A review of the mean field limits for Vlasov equations.

Pierre-Emmanuel Jabin *

Abstract

We review some classical and more recent results on the mean field limit and propagation of chaos for systems of many particles, leading to Vlasov or macroscopic equations.

Contents

*CSCAMM and Dpt of Mathematics, University of Maryland, College Park, MD 20742, USA. P.E. Jabin is partially supported by NSF Grant 1312142 and by NSF Grant RNMS (Ki-Net) 1107444.

1 Introduction

The focus of this article is on deterministic second order systems leading to the kinetic Vlasov equation. In addition of presenting the results, we attempt to show some proofs when possible. Due to the complexity of the question, those sometimes had to be simplified leading to results less optimal than the original one but hopefully keeping the main ideas.

The existing lecture notes on the subject have been very useful, in particular the classical book by Spohn [148], the more recent notes by Golse [73] and the seminar by Hauray [85].

1.1 The ODE system and the mean field scaling

Consider N indistiguishable point particles, and denote by $X_i \in \Omega$ and $V_i \in \mathbb{R}^d$ the position and momentum of particle number i. The space domain

Ω may be the whole $\mathbb{R}^d$, or the torus Π^d. The case of a bounded, smooth domain is trongly dependent on the boundary conditions but can sometimes be handled in a similar manner with some adjustments.

In the classical case of Newton dynamics, X_i and P_i satisfy

$$\begin{aligned} \dot{X}_i &= v(P_i), \\ \dot{P}_i &= E_N(X_i) = \alpha \sum_{j\neq i} F(X_i - X_j). \end{aligned} \tag{1.1}$$

In the simplest case, the velocity is equal to the momentum $v(P) = P$ but most of the results reviewed here are also valid in the more general case, including for instance special relativity, $v(P) = P/\sqrt{c^2+P^2}$. We only need to assume that $v(.) \in W^{1,\infty}(\mathbb{R}^d)$.

The most classical example of interaction kernel is the Poisson kernel, $F = C\,x/|x|^d$ in $\mathbb{R}^d$. This corresponds to particles under gravitational interaction for $C < 0$ or electrostatic interactions (ions in a plasma) for $C > 0$. Other Examples of interaction kernels are discussed in subsection 1.2.

The coefficient α includes the physical parameters of the particles. In the most classical example of gravitational or electrostatic interactions, then $\alpha = G/m$ (gravitational) or $\alpha = \varepsilon\, q^2/m$ (electrostatic) with q the charge of one particle and m its mass.

The system (1.1) is supplemented with initial conditions, chosen at $t = 0$ for simplicity

$$\begin{aligned} X_i(t=0) &= X_i^0, \quad i = 1\dots N, \\ P_i(t=0) &= P_i^0, \quad i = 1\dots N. \end{aligned} \tag{1.2}$$

The mean field scaling consists in assuming that $\alpha \sim 1/N$, that is in considering

$$\begin{aligned} \dot{X}_i &= v(P_i), \\ \dot{P}_i &= E_N(X_i) = \frac{1}{N} \sum_{j\neq i} F(X_i - X_j). \end{aligned} \tag{1.3}$$

At least in the case of classical mechanics with $v(P) = P$, it is possible to rescale (1.1) in position and time and therefore in velocity or momentum. By choosing the scalings appropriately, it thus seems to be possible to reduce (1.1) to (1.3).

However the rescaling changes the initial conditions in (1.2). Therefore the rescaling in position and time should instead be chosen so that the initial positions and velocities are of order 1.

In the specific case where F is homogeneous, $F(\lambda x) = \lambda^\alpha F(x)$, one obtains (1.1) with a coefficient α which incorporates both the physical parameters of each particles and the initial scales of the positions and velocities but which has no reason to be of order $1/N$. In this respect the mean field scaling, and subsequently the mean field limits, are only a particular situation. Still in this homogeneous setting, it can be argued that it is the first (or simplest) interesting scaling in the system: Formally, assuming F to be of order 1, the interaction term $\alpha \sum_{j\neq i} F(X_i - X_j)$ is of order $\alpha\, N$.

- If $\alpha << 1/N$, the acceleration term in the second equation of (1.1) is small and one expects that the momentum will mostly not change in time, leading to a not very fascinating regime of free transport as $N \to \infty$.

- If $\alpha >> 1/N$, the acceleration term is very large and one expects some sort of singular behavior; For example the momenta could become very large, or the particles could be distributed along precise patterns to create cancellations in $\alpha \sum_{j\neq i} F(X_i - X_j)$. The analysis is likely complex and heavily dependent on the structure of the interaction.

- Only if $\alpha \sim 1/N$, in the mean field scaling, should the acceleration term precisely be of order 1.

In the general case F is not homogeneous and may for instance have fast decay at infinity (due to a background charge in electrostatic interactions for instance). The discussion is even more complex as two dimensionless parameters are now needed and one finds after rescaling

$$\begin{aligned} &\dot{X}_i = v(P_i), \\ &\dot{P}_i = E_N(X_i) = \alpha \sum_{j\neq i} F(\beta\,(X_i - X_j)). \end{aligned} \tag{1.4}$$

There are now several interesting scalings other than the mean field. The most famous example is the Boltzmann-Grad limit as introduced in [77], which consists in taking $N\,\beta^{d-1} \sim 1$ and $\alpha\,\beta \sim 1$. For a force kernel with fast decay (integrable at infinity) β characterizes the range of the interaction, so $N\,\beta^{d-1}$ corresponds to the total cross-section.

The Boltzmann-Grad limit and the derivation of the Boltzmann equation are at least as important a physical question as the mean field limit. But we will not review in details the results deriving the Boltzmann equation here. The limit was obtained for hard-spheres and short times in [114] with some gaps in the proofs. The proofs were partially completed in [98], [40], (see also [28] and [146]). A full solution was however only given recently in the seminal [64], still only for a short time but with possibly more complex interactions of the type (1.4).

The derivation the Boltzmann equation is in many respects quite different from the mean field approaches; the limiting equation is not time reversible for instance. However many of the tools that are used for the mean field were initially developed in the Boltzmann-Grad setting.

We will also review the mean field limit for first order systems

$$\dot{X}_i = E_N(X_i) = \alpha \sum_{j\neq i} F(X_i - X_j). \tag{1.5}$$

with the corresponding initial data

$$X_i(t=0) = X_i^0, \qquad i = 1 \dots N. \tag{1.6}$$

Through a rescaling in time, the system (1.5) becomes

$$\dot{X}_i = E_N(X_i) = \frac{1}{N} \sum_{j\neq i} F(X_i - X_j), \tag{1.7}$$

which is the now familiar mean field scaling. In fact if F is homogeneous, the mean field scaling is now the only natural one as it always possible to scale (1.5) in (1.7) even if one also needs to rescale the initial data (1.6).

However in the general case of a non homogeneous F, if rescaling in space is necessary, then one obtains as before the more complicated

$$\dot{X}_i = E_N(X_i) = \alpha \sum_{j\neq i} F((X_i - X_j)/\beta). \tag{1.8}$$

Just as for the second order models, various interesting scalings leading to many different limits have been investigated. We only mention here as an example the case where $N^{-1/d} << \beta << 1$, $\alpha = \beta^{-d-1}$ and $F = -\nabla V$ is a short distance, repulsive potential: For instance, $V \geq 0$, $\hat{V} \geq 0$, and

$\int(1+|x|^2)\,V\,dx < \infty$. In that case one formally expects to derive the porous medium equation

$$\partial_t u = C\,\Delta u^2.$$

We refer to [52] for the introduction of this method and to [118] for a first step in the analysis of the convergence.

There are many other interesting mean field limits or related questions that we will not consider here. For instance

- Stochastic or Langevin models. For second order systems, propagation of chaos was shown in [123] for Lipschitz kernels. The propagation of chaos for the stochastic vortex system with independent noises was first proved in the eighties [132], and recently generalized in [63]. We also refer to [23].

- Quantum mechanics and the derivation of non linear Schrödinger equations (in particular Schrödinger-Poisson and Gross-Pitaevskii) from linear, N-particles Schrödinger; see for example [9], [8], [61], [62] and the references therein.

- Mean Field games: Deterministic or stochastic systems coupled through the optimization of an averaged utility; see for instance [81], [115].

- Swarming or consensus models with "auto-rescaling". Instead of (1.5), one considers for example

$$\dot{X}_i = \alpha \frac{\sum_j (X_i - X_j)\,\phi(X_i - X_j)}{\sum_j \phi(X_i - X_j)}.$$

  This is a time continuous variant, derived in [128], of the so-called Krause model, [111], [89] (see also the review [158]). The interaction is now automatically rescaled by the number of particles within range, so $\dot{X}_i$ is always of order 1.

- The study of systems of particles over time scales which are longer than the validity of the mean field limit. Note that those time scales are yet unclear: The best current mathematical results predict that the mean field limit for (1.3) should remain valid for times of order $\log N$. But this is conjectured to be very suboptimal, see [32, 33, 34] for an extension to polynomial times in a simple setting. The approach in physics revolves around tracking the fluctuations around the limit and what is often called as the Lenard-Balescu equation, introduced in [117] and [6]. We refer for example to [42], [112] for more recent advances on this problem which is still not very well understood mathematically.

- Corrections to the mean field limit, in particular for high field or large concentrations, *c.f.* the discussion about $\alpha >> 1/N$ after (1.4). This has been in particular developed for so-called Ostwald ripening, for example in [92, 93]. See also [15] in a numerical context for (1.5).

- Non linear transport equations with long range interactions. Those are typically models with two scales combining a long range interaction over distances of order 1 with strong repulsive effects at very short distance. A good example is the so-called Hughes model (see [56] or [150]).

From now on, we strictly focus on Systems (1.3) and (1.7). Note that the discussion would remain valid for many extensions of those models: Adding a simple velocity dependence in the momentum equation for instance (friction) or considering multi-species models (electrons and ions for example).

1.2 Some examples of interaction kernels

There are many, many examples of interaction kernels in the literature, and the purpose of this subsection is only to give a few classical ones together with some typical order of magnitude for the number of particles N.

• *The Poisson kernel.* This is the oldest kernel dating back to Newton's theory of gravitation.

It is still widely in cosmology and astrophysics to study the formation and evolution of galaxies, galaxy clusters since relativistic effects can often be neglected at large scales. Each particle in this context is a star (or even some larger structure). The number of particles in such physical systems depends much on the case under consideration, from 10^{10} to $10^{20} - 10^{25}$; some models of dark matter even predict up to 10^{60} particles. We refer to [1] for example.

In the repulsive case, the Poisson kernel corresponds to electrostatic interactions between particles. It is commonly used in plasma physics (often with several species or component), see for instance [18]. The number of particles is usually around $10^{20} - 10^{25}$.

The Poisson kernel is also used in first order models, for example in the context of chemotaxis (movement of bacteria or cells induced by a chemical potential) and introduced in [105], [135]. In that case the force field E_N can be interpreted as the gradient of the concentration $c(t,x)$ of a chemical produced by each particle. Neglecting the size of the particles, c should satisfy a simple diffusion equation

$$\tau \partial_t c - \Delta c = \sum_{i=1}^{N} \alpha \, \delta_{X_i}. \tag{1.9}$$

If the diffusion is fast, $\tau << 1$, then (1.9) can indeed be reduced to the Poisson equation. Note in addition that most of the techniques developed for the mean field limit could also be applied if the particles' dynamics was coupled through (1.9).

In general in applications to the Bio-sciences, the number of particles N is lower, typically between $10^8 - 10^9$ and at most 10^{15}.

