption (2.7) surely holds for $t < 10$. After $t = 10$, $g_{ij}$ slowly transforms from vertical shape (i.e., $g_{ij}$ is independent of $j$) and eventually becomes diagonal, see Fig. 6, left. In order to estimate accuracy of of (2.7) we introduce the quantity

\[
\mathcal{G} = h^2 \sum_{i,j} |g_{ij} - h \sum_i g_{ij}|
\]

For $t < 15$, $\mathcal{G}$ is small and oscillates, and after $t > 15$ it linearly grows, see Fig. 6. This gives another estimate to the time interval where the ansatz (4.2) is applicable: this interval belongs to $0 < t < 15$; beyond this interval the ansatz (4.2) is inaccurate.

**Convergence of numerical methods.** Here we show the convergence of numerical methods used in this section.

First, consider the calculations of marginals $f_1$ and $f_2$. Comparisons of numerical simulations for various spatial and time steps for $t = 1$, $t = 2$ and $t = 3$ are presented on Figures 7 and 8. Convergence of the numerical method in computing $\int_B f_2 dx$ and correlations $c(t)$ is observed on plots in Figure 9.

Next, we consider histograms $\tilde{f}_1$ and $\tilde{f}_2$. Plots on Figures 10 and 11 illustrate convergence of the method for histogram $\tilde{f}_1$ at times $t = 1$, $t = 2$ and $t = 3$, and for histogram $\tilde{f}_2$ summed
Many Particle Systems with Randomness

Figure 6. Left: Plot of $g_{ij}$ at $t = 60$ illustrates that $g_{ij}$ tends to have diagonal form for large times; Right: plot of $\bar{g}$ for $0 < t < 60$ shows that $\bar{g}$ grows.

Figure 7. Plots of marginal $f_1$ at $t = 1$ and Right: $t = 2$.

over the set $\mathcal{B}$. Several time steps $\Delta t$ are considered: $\Delta t = 0.02$, $\Delta t = 0.01$, $\Delta t = 0.001$. The number of realizations, $R = 10^5$, is chosen for the width of bin $h = 0.02$. It is seen on Figures 10 and 11 that such a number of realizations $R$ seems to be enough to have converging numerical solutions for $\tilde{f}_1$ and $\int_{\mathcal{B}} \tilde{f}_2$. To compute correlations $\tilde{c}(t)$ more realizations would be needed and plots in Figure 12 show that to estimate $\tilde{c}(t)$ we need more than $R = 5 \cdot 10^5$ realizations with $\Delta t = 0.002$. 

Numerical simulations presented above show that PDE system (4.3) not only preserves the qualitative properties of the probability distribution functions (like positivity, consistency, propagation of chaos, etc.), but also may serve for the study of saturation of correlations in such many particle systems. Correlations play an important role, e.g., in the description of collective motion (see, e.g., [7], where transition from individual to collective state is described via correlations).
7. Conclusions. We developed a numerical approach which allows for study correlations in the evolution of many particle systems with random initial conditions. This approach is implemented in a simple 1D settings (toy model). The complexity of solving PDE (4.3) only slowly grows as \( N \) goes to infinity. In other words, the dependence of the complexity on \( N \) in our approach is more 'innocuous', in sharp contrast with direct simulations when the complexity drastically increases as \( N \) grows.

We believe that this approach can be successfully applied to problems in biology, physics and economics.

REFERENCES


Many Particle Systems with Randomness

Figure 10. Left: plots of histogram $\tilde{f}_1$ at $t = 1$; Right: plots of histogram $\tilde{f}_1$ at $t = 2$

Figure 11. Left: plots of histogram $\tilde{f}_1$ at $t = 3$; Right: plots of histogram $\tilde{f}_2$ summed over set $B$


Correlations computed on histograms

Figure 12. Correlations computed on histograms


Acknowledgments. The work of Leonid Berlyand and Mykhailo Potomkin was supported by DOE grant DE-FG02-08ER25862. The work of Pierre-Emmanuel Jabin was partially supported by NSF grant DMS-1312142. LB and MP wish to thank V. Rybalko for his comments and suggestions which helped to improve the manuscript.

Appendix A. Proof of Lemma 2.1.

We need to derive equation (2.1) for the first order dynamics equation (1.1). It can be interpreted as a system in which velocities are determined by forces defined by RHS of (1.1). Then the equation (2.1) can be viewed as the continuity equation: \( \partial_t p + \nabla \cdot (vp) = 0 \) with \( v_i = F_i \).

Because of presence of randomness in the system (we recall that initial conditions are random) we first need to define the probability space \((\Omega, \mathcal{F}, \mathbb{P})\). An elementary outcome of the sample space \(\Omega\) is a trajectory of \(N\) particles:

\[(X_1(t), \ldots, X_N(t))_{t \geq 0}.\]

Since randomness is present in initial conditions only, the corresponding \(\sigma\)-algebra \(\mathcal{F}\) is generated by sets

\[A = \{(X_1(t), \ldots, X_N(t))_{t \geq 0} : (X_1(0), \ldots, X_N(0)) \in B(\mathbb{R}^N)\},\]

where \(B(\mathbb{R}^N)\) are Borel sets in \(\mathbb{R}^N\). The corresponding probability measure \(\mathbb{P}\) depends only on probability distribution function \(f^0_N(x_1, \ldots, x_N)\) for initial conditions \((X_1(0, \omega), \ldots, X_N(0, \omega))\):

\[\mathbb{P}(A) = f^0_N(A|t=0).\]

The joint probability distribution function \(f_2(t, x_1, x_2)\) for positions of any pair of particles \(X_i(t)\) and \(X_j(t)\) at time \(t\) satisfies the following equality

\[\mathbb{E}\{\varphi(X_1(t), X_2(t))\} = \int \varphi(x_1, x_2)f_2(t, x_1, x_2)dx_1dx_2\]

for all integrable functions \(\varphi(x_1, x_2)\).

Take time derivative of both sides of (A.4):

\[\partial_t \mathbb{E}\{\varphi(X_1(t), X_2(t))\} = \int \varphi(x_1, x_2)\partial_t f_2(t, x_1, x_2)dx_1dx_2.\]

On the other hand, due to linearity of expectation we obtain

\[\partial_t \mathbb{E}\{\varphi(X_1(t), X_2(t))\} = \mathbb{E}\{\partial_t \varphi(X_1(t), X_2(t))\} = \mathbb{E}\{\nabla_{x_1} \varphi \hat{X}_1(t) + \nabla_{x_2} \varphi \hat{X}_2(t)\}.\]

Using the property of conditional expectation \(\mathbb{E}\{X \mathbb{E}\{Y|Z\}\} = \mathbb{E}\{EY\}\) which holds for all random variables \(X, Y\) and \(Z\), we get

\[\partial_t \mathbb{E}\{\varphi(X_1(t), X_2(t))\} = \mathbb{E}\{\nabla_{x_1} \varphi F_1 + \nabla_{x_2} \varphi F_2\},\]
By definition of $f_2$, (A.4), we can rewrite (A.7) as follows

$$(A.8) \quad \partial_t \mathbb{E}\{\varphi(X_1(t), X_2(t))\} = \int (\nabla_{x_1} \varphi F_1 + \nabla_{x_2} \varphi F_2) f_2(t, x_1, x_2) \, dx_1 \, dx_2.$$ 

From (A.5) and (A.8) we derive the weak formulation of the equation (2.1) and, thus, the proof of Lemma 2.1 is complete.

**Appendix B. Motivating example: Study of correlations in bacterial suspensions.** In the introduction we mentioned that our work is motivated by experiments in bacterial suspensions and the emergence of a coarse collective scale described by the correlation length. In this section we describe the individual based model from [8] of the bacterial suspension which is proven to partially capture this phenomenon. We show that the correlation length is determined by one- and two- particle probability distribution functions. Thus, the approach presented in this paper may serve (i) as a more efficient method (than direct simulations) to compute the collective scale and (ii) as a first step of the analysis of collective (correlated) behavior of this kind.