The Poisson kernel is unbounded, not smooth, anti-symmetric $F(-x) = -F(x)$. It is a critical case in many respects as $F \notin BV$ but ∇F is bounded on L^p as a convolution kernel.

- *Point vortices.* This consists in taking $F = C\, x^\perp / |x|^2$ in dimension 2. The first order model (1.7) then corresponds to the dynamics of point vortices for the $2d$ incompressible Euler equations. In the strict framework of (1.7), all points would have the same vorticity and instead in that case the system is usually generalized to

$$\dot{X}_i = E_N(X_i) = \frac{1}{N} \sum_{j \neq i} \omega_i \, F(X_i - X_j), \tag{1.10}$$

with the ω_i fixed coefficients.

Because of its importance, both for numerical and theoretical purposes in statistical physics, this case has been extensively studied on its own, [76], [96], [75], [97], [144], [145] and [63]. For numerics, up to $10^9 - 10^{10}$ particles can typically be used.

The singularity of the kernel is obviously the same as in the Poisson case.

- *Polynomial potentials.* Take $F = -\nabla V$ with $V(x) = A\,|x|^a - B\,|x|^b$. The potential has an attractive part $-B\,|x|^b$ and a repulsive one $A\,|x|^a$. This is a common choice for many life science applications, in particular swarming and flocking: The collective motion of animals (birds, fishes) or other living organisms, see for instance [47], [48], [153]. The interaction should be repulsive at short range because individuals try to avoid collisions. But it is attractive at long range in order to keep the flock together.

Note that the actual interaction between individuals is probably extremely complex and unknown: Hence this simple choice of F, still capturing the main features.

The number N of individuals can range up to 10^{12} for bacteria or cells but can be much lower for animals, often too low for mean field limits to really apply.

The regularity of the kernel of course depends on the choices a and b.

• *"Pointy" kernels.* The interaction kernel is a smooth function of $|x|$, $F = \tilde{F}(|x|)$, or the gradient of one, $F = -\nabla V$ with $V = \tilde{V}(|x|)$. This is somewhat comparable to the previous example and often used in similar situations.

Because of the dependence on $|x|$ (instead of $|x|^2$ for instance), those force kernels are not necessarily smooth. In fact, unless $F(0) = 0$, $\tilde{F}(|x|)$ is at most Lipschitz even if $\tilde{F} \in C^\infty$; hence the name "pointy".

• *Particles in a fluid.* Each particle influences the others by modifying a fluid surrounding them which in turn affects all the particles. This leads to a whole range of models, varying in the complexity of the description of the interaction or the fluid dynamics: Navier-Stokes, Stokes or Euler, incompressible or sometimes compressible...

While in general the interaction is too complex to be exactly represented by a system like (1.3) or (1.7), it can sometimes be well approximated by such a reduced model for a large number N of particles.

For example, consider an incompressible Stokes flow with N rigid spheres. The complete model involves solving the Stokes system out of the volume occupied by the spheres, with a no-slip boundary condition on each sphere. The solution to the fluid system gives the force applied on each sphere by integrating the stress tensor on the surface of the sphere.

The interaction is in general extremely non linear and complex. However with the right scaling as N becomes large, the force acting on particle i is in fact approximated by, in dimension 3

$$-\lambda P_i + \frac{\mu}{N} \sum_{j \neq i} \left(\frac{Id}{|X_i - X_j|} - \frac{(X_i - X_j) \otimes (X_i - X_j)}{|X_i - X_j|^3} \right) .P_j,$$

fitting with the description of (1.3). At those larger scales, the interaction between particles has thus just become a sum of Stokeslet. The kernel F now depends on P_j and is again non-smooth with a singularity like $1/|x|$. We refer to [100] for the formal derivation, to [55] for a rigorous justification provided the particles are well distributed, and to [99] for a proof that the dynamics itself keeps the particles well distributed if they were so initially.

There are many applications from sedimentation to aerosols and the number N typically ranges from 10^{10} to 10^{15} or even 10^{20} for the smallest particles.

A similar question concerns self propelled particles in a fluid. It allows to consider micro-organisms like bacteria who can "swim" in the fluid. Though the modeling approach is roughly similar, the structure of the interactions and of the final model is changed as the particles add energy to the system. The Stokeslet are typically replaced by dipoles and the kernel F has a singularity in $1/|x|^2$ in dimension 3. See [82, 51] for examples of such modeling.

• *Kernels with cut-off.* Many of the kernels important for applications are singular, which poses problems both for the theory and for numerical simulations. For numerical purposes an easy remedy is to regularize the kernel. Thus instead of F, one considers F_N with a regularization depending on N.

There are of course several ways of achieving this, usually through the choice of a small scale ε_N. For instance one can take $F_N(x) = F(x)$ if $|x| \geq \varepsilon_N$ and some constant or smooth value for $|x| \geq \varepsilon_N$. In the case of the Poisson kernel, it is also possible to consider for example

$$F_N(x) = C \, \frac{x}{(|x|^2 + \varepsilon_N^2)^{d/2}}.$$

The delicate question is how to choose ε_N. Obviously the larger ε_N the smoother F_N will be, the better behaved is the system (less singularity when particles are close) and the easier it will be to show the convergence of (1.3) or (1.7). However one is not computing the real interaction and the smaller ε_N is, the closer F_N is to the real F and thus the better the approximation to the actual system.

Ideally one would be able to show regularity of the system (1.3) or (1.7) even if $\varepsilon_N = 0$. Unfortunately the present theory is unable to handle $\varepsilon_N = 0$ for most realistic force kernels F in the case of second order systems like (1.3). Note that the case of the first order (1.7) seems to be easier, see [37], [63], [75], [76], [84], [96], [97], [144], [145]...

Consequently the necessary balance between accuracy (small ε_N) and regularity (large ε_N) makes a difficult choice. The theory of convergence could inform this choice and part of the analysis we review here tries to do just that. However in practice, ε_N is usually chosen much lower than the "safe" value suggested by the theory.

In many respects, a critical scale is $\varepsilon_N \sim N^{-1/d}$ which is a sort of average minimal distance between N particles in dimension d. It is for instance the distance between two neighboring particles on a mesh. Heuristically if $\varepsilon_N << N^{-1/d}$ then it should be rather unlikely that two particles are at distance less than ε_N and the cut-off should not influence much the dynamics. On the other hand if $\varepsilon_N >> N^{-1/d}$, one would expect to see many particles with a distance less than ε_N. Note that although this argument is reasonable, a rigorous justification is out of reach, for the time being...

Unfortunately many of the mean field results for (1.3), require $\varepsilon_N >> N^{-1/d}$, even for particles initially on a mesh, see [66], Wollman [157] and Batt [12]. If particles are not initially well distributed, the assumption on ε_N is usually even worse as in [65].

• *The "typical" structure of F.* Let us summarize here which kind of assumptions one can reasonably make on F.

First of all, F is often non smooth. If it is singular though, it is usually singular only at $x = 0$ when two particles are very close. Therefore we assume that

$$\exists\, C > 0, \quad \forall\, x \in \mathbb{R}^d \backslash \{0\}, \quad |F(x)| \le \frac{C}{|x|^\alpha}\,, \quad |\nabla F(x)| \le \frac{C}{|x|^{\alpha+1}}, \tag{1.11}$$

for some α.

The behavior near $x = 0$ should be precised if a cut-off is used (see the previous point). In that case we may assume that

$$\begin{array}{ll} i) & F \text{ satisfies a } (S^\alpha) - \text{condition for some } \alpha < d-1, \\ ii) & \forall\, |x| \ge N^{-m},\ F_N(x) = F(x), \\ iii) & \forall\, |x| \le N^{-m},\ |F_N(x)| \le N^{m\alpha}. \end{array} \tag{1.12}$$

Note that these assumptions suggest that the singularity in the interaction is similar for (1.3) and (1.7) but it is not so. Consider the interaction between 2 particles i and j in the space of their relative position $X_i - X_j$ for (1.7) or their relative position $X_i - X_j$ and relative momentum $P_i - P_j$ for (1.3). In the case of (1.7), the singularity is a point, 0, of $\mathbb{R}^d$. In the case of (1.3), the singularity is a plane of dimension d of $\mathbb{R}^{2d}$, that is a much larger structure.

Finally note that in many, but not all, cases, F derives from a potential: $F = -\nabla V$. In general though, F is odd, $F(-x) = -F(x)$, as a consequence of the law of action-reaction: The force applied on particle i from particle j is the opposite of the force applied on particle j from particle i.

1.3 The limit: The Jeans-Vlasov equation

Formally the discrete system (1.3) is close to "continuous" model for large number of particles N. This model involves the distribution density of particles over $\Omega \times \mathbb{R}^d$, that is the distribution function $f(t, x, v)$ in time, position and velocity. The evolution of that function $f(t, x, v)$ is given by the Jeans-Vlasov equation (or collisionless Boltzmann equation)

$$\begin{aligned} &\partial_t f + v(P) \cdot \nabla_x f + E(t, x) \cdot \nabla_p f = 0\,, \\ &E(t, x) = \int_{\mathbb{R}^d} \rho(t, y)\, F(x - y)\, dy, \\ &\rho(t, x) = \int_{\mathbb{R}^d} f(t, x, p)\, dp, \end{aligned} \tag{1.13}$$

where here ρ is the spatial density and the initial density f^0 is given.

This equation was derived in moments form and in the context of stellar dynamics by Jeans in [102]. Under its present form, it was obtained by Vlasov, [155] and [156] for the English translation, in the context of plasmas and electrostatic interactions.

In the whole space in dimension $d \leq 3$ and in the classical case $v(P) = P$, the well posedness of the Vlasov system is now well established, up to singularity including the Poisson kernel $F = C\, x/|x|^d$. Weak solutions were obtained in [5], [60]; classical solutions for small initial data in [7]. The conditions to obtain strong solutions were formalized in [94]. Global strong solutions were finally obtained in [137], [143] (see also [95]) and at the same time in [119], see [68] and [71] as well. Strong solutions requires an initial data $f^0 \in L^1 \cap L^\infty(\mathbb{R}^{2d})$ with compact support or enough moment in velocity (see [120] for the best uniqueness condition).

Strong solutions in a bounded domain Ω are more delicate. The periodic case is handled in [134] with a new method completing [13]. The relativistic case, $v(P) = P/\sqrt{1+P^2}$, is still open. The Vlasov-Maxwell system then makes more sense (charged particles with relativistic speed create magnetic fields), see [57], [72] in dimension 2; strong solutions are still an open problem in dimension 3, we refer to [27] for the best result so far. The gravitational relativistic case with the Einstein-Vlasov system is even more delicate, see [4] for a positive result in the spherically symmetric case.

In dimension $d \geq 4$, strong solutions to Vlasov-Poisson typically only exist for short times (see again [94] for instance) and even weak solutions may have blow-up in finite time in the gravitational case, see [116]. If the force kernel is less singular then strong solutions may of course exist globally in time; for example $F \sim 1/|x|$ in [87].

For first order systems (1.7), the formal limit is the macroscopic equation

$$\partial_t \rho + \mathrm{div}(F \star \rho\, \rho) = 0, \tag{1.14}$$

on the macroscopic density $\rho(t, x)$. The well posedness theory strongly depends on the structure of F. Without any particular assumptions, we refer to [17] for a general L^p theory for aggregation equations, and to [16] for a proof of blow-up for kernels F with singularity.

If F is a gradient, $F = \nabla V$, then blow-up still occurs in general. However gradient flow techniques can be used effectively even if F is singular but typically assuming some type of convexity on V; see [38] and [39].

The special case $F = -x/|x|^d$ corresponding to the Patlak [135], Keller-Segel [105] model of chemotaxis has been extensively studied and we only quote a small subset of the relevant references here. Generally speaking blow-up may occur if the appropriate norms are above some critical value (the mass in dimension 2): See for example [90], [101], [130], [133], [139]. Below the critical value, global solutions exist, see [44], [35], [136].

In the case when $\mathrm{div}\, F = 0$ then Eq. (1.14) is in general well posed globally in time, at least for weak solutions thanks to the propagation of L^p bounds of ρ. The most important example is incompressible Euler in dimension 2, $F = C\, x^\perp/|x|^2$. Measure-valued solutions with finite energy exist globally in time, [54]. Uniqueness usually requires $\rho^0 \in L^\infty$, [103], [160] (see [138] for an extension to the critical Besov space). We also refer to [121], [43].

To summarize the discussion in this subsection, by possibly reducing the time interval $[0,\ T]$ over which we are working, we may assume that

Proposition 1. *Given $f^0 \in L^\infty(\Omega \times \mathbb{R}^d)$ and $\rho^0 \in L^\infty(\Omega)$ both with compact support, there exist a unique, compactly supported, $f \in L^\infty([0,\ T] \times \Omega \times \mathbb{R}^d)$ and $\rho \in L^\infty([0,\ T] \times \Omega)$ solutions in the sense of distribution to respectively* (1.13) *and* (1.14)*.*

1.4 The choice of the initial data

It is very delicate to choose the initial data for very large systems like (1.3) or (1.7). The question is framed very differently when one solves the system for numerical purposes or investigate the behavior of a large physical system.

When (1.3) or (1.7) is used, as particles' method, in order to approximate numerically the solution to (1.13) or (1.14), one is free to choose the initial positions and velocities provided they yield a good approximation of the initial data. One of the simplest and most common choice is to take the particles on a regular mesh. Another possibility, sometimes used for particles in a cell, is to choose randomly the positions and velocities of the prescribed number of particles in the cell under consideration.

In essentially every case, the resulting distribution of the particles is regular. The method can be repeated giving for each choice of N. This produces a sequence of initial data $(X_1^{0,N},\ P_1^{0,N}, \ldots, X_N^{0,N}, P_N^{0,N})$ or $(X_1^{0,N}, \ldots, X_N^{0,N})$, indexed by the number of particles. For simplicity we denote by $Z_i^{0,N}$ the vector $(X_i^{0,N},\ P_i^{0,N})$ or $X_i^{0,N}$ depending on whether (1.3) or (1.1) is considered. The whole vector of initial data is denoted by $Z^{0,N}$ and we will use similarly $Z_i^N(t)$ and $Z^N(t)$. The aim is to show that the corresponding sequence of solutions $Z^N(t)$ to (1.3) or (1.1) converges in an appropriate sense to the solution to (1.13) or (1.14). It is even better if good rates of convergence can be obtained.

On the other hand, in any actual physical setting, one cannot choose the initial data (or the number of particles). But experiments and observations cannot provide accurate positions and velocities for 10^{10} or more particles; they can at best give good statistical informations about the distribution of particles.

The question in this case is usually formulated in terms of the joint law of the initial positions denoted here by $f_N^0(z_1, \ldots, z_N)$. If the initial condition is deterministic then f_N^0 is simply a Dirac mass.

For indistinguishable particles, it is natural to assume that the law is invariant by permutation of the particles leading to the notion of exchangeability

Definition 1. *A law with N component $f_N^0(z_1,\ldots,z_N)$ is exchangeable if for any permutation σ, $f_N^0(z)=f_N(z_\sigma)=f_N(z_{\sigma(1)},\ldots,z_{\sigma(N)})$.*

This is of course insufficient to characterize f_N^0. Instead a usual assumption is that f_N^0 is chaotic

Definition 2. *A law f_N is chaotic if $f_N^0(z_1,\ldots,z_N)=\Pi_{i=1}^N f_{N,1}^0(z_i)$.*

That means that the initial positions of each particle is randomly and independently distributed with the 1 component law $f_{N,1}$. This is in general legitimate and can be justified in some cases.

In general the initial condition is itself the result of some dynamics. This dynamics can correspond to a different model: In experimental settings for instance, it is the result of whichever design lead to the experiment. But it can also be the same model: In cosmology for example the initial data is itself the result of the dynamics of particles (stars, galaxies) in gravitational interaction.

Therefore it is reasonable to take as initial data the equilibrium measure or a fluctuation around the equilibrium measure of a dynamical system, similar but possibly different from (1.3) or (1.1); the question is now whether the corresponding measure satisfies (2). In general this only occurs in some asymptotic sense as $N\to\infty$ but not exactly for any finite N, leading to the notion of chaotic sequences of initial data as introduced in [104] (see also [36, 88]

Definition 3. *Let E be a measurable metric space (here $E=\Omega\times\mathbb{R}^d$ or $E=\Omega$), and f a probability measure on E. A sequence $(f_N)_{N\in\mathbb{N}}$ of exchangeable probabilities on E^N is said to be f-chaotic, if one of the following equivalent properties holds:*

i) for all $k\in\mathbb{N}$, the k-marginals of f_N, defined as

$$f_{N,k}(t,z_1,\ldots,z_k)=\int_{\Omega^{N-k}\times R^{d(N-k)}} f_N(t,z_1,\ldots,z_N)\,dz_{k+1}\ldots dz_N,$$

converges weakly towards $f^{\otimes k}$ as N goes to infinity: $f_{N,k}\rightharpoonup f^{\otimes k}$, *ii) the second marginal $f_{N,2}$ converges weakly towards $f^{\otimes 2}$: $f_{N,2} \rightharpoonup f^{\otimes 2}$.*

The study of the chaoticity of equilibrium measures was performed in [30, 31, 106, 107, 124, 108]. Quantitative estimates could even be obtained in [140] for the Coulombian interaction.

1.5 The questions to solve

As a consequence of the previous discussion on initial data, there are different ways of formulating the question of the convergence of (1.3)-(1.7) to (1.13)-(1.14).

The mean field limit *per se* means showing the convergence for a specific, deterministic sequence of initial data $Z^{0,N}$. Of course the answer could depend on the choice of the sequence as the convergence could hold for some and not for others. In fact when F is singular, this is bound to happen as it is easy to choose $Z^{0,N}$ s.t. (1.3) or (1.7) is ill posed (just have all particles occupy the same initial position).

Therefore this mean field limit question can be reformulated as identifying criteria on the sequence $Z^{0,N}$ s.t. the convergence holds.

For the propagation of chaos, one needs to prove that for random initial data chosen according to Definitions 2 or 3, the (random) solutions to (1.3) or (1.7) converge to (1.13) or (1.14) with probability 1 asymptotically as $N \to \infty$. Propagation of chaos has for consequence that the solution to (1.3)-(1.7) has in fact vanishing random fluctuations around its mean, being asymptotically close to the deterministic solutions to the PDE's (1.13)-(1.14).

The two formulations are somewhat connected as for instance the propagation of chaos implies the mean field limit for almost all initial data according to the law determined by Definitions 2 or 3. Reciprocally the common strategy to obtain the propagation of chaos consists in proving the mean field limit for a large class of initial data; large enough so that initial conditions chosen according to 2 or 3 belong to it with probability close to 1.

Nevertheless many interesting mean field results do not imply the propagation of chaos: Showing the convergence for particles initially on a mesh is important for numerical purposes but irrelevant from a propagation of chaos point of view.

Apart from this discussion between mean field limit and propagation of chaos, one can also distinguish between compactness methods where only some abstract convergence is proved; and quantitative estimates explicitly bounding some distance between solutions to the ODE system and the limit. The second type of results are of course much more useful and are the only meaningful results in a non numerical context. The number N is then fixed and determined by the problem so an abstract convergence as N increases to ∞ does not imply much...

1.6 Why the mean field limit: The complexity of systems (1.3) or (1.1)

The complexity of large systems of ODE's typically increases with the dimension: They become more and more costly to solve numerically, more sensitive to changes in the initial data...

This would be a problem for systems like (1.3) and (1.7). The best direct numerical methods to solve them are probably fast particles methods as introduced in [78, 79] (see also [53] and [80] for particle-in-cell methods). They can handle up to 10^{10} particles in the right conditions. This is remarkable but still much lower than the 10^{25} particles that some applications would require. In addition as the initial data are often random as per the discussion in 1.4, one would possibly require many realizations of the solution to (1.3) or (1.7). However in practice, one notices that one realization of the system with far fewer particles is usually enough.

Our main goal is precisely to justify this fundamental reduction in complexity by proving that, with large probability, any realization of a solution to (1.3) or (1.7) is close to the solution to (1.13) or (1.14).

Note that it can only be true in some statistical sense that (1.3)-(1.7) depend only weakly on the number of particles or their exact initial positions. Obviously the trajectory of a given fixed particle will strongly depend on the starting point of the said particle. But the trajectory of most other particles and the force field E_N will not be much affected.

Second this reduction in complexity can only be true for some limited time. The behavior in large times of (1.3) or (1.7) is in general very different from the behavior of (1.13) or (1.14).

Limiting the discussion to the Hamiltonian case, (1.3) with $F = -\nabla V$, one expects the long time behavior of (1.3) to be described by some equilibrium measure, unique in the ergodic case. In particular this measure should be the same for $t \to -\infty$ and $t \to +\infty$. The typical example is the Gibbs equilibrium

$$\frac{1}{Z_N}\exp(-N\,H_N), \tag{1.15}$$

with $N\,H_N$ the total energy of the system.

However Eq. (1.13) usually has more possible equilibria. In addition even though it is formally reversible in time, (1.13) exhibits some damping of the solution to the equilibrium. This famous Landau damping, first surmised in [113], was eventually proved in [129]. This phenomenon also occurs for first order systems, like the 2D incompressible Euler as was recently shown in [14].

The nature of the long time behavior of (1.13) is hence very different from (1.3). It is further demonstrated by the fact that the limit of the solution to (1.13) is different for $t \to -\infty$ and $t \to +\infty$.

2 Well posedness for a finite number of particles

This section presents some of the well posedness results available for systems of ODE's as (1.3)-(1.7) for a fixed number of particles with two goals

- Explain in what sense we may have solutions to (1.3) or (1.7) when the interaction kernel F has some singularity. Global solutions would be ideal but if one only obtains existence for a fixed time interval then that time has to be bounded from below uniformly in N.
- Study how quantitative estimates, developed for well posedness, change as N increases. Stability estimates which are independent of N would be very useful for the mean field limit.

2.1 The Cauchy-Lipschitz theory

The Cauchy-Lipschitz theory provides the existence and uniqueness of a maximal solution to (1.3) or (1.7) if $F \in W^{1,\infty}_{loc}$. If in addition F is bounded then the solution is global in time.

Note however that if F increases too fast at ∞ or near $\partial\Omega$, the solution could diverge to ∞ or $\partial\Omega$ in finite time T and moreover that time T would in general depend on N and the initial data, which is not satisfactory.

At the heart of the Cauchy-Lipschitz theory is the Gronwall estimate. Assume from now on that $F \in W^{1,\infty}(\Omega)$ globally (so in particular a global solution exists). Consider two solutions (X, P) and (Y, Q) to (1.3) with $X = (X_1, \ldots, X_N)$, $P = (P_1, \ldots, P_N)$, and a similar notation for Y and Q. In order to compare the two solutions, one needs a norm on $\mathbb{R}^{N\,d}$, typically a p norm

$$\|U\|_p = \left(\frac{1}{N}\sum_i |U_i|^p\right)^{1/p}, \tag{2.1}$$

which is normalized here with N.