Interacting bacteria are modeled by:

$$\begin{align*}
\dot{x}^i &= -d^i + \alpha \sum_{j \neq i} u(x^j - x^i, d^j) + \alpha \sum_{j \neq i} L_{LJ}(x^j - x^i), \\
\dot{d}^i &= \alpha d^i \times \left[ \sum_{j \neq i} (\omega(x^j - x^i, d^j) + B d^i \times \sum_{j \neq i} E(x^j - x^i, d^j)d^i) \right],
\end{align*}$$

where $x^i$ and $d^i$ ($|d^i| = 1$) are positions and orientations of bacteria, $u(x, d)$ is a vector field describing hydrodynamic interactions defined as the solution of the Stokes problem

$$\Delta u - \nabla p = \nabla \cdot \Sigma, \quad \nabla \cdot u = 0 \text{ in } \mathbb{R}^3, \quad |u| \to 0 \text{ as } |x| \to \infty,$$

where $\Sigma = (\ddot{d}d^T - I/3)\delta(x)$; $L_{LJ}$ is the standard Lennard-Jones interaction; $\omega = \nabla \times u$, $2E = \nabla u + (\nabla u)^T$ are the vorticity and the rate of strain due to interactions, and parameter $\alpha$ depends on the strength of interactions. After proper rescaling, the equations (B.1)-(B.2) can be rewritten in the form similar to (1.1):

$$\dot{X}_i = S(X_i) + \frac{\alpha}{N} \sum_{j \neq i} K(X_i, X_j).$$

Note that the method of truncation proposed in this paper does not rely on a specific structure of interactions $K$, and the choice of translation invariant $K (K(x_1, x_2) = K(x_1 - x_2))$ is made for simplicity of presentation.

Next, we define the correlation length for the system (B.3) with $S(x) \equiv 0$, for the sake of simplicity. This definition is analogous to the one from [8]. For given $r > 0$, define the auto-correlation function $C(r)$ by

$$C(r) = \int_{D \times S} \frac{\mathbb{E}\{V(x + r s) \cdot V(x)\} - (\mathbb{E}\{V(x)\})^2}{\mathbb{E}\{V^2(x)\} - (\mathbb{E}\{V(x)\})^2} \, dx \, ds.$$
Here $D = [0, 1]$ is the domain where particles are located; $s \in S$, where $S$ is unit $d$-dimensional sphere, the velocity field $V(x)$ is the velocity field created by all particles $X_j(t), j = 1, ..., N$ in point $x$:

$$V(x) = \frac{\alpha}{N} \sum_{j=1}^{N} K(X_j(t), x).$$

The correlation length $L_{\text{corr}}$ is extracted by fitting the auto-correlation $C(r)$ to an exponential function of the form $e^{-r/L_{\text{corr}}}$.

Finally, we show that the auto-correlation $C(r)$ (and, thus, the correlation length $L_{\text{corr}}$) is determined by $f_1$ and $f_2$. Indeed, since particles are identical we obtain from (B.5):

$$E\{V(x + rs) \cdot V(x)\} = \frac{\alpha^2}{N^2} \sum_{i,j=1}^{N} \int K(x_i, x + rs)K(x_j, x)f_N(x_1, ..., x_N)dx_1...dx_N$$
$$= \frac{\alpha^2(N-1)}{N} \int K(x_1, x + rs)K(x_2, x)f_2(x_1, x_2)dx_1 dx_2$$
$$+ \frac{\alpha^2}{N} \int K(x_1, x + rs)K(x_1, x)f_1(x_1) dx_1,$$

$$E\{V(x)\} = \alpha \int K(x_1, x)f_1(x_1) dx_1,$$

$$E\{V^2(x)\} = \frac{\alpha^2 N - 1}{N} \int K(x_1, x)K(x_2, x)f_2(x_1, x_2) dx_1 dx_2$$
$$+ \frac{\alpha^2}{N} \int |K(x_1, x)|^2 f_1(x_1) dx_1.$$

Substituting the above formula into (B.4) we arrive to the conclusion that $C(r)$ is determined by $f_1$ and $f_2$. 