Then since $v(p)$ is Lipschitz

$$\frac{d}{dt}\|X - Y\|_p \leq \|\dot X - \dot Y\|_p = \|v(P) - v(Q)\|_p \leq C\,\|P - Q\|_p.$$

And

$$\begin{aligned}
&\frac{d}{dt}\|P - Q\|_p \leq \|\dot V - \dot W\|_p \\
&\leq \frac{1}{N}\sum_{j=1}^N \|(F(X_1 - X_j) - F(Y_1 - Y_j), \ldots, F(X_N - X_j) - F(Y_N - Y_j))\|_p \\
&\leq \frac{1}{N}\sum_{j=1}^N \|\nabla F\|_{L^\infty}\,\|(|X_1 - Y_1| + |X_j - Y_j|, \ldots, |X_N - Y_N| + |X_j - Y_j|)\|_p \\
&\leq \|\nabla F\|_{L^\infty}\,\left(\|X - Y\|_p + \|X - Y\|_1\right).
\end{aligned}$$

Since $\|U\|_1 \leq \|U\|_p$, one deduces that

$$\begin{aligned}\|X - Y\|_p + \|P - Q\|_p \leq \big(&\|X^0 - Y^0\|_p + \|P^0 - Q^0\|_p\big) \\ &\exp(t\,(1 + 2\,\|\nabla F\|_{L^\infty})).\end{aligned} \tag{2.2}$$

This estimate provides well posedness for (1.3) but even more importantly for our purpose it gives a quantitative stability estimate which is completely independent of N. A similar control is available for (1.7).

Those estimates are at the heart of the first rigorous results on the mean field limit in [29], [60], [131].

2.2 The system (1.3) with repulsive potentials

The Cauchy-Lipschitz theory can be extended to some cases with singular interactions kernel F. Because of assumption (1.11), the singularity is concentrated on the configurations with particles too close to one another: $\liminf_{t\to t_0} |X_i - X_j| = 0$ for some $i \neq j$ at some time t_0. If it is possible to show that such singularity never occur then one obtains well posedness.

The classical example is the case with repulsive potential: F is odd, $F = -\nabla V$ with $V \geq 0$ and

$$V(x) \longrightarrow +\infty, \qquad \text{as } |x| \to 0. \tag{2.3}$$

In that situation, one uses the conservation of the total energy

$$H_N(t) = \frac{1}{N}\sum_{i=1}^N e_K(P_i) + \frac{1}{2\,N^2}\sum_{i=1}^N \sum_{j\neq i} V(X_i - X_j) = H_N(t=0), \tag{2.4}$$

with $e_K(P)$ the kinetic energy of a particle with momentum P: $e_K(P) = |P|^2/2$ in the classical case, $e_K(P) = \sqrt{1+|P|^2}$ in the relativistic case and in general $\nabla_P e_K(P) = v(P)$.

Denoting the minimal distance in physical space

$$d_{N,x}(t) = \min_{i\neq j} |X_i - X_j|,$$

then (2.4) implies that

$$V(d_{N,X}(t)) \geq 2\,N^2\,H_N(t=0).$$

The combination of (1.11) and (2.3) then guarantees that the interaction remains smooth and that (1.3) is well posed for any initial data s.t. $H_N(t = 0) < \infty$.

A similar analysis may be performed for (1.7) depending on the exact structure. For instance in the gradient flow case, $F = -\nabla V$ then the potential energy

$$\frac{1}{2\,N^2}\sum_{i=1}^N \sum_{j\neq i} V(X_i - X_j)$$

is dissipated, yielding the same control on $d_{N,X}(t)$.

Let us observe that while they give abstract well posedness for a fixed N, this type of techniques do not provide any reasonable quantitative stability estimates, contrary to the previous Lipschitz case. Take as an example the Coulombian potential for (1.3). Then

$$V(x) \geq \frac{1}{C\,|x|^{d-2}},$$

except in dimension $d = 2$ where the divergence is logarithmic. Therefore this only implies

$$d_{N,X}(t) \geq \frac{(N^2\,H_N(t=0))^{1/(d-2)}}{C}.$$

Assuming that $H_N(t=0) \sim 1$, then

$$|\nabla F(X_i - X_j)| \leq \frac{C}{N^{2d/(d-2)}},$$

which combined with (2.2) yields

$$\|X - Y\|_p + \|P - Q\|_p \leq \big(\|X^0 - Y^0\|_p + \|P^0 - Q^0\|_p\big) \\ \exp(C\,t\,N^{2d/(d-2)}).$$

This estimate is not only useless for mean field limit, it is actually already extremely bad for values of N that are not so large: In dimension 3, at time $t = 1/C$, with only $N = 10$ particles, it would control the growth of the difference between (X, P) and (Y, Q) by a factor $\exp(10^6) \geq 10^{10^5}$ which for all practical purposes could just as well be $+\infty$...

2.3 The Liouville equation

In the non repulsive but singular cases, it does not seem possible to avoid collisions between particles from any initial configuration. In the oldest and most classical example of two particles under gravitational interaction, a singularity will happen in finite time only if the relative position of the particles and their relative velocities are parallel. Therefore even though a blow-up may occur, it is only the case for a set of initial data of measure 0.

A natural idea would therefore be to obtain well posedness for almost every initial configuration. This leads to the so-called Liouville equation which gives the evolution in time of the law $f_N(t, x_1, p_1 \ldots, x_N, p_N)$ of the distribution of the particles

$$\partial_t f_N + \sum_{i=1}^{N} v(p_i) \cdot \nabla_{x_i} f_N + \sum_{i=1}^{N} \frac{1}{N} \sum_{j \neq i} F(x_i - x_j) \cdot \nabla_{p_i} f_N = 0, \qquad (2.5)$$

for (1.3) and

$$\partial_t f_N + \sum_{i=1}^{N} \frac{1}{N} \sum_{j\neq i} \mathrm{div}_{x_i} \left(F(x_i - x_j)\, f_N\right) = 0, \tag{2.6}$$

for (1.7).

These equations were actually derived by J. W. Gibbs (see [69] and [70]) but are based on Liouville's earlier observation that Hamiltonian systems preserve volume.

For simplicity, we limit ourselves here to the case where $\mathrm{div}\, F = 0$ if (1.7) or (2.6) is considered. This can easily be extended to $\mathrm{div}\, F \in L^\infty$ but the situation can be very different in the other cases with possible concentrations in finite time.

Eqs (2.5) and (2.6) have two straightforward a priori estimates, which are essentially the preservation of volume noticed by Liouville

$$\|f_N(t)\|_{L^1} = \|f_N^0\|_{L^1}, \qquad \|f_N(t)\|_{L^\infty} = \|f_N^0\|_{L^\infty}, \tag{2.7}$$

yielding for instance if f_N^0 satisfies Def. 2

$$\|f_N(t)\|_{L^1} = \|f^0\|_{L^1}^N, \qquad \|f_N(t)\|_{L^\infty} = \|f^0\|_{L^\infty}^N. \tag{2.8}$$

Assume $f^0 \in L^\infty$ (or at least $f^0 \in L^1$ and not only a measure) and denote by $\mathcal{C}_N$ the configurations with singularity

$$(X, P) = (X_1, P_1, \ldots, X_N, P_N) \in \mathcal{C}_N \qquad \text{iff} \quad X_i = X_j \quad \text{for some } i \neq j.$$

Because, until the time of first collision, the dynamics is continuous in time, it is enough to show that for almost all initial configurations and any rational time $t \in \mathbb{Q}$ then $(X(t), V(t)) \notin \mathcal{C}_N$. As $\mathbb{Q}$ is countable, it would be enough to show that $\mathcal{C}_N$ is of negligible measure. Unfortunately $\mathcal{C}_N$ is unbounded, so if one defines for instance $\mathcal{C}_{N,\varepsilon}$ by

$$(X, P) = (X_1, P_1, \ldots, X_N, P_N) \in \mathcal{C}_{N,\varepsilon} \qquad \text{iff} \quad |X_i - X_j| \leq \varepsilon \quad \text{for some } i \neq j,$$

then $|\mathcal{C}_{N,\varepsilon}| = +\infty$ because of possible unbounded positions or velocities. In particular one does not have that $|\mathcal{C}_{N,\varepsilon}| \to 0$ as $\varepsilon \to 0$.

The best that one may say in general is the following if-theorem

Theorem 1. *If for any compact $K \subset \Omega^N \times \mathbb{R}^{dN}$, one has that*

$$\left|\{(X^0,P^0)\in K \mid \exists t\in\mathbb{Q}\cap[0,\ 1]\quad (X,P)|_t\in\mathcal{C}_{N,\varepsilon}\}\right| \longrightarrow 0, \qquad \text{as } \varepsilon\to 0,$$

then the system (1.3) *is well posed for almost every initial configuration.*

Because $\dot X_i = v(P_i)$, the bound on the positions is mostly irrelevant and what Theorem 1 implies is that one has to control the singularities where at least one particle has a large momentum. We give some examples in the next subsection but such a control is not easy in general even when the trajectories are close to lines ([11]).

Note that a similar theorem may be obtained for (1.7), provided again $\operatorname{div} F = 0$. The conclusion is sometimes easier in that setting: If Ω is bounded for instance then well posedness is automatic.

2.4 Well posedness in the attractive case with bounded or logarithmic potential

We briefly present as an application of Theorem 1 the classical argument about well posedness for (1.3) in the Hamiltonian case $F = \nabla V$ with V bounded or weakly singular. For simplicity assume that we are in the classical case $v(P) = P$ with periodic positions, $\Omega = \Pi^d$.

We again rely on the conservation of energy (2.4). In the case of V bounded from below by some constant C this directly implies that all momenta P_i are in a ball of diameter $\sqrt{N\,H_N(0) + N\,C}$.

Therefore if (X,P) are in a compact set of radius R then $H_N(0) \le C\,R$ and for any fixed t

$$\left|\{(X^0,P^0)\in K \mid (X,P)|_t\in\mathcal{C}_{N,2\,\varepsilon}\}\right| \le C\,N^2\,N^{d\,N/2}\,R^{d\,N/2}\,\varepsilon^d.$$

Furthermore because the velocities are bounded, one does not need to consider all $t\in\mathbb{Q}$ but only time steps of length $\varepsilon/(C\,N\,R)$. Indeed if $(X,P)|_t \in \mathcal{C}_{N,\varepsilon}$ then $(X,P)|_s \in \mathcal{C}_{N,2\,\varepsilon}$ for any $|s-t| \le \varepsilon/(C\,N\,R)^{1/2}$. Finally

$$\begin{aligned}\left|\{(X^0,P^0)\in K \mid \exists t\in\mathbb{Q}\cap[0,\ 1]\quad (X,P)|_t\in\mathcal{C}_{N,\varepsilon}\}\right|&\\ \le C\,N^{3+d\,N/2}\,R^{1+d\,N/2}\,\varepsilon^{d-1} \longrightarrow 0,&\end{aligned}$$

as $\varepsilon\to 0$. Note that this only works in dimension $d \ge 2$ and it is indeed well known that collisions between particles occur generically in dimension 1.

The previous calculation can be extended to the case where $|V(x)| \leq C \log|x|$. Consider only those initial configurations (X^0, P^0) which are not in $\mathcal{C}_{N,\varepsilon}$. This excludes an initial set of measure less than $N^2 \varepsilon^d$ and we may now assume that

$$H_N^0 \leq C\,(R + |\log\varepsilon|).$$

Following one trajectory, denote by $\bar{t}$ the first time when $(X, P) \in \mathcal{C}_{N,\varepsilon}$. Until $\bar{t}$, $|X_i - X_j| \geq \varepsilon$ for all $i \neq j$ and therefore by (2.4)

$$|P_i|^2 \leq N\,(H_N^0 + C\,|\log\varepsilon|) \leq C\,N\,(R + |\log\varepsilon|).$$

Denote $t_n = n\,\varepsilon/(C\,N\,(R + |\log\varepsilon|)^{1/2}$ and choose n s.t. $t_n \leq \bar{t} < t_{n+1}$. Then at t_n, (X, P) has to be in $\mathcal{C}_{N,2\,\varepsilon}$ and therefore

$$\begin{aligned}&\{(X^0, P^0) \in K \mid \exists t \in \mathbb{Q} \cap [0,\ 1] \quad (X,P)|_t \in \mathcal{C}_{N,\varepsilon}\}\\&\qquad\subset \bigcup_n \{(X,P)|_{t_n} \in \mathcal{C}_{N,2\,\varepsilon}, \quad |P(t_n)| \leq (C\,N\,(R + |\log\varepsilon|)^{1/2}\}.\end{aligned}$$

The conclusion is that

$$\begin{aligned}&\{(X^0, P^0) \in K \mid \exists t \in \mathbb{Q} \cap [0,\ 1] \quad (X,P)|_t \in \mathcal{C}_{N,\varepsilon}\}\\&\qquad\qquad \leq C\,N^{3+d\,N/2}\,(R + |\log\varepsilon|)^{1+d\,N/2}\,\varepsilon^{d-1},\end{aligned}$$

which still converges to 0, thus proving the well posedness. Consequently one has

Theorem 2. *Assume that $F = -\nabla V$, that F satisfies* (1.11) *and that $|V(x)| \leq C\,(1 + |\log x|)$. Then for a.e. $Z^0 \in \Pi^d \times \mathbb{R}^d$, there exists a unique solution to* (1.3).

A similar result can be obtained if $\Omega = \mathbb{R}^d$, provided some appropriate growth conditions at ∞ on V or F are assumed.

Let us observe that in the previous argument a polynomial blow-up of the potential V would not work. Given Assumption 1.11, it would be natural to assume in the attractive case that $V \sim -|x|^{1-\alpha}$. However one then obtains in the previous estimate a factor

$$(R + |\varepsilon|^{1-\alpha})^{1+d\,N/2}\,\varepsilon^{d-1}$$

which blows up as $\varepsilon \to 0$ whenever $1 - \alpha < 0$ just by choosing N large enough.

In particular while it applies to gravitational interaction in dimension $d = 2$, the previous argument fails if $d \geq 3$. It is rather surprising that a global well posedness for almost all initial data is up to now only available for $N \leq 4$, see [141, 142], in the oldest and most classical example: Gravitational, Newtonian interactions in dimension 3.

The discussion in that case resolves around the notion of collisional and non collisional singularities. A singularity corresponds to our definition here: At some t_0 the distance between two particles vanishes, $\liminf_{t\to t_0} |X_i(t) - X_j(t)| = 0$ for some $i \neq j$. The singularity is called collisional if $\lim_{t\to t_0} X_i(t)$ exists for every i. As one can readily imagine, it is a simpler case to deal with and it has been proved, see again [141], that collisional singularity occur only for negligible sets of initial data.

The difficult lies in the non collisional singularities which are proved to exist, see [159], and where at least one particle oscillates wildly, diverges to ∞ or both as $t \to t_0$, .

2.5 Renormalized solutions

The theory of renormalized solutions is now the main tool to study transport equations or dynamical systems with singular force terms. A thorough presentation would carry us far from the main scope of those notes though. In our specific case, it also turns out that renormalized solution are not necessary because of Assumption (1.11) which explicitly identifies the set where the interaction is singular. This simple (from a geometric point of view) set of singular configurations enabled us to do most of the analysis of well posedness in an elementary fashion.

However renormalized solutions allow to study the well posedness for more complicated (geometrically) and hence more general interactions. Renormalized solution are concerned with the Liouville equation (2.5)-(2.6). More precisely a solution in the sense of distributions, $f_N \in L^1 \cap L^\infty$, to (2.5) or (2.6) (provided $\operatorname{div} F = 0$) is renormalized iff for any smooth and bounded β, $\beta(f_N)$ is also a solution in the sense of distributions to (2.5) or (2.6).

If all $L^1 \cap L^\infty$ solutions are renormalized then the equation itself is said to have the renormalization property. This in particular implies that there exists a unique solution in the sense of distributions for every initial data f_N^0.

The renormalization property implies the well posedness of the flow which is a parameterized family $X(t, Z^0), P(t, Z^0)$, where as before we denote $Z^0 =$

$(X_1^0, P_1^0, \dots, X_N^0, V_N^0)$, solving (1.3) in the sense that for any t and for *a.e.* Z^0

$$\begin{aligned} X_i(t, Z^0) &= X_i^0 + \int_0^t v(P_i(s, Z^0))\, ds, \\ P_i(t, Z^0) &= P_i^0 + \int_0^t \frac{1}{N} \sum_{j \neq i} F(X_i(t, Z^0) - X_j(t, Z^0))\, ds. \end{aligned} \tag{2.9}$$

However the well posedness of the flow does not imply the existence or uniqueness for (1.3) in the sense that we used before. In other words, for any t there exists a set ω_t s.t. $|\omega_t|^c = 0$ and the equality in (2.9) is satisfied for any $Z^0 \in \omega_t$. But the set ω_t depends on t in general while in the previous subsections we had a set ω with $|\omega|^c = 0$ and s.t. for any $Z^0 \in \omega$, (2.9) is satisfied for all t.

Renormalized solution were introduced in [58] for force terms which are of bounded divergence, in $W^{1,1}_{loc}$ and globally in some L^p. This was extended to the BV_{loc} case in [26] for second order systems like (1.3), and later in [2] for the general case. It is also possible to obtain well posedness directly on the flow as in [46]. We finally refer to [3] and [50] for a more precise presentation of this subject.

For our purpose those results are very useful for (1.7) guaranteeing well posedness for a large range of interaction kernels F: $F \in L^p \cap BV_{loc}$ with $\operatorname{div} F = 0$. They do not work so well for (1.3) however because the interaction term contains $v(P)$ which does not belong to any L^p.

One of the best results so far for (1.3) or (2.5) is [83]: It requires $F \in BV_{loc}(\mathbb{R}^d \setminus \{0\})$, $F \in L^1_{loc}$ with a Hamiltonian structure $F = -\nabla V$ and a lower bound $V \geq -C\,(1+|x|^2)$. Because of the lower bound on V, it still does not apply to Newtonian gravitational interactions but it can handle kernels F with complex singularities, provided they are at least in BV_{loc}.

2.6 Conclusion on the well posedness of (1.3) and the mean field limit

The existence results discussed before cannot handle some of the most interesting kernels we wish to consider. The usual solution is to consider a truncation of the form (1.12). For any fixed N, existence to (1.3) or (1.7) is then ensured.

This is satisfactory as long as the mean field limit or propagation of chaos can be obtained without any restriction on m, that is no matter how large

m is chosen.

As a matter of fact well posedness for (1.3) can sometimes be deduced from results of propagation of chaos by using this approach and letting $m \to \infty$, see for instance [87].

Finally (2.2) is the only quantitative estimate independent of N obtained so far and which can hence be used for mean field limits. It requires $F \in W^{1,\infty}_{loc}$.

Another quantitative estimate, uniform in N but working for singular kernels F, was obtained in [10]. Unfortunately it requires that the initial law f_N^0 be chosen equal to the Gibbs equilibrium (1.15) and thus is of no use for mean field limits.

3 The main tools

We review in this section the main objects used to compare (1.3) with (1.13): The empirical measure and the marginals. Because those objects are naturally measures, we then present the classical weak distances on measures used in this setting: The $W^{-1,1}$ norm and the Monge-Kantorovich-Wasserstein (or MKW) distances.

As a first example of application, we use those concepts to compare the initial conditions of (1.13) and (1.3) in the framework of Def. 2 or Def. 3. We conclude the section with a short summary of the ways to quantify how close a distribution of particles is of being chaotic in the spirit of [126]-[88].

3.1 The empirical measure

Given particles' positions and momenta (X_i, P_i) the empirical measure is defined as

$$\mu_N(t,x,p) = \frac{1}{N} \sum_{i=1}^{N} \delta(x - X_i(t))\, \delta(p - P_i(t)). \tag{3.1}$$

Also defined as the counting measure (without the $1/N$ factor usually then), it is a probability measure counting the proportion of particles in a domain $O \subset \Omega^N \times \mathbb{R}^{dN}$

$$\#\{i \mid (X_i(t), P_i(t)) \in O\} = N \int_O \mu_N(t, dx, dp).$$

The empirical measure can similarly defined for first order systems like (1.7)

$$\mu_N(t,x) = \frac{1}{N}\sum_{i=1}^N \delta(x - X_i(t)). \tag{3.2}$$

The only important difference is the case of $2d$ incompressible Euler and (1.10) where for convenience one usually defines

$$\mu_N(t,x) = \frac{1}{N}\sum_{i=1}^N \omega_i\, \delta(x - X_i(t)). \tag{3.3}$$

In that case the empirical measure is a signed and not a probability measure, but solves exactly (1.14).

From (3.1), one can recover the full vector $(X_i, P_i)_{i=1\ldots N}$ from μ_N up to a permutation in the indices. Since the particles are here indistinguishable, the empirical measure thus contains all the information on the system.

Some early definition and use of the empirical measure can be found in [149] for instance. Its usefulness for the Boltzmann-Grad limit is restricted.

However in the mean field scaling, it enjoys a remarkable property.

Proposition 2. *Assume that F satisfies* (1.11) *and define $F(0) = 0$. Then $(X_i, P_i)_{i=1\ldots N}$ solve* (1.3) *iff μ_N defined by* (3.1) *solves* (1.13) *in the sense of distribution. Similarly the $(X_i)_{i=1\ldots N}$ solve* (1.7) *iff μ_N defined by* (3.2) *solves* (1.14) *in the sense of distribution*

In particular Prop. 2 implies that the well posedness theory for (1.13)-(1.14) for initial data which are a sum of Dirac masses is exactly the same as the well posedness for (1.3)-(1.7) described in the previous section.

The empirical measure allows us to precisely define the notion of mean field limit as introduced in Subsection 1.5

Definition 4. *Consider a particular sequence of initial data $Z^{0,N}$ or equivalently of initial empirical measure μ_N^0 s.t. $\mu_N^0 \longrightarrow \mu^0$ in the sense of distribution or equivalently for the weak-* topology of measures. The mean field limit holds iff the empirical measure $\mu_N(t)$ converges in the sense of distribution to a probability measure μ which solves* (1.13) *or* (1.14) *with initial data μ^0.*

Note that the definition in particular implies that one has to be capable of giving a meaning to μ solving the limit equation. For singular F that typically means being able to prove that $\mu \in L^p$ for p large enough.

The fact that μ_N solves the same equation (1.13) as the conjectured limit suggests one ideal way of attacking the mean field problem: Just obtain well posedness (with quantitative stability estimates if possible) for (1.13) for measure valued solutions.

Until now though, the only such well posedness result requires F to be Lipschitz. In the realistic cases where F is singular, well posedness for any measure valued solution seems completely out of reach, much harder than proving the mean field limit and may even be false. Let us emphasize here that solving (1.13) for *any* measure is very different and much more general than solving it only for sums of Dirac masses. For instance it requires some sort of uniform control as two Dirac masses merge into 1; this exactly corresponds to a collision of two particles which no one knows how to handle for (1.3) unless F is smooth.

3.2 The BBGKY hierarchy and the marginals

The marginals f_k^N were very useful in the discussion for random initial data and the definition of chaos. They remain important to understand the mean field limit.

We recall that they can be defined from the N particle distribution f_N by

$$f_{N,k}(t, z_1, \ldots, z_k) = \int_{\Omega^{N-k}\times\mathbb{R}^{d(N-k)}} f_N(t, z_1, \ldots, z_N)\, dz_{k+1} \ldots dz_N. \tag{3.4}$$

From Eqs (2.5)-(2.6) on f_N, it is possible to deduce equations on each $f_{N,k}$. Using the fact that particles are indistinguishable and the appropriate permutation, one obtains

$$\begin{aligned}
&\partial_t f_{N,k} + \sum_{i=1}^k v(p_i)\cdot\nabla_{x_i} f_{N,k} + \frac{1}{N}\sum_{i=1}^k \sum_{j=i\ldots k,\ j\neq i} F(x_i - x_j)\cdot\nabla_{p_i} f_{N,k} \\
&+ \frac{N-k}{N}\sum_{i=1}^k \int_{\Omega\times\mathbb{R}^d} F(x_i - y)\ \cdot\nabla_{p_i} f_{N,k+1}(t, x_1, p_1, \ldots, x_k, p_k, y, q)\, dy\, dq = 0,
\end{aligned} \tag{3.5}$$

for (1.3) and

$$\partial_t f_{N,k} + \frac{1}{N}\sum_{i=1}^{k}\sum_{j=i\dots k,\ j\neq i} \mathrm{div}_{x_i}\left(F(x_i - x_j)\, f_{N,k}\right)$$
$$+ \frac{N-k}{N}\sum_{i=1}^{k}\int_{\Omega} \mathrm{div}_{x_i}\left(F(x_i - y)\, f_{N,k+1}(t, x_1, \dots, x_k, y)\right)\, dy = 0, \tag{3.6}$$

for (1.7).

Those equations are not closed: The equation on $f_{N,k}$ involves the next marginal $f_{N,k+1}$. The BBGKY hierarchy, (3.5) or (3.6), was derived in a series of articles by Yvon [161], Bogolubiov [19, 20], Born and Green [25] and Kirkwood [109, 110].

Contrary to f_N which is defined on a space depending on N, each marginal $f_{N,k}$ is defined on a fixed space, depending only on k. Therefore one may consider the limit of $f_{N,k}$ as $N \to \infty$ but for a fixed k. Formally one obtains from (3.7), the Vlasov hierarchy for the limits $f_{\infty,k}$

$$\partial_t f_{\infty,k} + \sum_{i=1}^{k} v(p_i)\cdot \nabla_{x_i} f_{\infty,k}$$
$$+ \sum_{i=1}^{k}\int_{\Omega\times\mathbb{R}^d} F(x_i - y)\ \cdot \nabla_{p_i}\, f_{\infty,k+1}(t, x_1, p_1, \dots, x_k, p_k, y, q)\, dy\, dq = 0, \tag{3.7}$$

and from (3.6)

$$\partial_t f_{\infty,k} + \sum_{i=1}^{k}\int_{\Omega} \mathrm{div}_{x_i}\left(F(x_i - y)\, f_{\infty,k+1}(t, x_1, \dots, x_k, y)\right)\, dy = 0. \tag{3.8}$$

The limit $f_{\infty,k}$ correspond to the joint law of any k particles and therefore one can use the marginals to precisely define the notion of propagation of chaos

Definition 5. *Consider initial data which are chaotic as per Def. 2 or f^0-chaotic per* (3)*. Propagation of chaos holds for* (1.3) *or* (1.7) *on the time interval* $[0,\ T]$ *iff the limit $f_{\infty,k}(t)$ of each marginal in the sense of distribution (or for the weak-* topology of measures) is chaotic*

$$f_{\infty,k}(t) = \Pi_{i=1}^{k} f(t, z_k),$$

and $f(t)$ solves (1.13) *or* (1.14) *with initial data f^0.*

Just as for Def. 3, it is enough to show that $f_{\infty,2} = f(t,z_1)\,f(t,z_2)$ as this implies the equality for all others $f_{\infty,k}$. A natural approach to show propagation of chaos would be to use the BBGKY hierarchy by

- Proving rigorously that the limit $f_{\infty,k}$ solve (3.7) or (3.7), typically by passing to the limit in (3.5) or (3.6). For singular F, this requires additional estimates on the $f_{\infty,k}$, even though the hierachy is a linear system, to make sense of $F\,f_{\infty,k}$.
- Showing uniqueness of solutions to the limiting hierarchies (3.7)-(3.8), likely by requiring additional regularity estimates.

The key to this approach is that Prop. 1 guarantees the existence of a strong solution f to (1.13) or ρ to (1.14). Therefore we know one solution to the hierarchy (3.7) which is simply $\Pi_{i=1}^k f(t,z_k)$. Consequently if the $f_{\infty,k}$ solve (3.7) per the first step and the solution is unique per the second then $f_{\infty,k} = \Pi_{i=1}^k f(t,z_k)$ which is our goal.

This strategy was successfully implemented for the Boltzmann-Grad limit but it has yielded disappointing results for the mean field scaling. The problem is the lack of good estimates for (3.5) or (3.7). For instance most well posedness results for (3.7) require F *analytic*. The best result so far requires $F \in W^{1,\infty}$, see [147] for instance (and the nice presentation in [73]), and actually uses the system of particles (1.3) and (2.2) at the heart of the proof.

Many successful mean field or propagation of chaos approaches first work on the empirical measure or the ODE system (1.3) or (1.7). Because of their conceptual importance, one can then use and interpret the results and estimates in terms of the marginals.

It should be pointed out that it is simple to recover the marginals from the expectation of moments of the empirical measure. Denote by $\mathbb{E}$ the expectation with respect to the joint law f_N of the particles. For example for $k=2$ and any test function $\Phi(x_1,p_1,x_2,p_2)$

$$\begin{aligned}
&\mathbb{E}\int_{\Omega^2\times\mathbb{R}^{2d}} \Phi(x_1,p_1,x_2,p_2)\,\mu_N(t,dx_1,dp_1)\,\mu_N(t,dx_2,dp_2)\\
&= \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i}\Phi(X_i(t),P_i(t),X_j(t),P_j(t))\\
&\quad + \frac{1}{N^2}\sum_{i=1}^N \Phi(X_i(t),P_i(t),X_i(t),P_i(t)).
\end{aligned}$$

Using the fact that particles are indistinguishable, one deduces

$$\mathbb{E}\int_{\Omega^2\times\mathbb{R}^{2d}} \Phi(x_1,p_1,x_2,p_2)\,\mu_N(t,dx_1,dp_1)\,\mu_N(t,dx_2,dp_2)$$
$$=\frac{N-1}{N}\int_{\Omega^2\times\mathbb{R}^{2d}} \Phi(x_1,p_1,x_2,p_2)\,f_{N,2}(t,dx_1,dp_1,dx_2,dp_2)+O\left(\frac{1}{N}\right).$$

The calculation can be generalized for any k and leads to what is called Grunbaum lemma (see for instance [149] or [148])

$$f_{N,k}(t,z_1,\ldots,z_k)=\mathbb{E}\otimes_{i=1}^k\mu_N(t,z_i)+O\left(\frac{1}{N}\right). \tag{3.9}$$

Finally even though we only consider chaotic initial data, at least at the limit $f^0_{\infty,k}=\otimes_{i=1}^k f^0$, there is some generality in those initial conditions. Any initial hierarchy $f^0_{\infty,k}$ of initial data can be represented as a superposition of chaotic initial data, that is for some measure m on the space of probability measure $\mathcal{P}(\Omega\times\mathbb{R}^d)$

$$f^0_{\infty,k}=\int_{\mathcal{P}(\Omega\times\mathbb{R}^d)}\otimes_{i=1}^k f^0\,m(df^0), \tag{3.10}$$

as was famously proved in [91].

3.3 Distances on measures, the MKW distances

The empirical measure cannot be in any smoother space than probability measures (it is a sum of Dirac masses), even though its limit may be. The marginals may be smoother but as laws are still naturally probability measures. This has for consequence that the topology of the spaces of probability measures, denoted here by $\mathcal{P}(\Omega\times\mathbb{R}^d)$ for instance, and the various quantitative distances that metrize it are crucial for the mean field limit.

Recall that a sequence ν_N of probability measures converges for the $weak-*$ topology of measures to ν iff for any ϕ continuous with compact support

$$\int\phi\,\nu_N(dz)\longrightarrow\int\phi\,\nu(dz).$$

This can be easily extended if a time variable is present, as is the case for the empirical measure μ_N here. Then a sequence $\nu_N\in L^\infty([0,\ T],\ \mathcal{P}(E))$, where E is in general a space of the form $\Omega^k \times \mathbb{R}^{k\,d}$, converges weak-* to ν, iff for any $\phi \in L^1([0,\ T],\ C_c(E))$

$$\int_0^T \int_E \phi(t,z)\,\nu_N(t,dz)\,dt \longrightarrow \int_0^T \int_R \phi(t,z)\,\nu(t,dz)\,dt.$$

For a sequence of probability measures (or any sequence of measures with uniformly bounded total mass) the *weak* $-*$ topology is equivalent to convergence in the sense of distribution. This convergence does not imply that the limit ν is a probability measure (some mass could have been lost); instead this requires an additional tightness condition like

$$\sup_N \int_{|z|>R} \nu_N(dz) \longrightarrow 0, \quad \text{as } R \to \infty, \tag{3.11}$$

which is usually provided by control on moments

$$\sup_N \int_E |z|^k\,\nu_N(dz) < \infty. \tag{3.12}$$

One commonly used norm on $\mathcal{P}(E)$ is the $W^{-1,1}$ norm which can be defined as

$$\|\nu\|_{W^{-1,1}(E)} = \sup_{\|\phi\|_{W^{1,\infty}} \leq 1} \int_E \phi(z)\,\nu(dz). \tag{3.13}$$

The $W^{-1,1}$ norm metrizes the tight convergence of measures, *i.e.* for ν_N a sequence of probability measures, $\|\nu_N - \nu\|_{W^{-1,1}} \to 0$ iff $\nu_N \to \nu$ in the *weak* $-*$ topology and ν_N is tight as per (3.11).

The Monge-Kantorovich-Wasserstein distances are also widely used. We only summarize the main definitions and properties here and refer to [154] for example for more explanations. We first need the notion of transference plane

Definition 6. *Given two probability measures μ and ν on $\mathcal{P}(E)$, a transference plane π is a probability measure on $E \times E$ s.t.*

$$\int_E \pi(z,dw) = \mu(z), \quad \int_E \pi(dz,w) = \nu(w).$$

We denote $\Pi(\mu,\nu)$ the set of such transference planes.

The various MKW distances can then be defined

Definition 7. *The p MKW distance, denoted by* $W_p(\mu,\nu)$*, between two probability measures* μ *and* ν *is*

$$W_p^p(\mu,\nu)=\inf_{\pi\in\Pi(\mu,\nu)}\int_{E\times E}(d(z,w))^p\,\pi(dz,dw).$$

$W_\infty(\mu,\nu)$ *is simply defined as*

$$W_\infty(\mu,\nu)=\inf_{\pi\in\Pi(\mu,\nu)}\pi-esssup\ d(z,w).$$

The distance $d(z,w)$ is the natural distance on the space E, so $d(z,w)=|z-w|$ in general. But if $\Omega=\Pi^d$, the d-dimensional torus, then the appropriate distance should be used, and similarly if it is a Riemannian manifold or other more complex spaces.

There is a strict hierarchy in the MKW distances: Since π is a probability measure, by Hölder estimates for $p\leq q$

$$\int_{E\times E}(d(z,w))^p\,\pi(dz,dw)\leq\left(\int_{E\times E}(d(z,w))^q\,\pi(dz,dw)\right)^{p/q}.$$

The optimal plane for W_p may be different from the one for W_q but still by taking the infimum on the right-hand side

$$W_p(\mu,\nu)\leq W_q(\mu,\nu). \tag{3.14}$$

While it is not completely obvious at first glance, the W_p are indeed quasi-distances and satisfy the usual triangle inequality. Note as well that the infimum in the definition is realized on at least one transference plane but that this optimal plane is not unique in general.

Since $W_p(\mu,\nu)$ is not necessarily finite, it is not a distance in the strict sense. In order to guarantee finiteness, one usually demands that the p moment of each measure be finite. We hence denote by $\mathcal{P}_p(E)$ the set of probability measures ν with

$$\int_E|z|^p\,\nu(dz)<\infty.$$

If $p=\infty$ then $\mathcal{P}_\infty(E)$ is the set of probability measures with compact support.

It is sometimes easier to understand and handle the W_p distance in terms of transport maps

Definition 8. *Given two probability measures μ and ν on $\mathcal{P}(E)$, a transport map T is a measurable function $E \to E$ s.t. $T_\# \mu = \nu$ where we recall that*

$$T_\# \mu(O) = \mu(T^{-1}(O)), \quad \textit{for any measurable set } O.$$

If T is a transport map then $(Id, T)_\# \mu$ is a transference plane but obviously most transference planes cannot be represented in terms of transport maps. However if μ is absolutely continuous with respect to the Lebesgue measure then the MKW distance is realized on transport maps

Proposition 3. *Assume that μ is absolutely continuous with respect to the Lebesgue measure. Then one optimal transference plane can be represented by a transport map,* i.e. *there exists $T : E \to E$ with $T_\# \mu = \nu$ s.t.*

$$W_p^p(\mu, \nu) = \int_E (d(z, T\, z))^p \, \mu(dz).$$

If $p = \infty$ then

$$W_\infty(\mu, \nu) = \mu - esssup\ d(z - T\, z).$$

We again refer to [154] for the case $p < \infty$. The W_∞ distance is more complex and was less extensively studied; the corresponding result was only obtained recently in [41].

The most commonly used distances are the W_1, W_2 and more recently for systems of particles the W_∞ MKW distances. The W_2 distance is important in cases where the additional pseudo-Riemannian structure is useful, such as models with diffusion or gradient flow. In a deterministic setting like (1.3), it does not seem (for the moment!) to help much.

The W_1 distance is usually called the Kantorovich-Rubinstein distance and is comparable to $W^{-1,1}$ norm on compact sets; their behaviors at ∞ are different as W_1 includes a first moment. As a matter of fact, one has the duality formula for the Kantorovich-Rubinstein distance

$$W_1(\mu, \nu) = \sup_{\|\nabla \phi\|_{L^\infty} \leq 1} \int_E \phi\, (\mu(dz) - \nu(dz)). \tag{3.15}$$

Comparing (3.13) to (3.15), one sees that the only difference is that the $W^{1,\infty}$ norm of ϕ is replaced by the L^∞ norm of $\nabla \phi$.

One of the reasons why the MKW distances are so useful for systems of particles is that they generalize in some sense the p distances on $\Omega^N \times \mathbb{R}^{dN}$. They can be used to compare two empirical measures with different number of particles for instance. When the number of particles is the same in both empirical measures and the positions, momenta are close enough then the MKW distance is exactly the p distance:

Proposition 4. *Consider μ_N and ν_N two empirical measures built with* (3.1) *from two distributions of particles (X_i, P_i) and (Y_i, Q_i). Denote*

$$\delta_1 = \inf_{i\neq j} (|X_i - X_j| + |P_i - P_j|), \quad \delta_2 = \inf_{i\neq j} (|Y_i - Y_j| + |Q_i - Q_j|).$$

Assume that there exists a permutation $\bar\sigma \in \mathcal{S}_N$ s.t. for each i, $|X_i - Y_{\bar\sigma(i)}| + |P_i - Q_{\bar\sigma(i)}| < \inf(\delta_1, \delta_2)$. Then one has

$$W_p(\mu_N, \nu_N) = \inf_{\sigma\in\mathcal{S}_N} \|(X_i - Y_{\sigma(i)}, P_i - Q_{\sigma(i)})\|_p = \|(X_i - Y_{\bar\sigma(i)}, P_i - Q_{\bar\sigma(i)})\|_p.$$

Proof. For simplicity, let us assume that $\bar\sigma = Id$ (otherwise one can just relabel the indices on (Y_i, Q_i)). We may define a transport map simply by taking $T(X_i, P_i) = (Y_i, Q_i)$ for any i. We have not proved yet that this map is optimal but this already provides

$$W_p(\mu_N, \nu_N) \leq \inf_{\sigma\in\mathcal{S}_N} \|(X_i - Y_{\sigma(i)}, P_i - Q_{\sigma(i)})\|_p. \tag{3.16}$$

Consider any transference plane π for μ_N and ν_N Now because of the definition of transference plane, we may decompose

$$\pi(x,p,y,q) = \frac{1}{N^2} \sum_{i,j} \pi_{ij}\, \delta(x - X_i)\, \delta(p - P_i)\, \delta(y - Y_i)\, \delta(q - Q_i),$$

The transference plane corresponding to the transport map T is simply given by $\bar\pi_{ij} = \delta_{j=\bar\sigma(i)} = \delta_{ij}$.

Note that

$$\int_{E\times E} (|x-y| + |p-q|)^p\, \pi(dx, dp, dy, dq) = \frac{1}{N^2} \sum_{i,j} \pi_{ij}\, (|X_i - Y_j| + |P_i - Q_j|)^p.$$

But now by the assumption in the proposition

$$(|X_i - Y_j| + |P_i - Q_j|)^p \geq (|X_i - Y_i| + |P_i - Q_i|)^p,$$

with strict inequality if $j \neq i$. Therefore the optimal plane is $\bar\pi$. $\square$

It should be pointed out that the inequality (3.16) always holds without any assumption on the distributions of particles. The exact equality is not true in general though without such additional assumptions as in Prop. 4.

The W_∞ distance was first used in [122] and has started to be used extensively for mean field limit (see [37] or [87] for example). The W_1 distance corresponds to the 1 norm and therefore measures a sort of averaged distance between all the particles. But the W_∞, which is close to the ∞ norm, offers a more precise, and pointwise, control on the particles. It also automatically ensures that the empirical measures have compact support.

3.4 The distance between μ_N^0 and f^0

Define the discrete scale ε_N of the problem by

$$\text{for System (1.3), } \varepsilon_N = N^{-1/2d} \quad ; \qquad \text{for System (1.7), } \varepsilon_N = N^{-1/d}. \tag{3.17}$$

This scale is the minimum distance between an empirical measure and a smooth f^0

Proposition 5. *Let $f^0 \in \mathcal{P}(E)$ with $E = \Omega \times \mathbb{R}^d$ or $E = \Omega$ be a smooth function in $C(E)$. There exists a constant C_{f^0} s.t. for any empirical measure defined through* (3.1) *or* (3.2) *and any p*

$$W_p(f^0, \mu_N) \geq \frac{\varepsilon_N}{C_{f^0}}.$$

Proof. Since $f^0 \neq 0$ then one can find a ball $B(z_0, r)$ and a constant C s.t. $f^0 \geq 1/C$ on $B(z_0, r)$. Now for any choice of particles positions, there exist N disjoint balls $B(z_i, c_d\, r\, \varepsilon_N) \subset B(z_0, r)$ where there are no particle. The position z_i of the ball depends on the choice of the positions but not its radius (the constant c_d only depends on the dimension).

Choose for instance the test function ϕ vanishing out of $\bigcup_i B(z_i, c_d\, r\, \varepsilon_N)$ and with value $c_d\, r\, \varepsilon_N - |z - z_i|$ inside. By (3.15)

$$W_1(f^0, \mu_N) \geq \sum_i \int_{B(x_i, c_d\, r\, \varepsilon_N)} f\, \phi \geq N\, \frac{(c_d\, r\, \varepsilon_N)^{\delta+1}}{C\, 2^{\delta+1}} \geq \varepsilon_N / \tilde{C}_{f^0},$$

where $\delta = d$ for (1.7) and $\delta = 2d$ for (1.3).

This shows the proposition for the W_1 distance. However by (3.14), all the other distances are bounded from below by W_1. □

Of course if one can choose the initial distribution of particles, then it is always possible to do so in order to have $W_1(f^0, \mu_N) \sim \varepsilon_N$. Because of Prop. 5 it is not possible to have better. Note nevertheless that by using weaker norms or distances then it is possible to have higher order approximation. In general it is possible to choose the (X_i^0, P_i^0) so that

$$\|f^0 - \mu_N^0\|_{W^{-k,1}} \leq C\, \varepsilon_N^k.$$

However those weaker distances do not work well with systems of particles.

The question is more complex when the initial data is not chosen but determined randomly through Def. 2 or Def. 3. In the first case ε_N is still the right scale.

Assume that the initial distribution of particles is given by a chaotic law as per Def. 2. It was already pointed out in [152] that μ_N^0 converges to f^0 as $N \to \infty$ but quantitative estimates are now available. One has for instance from [59] that in dimension $d \geq 2$ provided f^0 is supported in a ball of radius R there exists a constant C_d depending only on the dimension s.t.

$$\mathbb{E}(W_1(\mu_N^0, f^0)) \leq C_d\, R\, \varepsilon_N. \tag{3.18}$$

This estimate was refined in [21, 22] where the deviation to the expectation is bounded: Under the same assumptions on f_N^0 and f^0, for a constant C_d, one has that

$$\mathbb{P}\left(W_1(\mu_N^0, f^0) \geq \mathbb{E}(W_1(\mu_N^0, f^0)) + \eta\right) \leq \exp(-C_d\, N\, \eta^2). \tag{3.19}$$

Unfortunately when the initial distribution is given by the more general Def. 3 then no such estimates are available. In that respect Def. 3 is not very satisfactory and it is not specific enough to say much about μ_N^0.

3.5 Some additional comments on the discrete scale ε_N

The scale ε_N is also manifested through the study of smoothing of μ_N^0. Choosing a smooth positive kernel ϕ with compact support and total mass 1, we may define as usual, for any parameter ε, the convolution kernel $\phi_\varepsilon = \varepsilon^{-\delta}\, \phi(./\varepsilon)$ with $\delta = d$ for (1.7) and $\delta = 2d$ for (1.3).

As a further illustration of the Wasserstein distances, we recall a proposition from [87] which shows that ε is then the order of the distance between μ_N^0 and its smoothed version

Proposition 6. *For any function $\phi : \mathbb{R}^{2d} \to \mathbb{R}$ radial with compact support in $B_{2d}(0,1)$ and total mass one we have for any $\mu_N^0 = \frac{1}{N}\sum_{i=1}^N \delta_{(X_i^0,P_i^0)}$*

$$W_\infty(\phi_\varepsilon \star \mu_N^0, \mu_N^0) = c_\phi \varepsilon,$$

where c_ϕ is the smallest c for which $Supp\,\phi \subset \overline{B_{2d}(0,c)}$.

A natural question is up to which scale ε, $\phi_\varepsilon \star \mu_N^0$ may still be smooth. We will limit ourselves to the analysis of $\|\phi_\varepsilon \star \mu_N^0\|_{L^\infty}$ and wish to know which is the smallest, critical ε for which this norm is still of order 1.

Let us start by the trivial bound from below: There exists a constant C_ϕ s.t. for any μ_N^0 defined through (3.1) or (3.2)

$$\|\phi_\varepsilon \star \mu_N^0\|_{L^\infty} \geq C_\phi\, \frac{\varepsilon_N^\delta}{\varepsilon^\delta}. \tag{3.20}$$

The proof of (3.20) is straightforward: Just evaluate $\phi_\varepsilon \star \mu_N^0$ at a point (X_i, P_i) or X_i to find

$$\phi_\varepsilon \star \mu_N^0(X_i^0, P_i^0) \geq \frac{\varepsilon^{-\delta}}{N}\,\phi(0).$$

This bound from below strongly suggests that ε_N is again the critical scale. If the initial positions and momenta are freely chosen, then this is easy to check. For instance if the (X_i^0, P_i^0) are taken on a mesh of size $r\,\varepsilon_N$ then

$$\|\phi_{r\,\varepsilon_N} \star \mu_N^0\|_{L^\infty} = 1. \tag{3.21}$$

But as before the case of random initial data is more complex. If they are given through the weaker Def. 3 then nothing is known. If instead they are obtained from Def. 2, the following result was proved in [87] (Prop. 8 of that article, see also [67], [24] for similar estimates)

Proposition 7. *Assume that $f^0 \in \mathcal{P}(\mathbb{R}^{2d})$ is bounded and with support in $B_{2d}(0,R)$. Assume that ϕ is bounded with support in $B(0,L)$. Assume finally that the $(X_i^0, P_i^0)_{i=1\ldots N}$ are distributed according to $f^{\otimes N}$. Consider for some $\gamma < 1$, $\varepsilon = N^{-\frac{\gamma}{2d}}$ then with $c_\phi = (4L)^{2d}\,\|\phi\|_\infty$ and $c = (2R+2)^{2d}(2L)^{-n}$*

$$\forall \lambda > 1, \quad \mathbb{P}\left(\|\phi_\varepsilon \star \mu_N^0\|_\infty \geq \lambda\, c_\phi\, \|f\|_\infty\right) \leq c_0\, N^\gamma\, e^{-(\lambda \ln \lambda - \lambda + 1)\,(4L)^{2d}\,\|f\|_\infty N^{1-\gamma}}.$$

Prop. 7 does not exactly give ε_N as the critical scale. It shows that the desired inequality holds for any $\varepsilon = \varepsilon_N^\gamma$ with $\gamma < 1$ but maybe not for $\varepsilon = \varepsilon_N$ and $\gamma = 1$.

We conclude with some comments on the connection between ε_N and the minimal distance between the particles at the initial time

$$d_N^0 = \min_{i\neq j} |X_i^0 - X_j^0| + |P_i^0 - P_j^0|.$$

If the particles are chosen on a mesh, then $d_N^0 \sim \varepsilon_N$. However if they are random, through Def. 2 for instance, then $\varepsilon_N >> d_N^0$ in general. In fact it is straightforward to see that if $\eta_N >> \varepsilon_N^2$ then the probability that $d_N^0 \geq \eta$ is 0 asymptotically in N.

In this random case d_N^0 is naturally of the order ε_N^2 as the opposite inequality from [84] shows

$$\mathbb{P}(d_N^0 \geq r\,\varepsilon_N^2) \geq e^{-C\,\|f^0\|\,r^d}. \tag{3.22}$$

3.6 Quantifying chaos

The result referred to here were obtained in [88], [126], [127]. The main difficulty and novelty in these articles is not however to prove estimates (as are presented below) on the initial data but to propagate them. We also refer to the nice summary in [125].

As we saw before the definition 3, initially due to Kac, is rather weak and does not allow a very precise control on the initial data. For this reason, it can be useful to consider stronger notions of chaos.

First of all coming back to Def. 3, it was observed that it is enough to check the convergence on $f_{N,2}^0$ as it was implying the convergence to a chaotic limit of all the other marginals. It is possible to quantify this (Theorem 2.1 in [125])

Theorem 3. *There exist constants $a,\ b \in (0,\ 1)$ s.t. for any $k,\ l \geq 2$, any initial $f_N^0 \in \mathcal{P}(E^N)$ with finite second moment and any $f^0 \in \mathcal{P}(E)$*

$$W_1(f_{N,k}^0, \otimes_{i=1}^k f^0) \leq C\,\left(W_1(f_{N,l}^0, \otimes_{i=1}^l f^0)^a + N^{-b}\right),$$

where C only depends on the second moments of $f_{N,1}^0$ and f^0.

Note that one does not need $k \leq l$ or anything of the kind. But the assumption $l \geq 2$ is crucial: Chaos can be controlled by marginals after the second but not by the first.

There are two important physical quantities that can also help quantifying chaos: The entropy and the Fisher information. The entropy is defined as

$$H_N(f_N^0) = \frac{1}{N} \int_{E^N} f_N^0 \log f_N^0 \, dz_1 \dots dz_N, \tag{3.23}$$

with $E = \Omega \times \mathbb{R}^d$ for (1.3) and $E = \Omega$ for (1.7) as before. The Fisher information is

$$I_N(f_N^0) = \frac{1}{N} \int_{E^N} \frac{|\nabla f_N^0|^2}{f_N^0} \, dz_1 \dots dz_N, \tag{3.24}$$

Note that both quantity are normalized so that for $f^0 \in \mathcal{P}(E)$

$$H_N(\otimes_{i=1}^N f^0) = H_1(f^0), \quad I_N(\otimes_{i=1}^N f^0) = I_1(f^0).$$

The entropy and Fisher information play a crucial role in collisional models and for many other dissipative systems. They do not seem to have such a special role with respect to the Gibbs Eq. (1.15); the entropy is connected to the Gibbs equilibrium and the long time behavior for (1.15). It is propagated by the equation but so are all the quantities based on the level sets of f_N. The Fisher information is not propagated in general by (1.15) and bounding it requires additional assumption on F.

The use of those quantities leads to alternative and stronger definitions of a $f^0 - chaotic$ sequence

Definition 9. *Consider the three notions*

i. f_N^0 *is* f^0*-Fisher chaotic in the sense that* $I_N(f_N^0) \to I_1(f^0)$ *and* $I_1(f^0) < \infty$.

ii. $I_N(f_N^0)$ *is uniformly bounded in* N *and* f_N^0 *is* $f^0 - chaotic$ *as per Def. 3.*

iii. f_N^0 *is* f^0*-entropy chaotic in the sense that* $H_N(f_N^0) \to H_1(f^0)$ *and* $H_1(f^0) < \infty$.

There is a strict hierarchy between these notions as per (this is for instance Theorem 2.2 in [125])

Theorem 4. *Consider any initial $f_N^0 \in \mathcal{P}(E^N)$ with uniformly bounded k-th moment for some $k > 2$ and any $f^0 \in \mathcal{P}(E)$ s.t. $f_{N,1}^0 \to f^0$ as $N \to \infty$. Then i. in Def. 9 implies ii. which implies iii which in turn implies that f_N^0 is $f^0 -$ chaotic as in Def. 3.*

To further emphasize the difficulty in Theorem (4), we emphasize that for instance $H_N(f_N^0)$ does not control $H_1(f_{N,1}^0)$ in general. There are many examples, one of them is

$$f_N^0 = \mathbb{I}_{[0,\ 1]^d} + \sum_i^N g_N(z_i),$$

with $g_N(z) = 1/4\eta_N^\delta$ for $|z| \leq \eta_N$, $g_N = -1/2$ on some fixed (independent of N) subset of $[0,\ 1]^d$ so that $\int_E g_N = 0$. Note that $f_{N,1}^0 = 1 + g(z_1)$, hence

$$H_1(f_{N,1}^0) \sim |\log \eta_N|.$$

However

$$H_N(f_N^0) \leq C\left(1 + \frac{|\log \eta_N| + \log N}{N}\right).$$

Choosing $\eta_N \to 0$ so that $|\log \eta_N| \leq N$ concludes.

In particular one could have a uniform bound on $H_N(f_N^0)$ but nevertheless have that $H_1(f^0) = \infty$. This is unfortunate as otherwise, since H_N is preserved by Gibbs, we would obtain a "free" bound on $H_1(f_{\infty,1}(t)$ at any later time t.

4 Some of the main results on mean field limits

4.1 The case F Lipschitz

The case F Lipschitz is critical in many respects. We saw with (2.2) that it is the only known case where we can have stability estimates on (1.3) which are independent of the number of particles. It is, in large part for this reason, the case where the classical results of mean field limits and propagation of chaos were obtained in [29] and then in [60], [131]. And for second order systems like (1.3), it is the only case for which mean field limits and propagation of chaos can be said to be more or less fully understood.

The key is that if F is Lipschitz then one has a well posedness theory for (1.13) in the space of probability measures. We present here the main estimate from [60]

Theorem 5. *Let $f,\ g \in \mathcal{P}(\Omega \times \mathbb{R}^d)$ be two solutions to* (1.13) *in the sense of distributions. Then*

$$W_1(f(t), g(t)) \leq W_1(f^0, g^0)\ \exp(t\,(1 + 2\,\|\nabla F\|_{L^\infty})).$$

The same result is available for (1.14) on $\mathcal{P}(\Omega)$; the exponential is just $\exp(t\,\|\nabla F\|_{L^\infty}$ in that case.

Note also that for F continuous (and a fortiori Lipschitz), measure valued solutions are straightforward to define for (1.13) or (1.14). Indeed the force term $F \star_x \int f dv$ is continuous and $f\, F \star_x \int f dv$ is then well defined.

Proof. The result can be seen as an extension of (2.2), MKW distances having replaced p norms. It can be proved either by using the characteristics or equivalently a duality method, and the formulation (3.15) of the W_1 distance.

Consider now any 1-Lipschitz $\bar\phi(x, p)$ and any time t_0. Define $\phi(t, x, p)$ the solution to

$$\partial_t \phi(t, x, p) + v(p) \cdot \nabla_x \phi + E_f(t, x) \cdot \nabla_p \phi = 0, \quad \phi(t = t_0) = \bar\phi, \tag{4.1}$$

with $E_f = F \star_x \int f\, dv$. Note that the equation on ϕ is exactly the dual of (1.13) once the force term E_f is "frozen".

If F is Lipschitz then E_f and E_g are both also Lipschitz with constant less than $\|\nabla F\|_{L^\infty}$ as both f and g have total mass 1. In particular, differentiating (4.1)

$$\begin{aligned}
&\partial_t \nabla_x \phi(t, x, p) + v(p) \cdot \nabla_x \nabla_x \phi + E_f(t, x) \cdot \nabla_p \nabla_x \phi = -\nabla_x E_f \cdot \nabla_p \phi, \\
&\partial_t \nabla_p \phi(t, x, p) + v(p) \cdot \nabla_x \nabla_p \phi + E_f(t, x) \cdot \nabla_p \nabla_p \phi = -\nabla_p v(p) \cdot \nabla_x \phi.
\end{aligned}$$

Therefore

$$\frac{d}{dt} \|\nabla \phi(t, ., .)\|_{L^\infty} \leq (1 + \|\nabla F\|_{L^\infty})\ \|\nabla \phi\|_{L^\infty},$$

implying that

$$\|\nabla \phi(t = 0)\|_{L^\infty} \leq \exp\left(t\ (1 + \|\nabla F\|_{L^\infty})\right). \tag{4.2}$$

Using (4.1) and the Lipschitz regularity of ϕ, one obtains

$$\int \bar\phi\,(f(t_0,dx,dp)-g(t_0,dx,dp)) = \int \phi(0,x,p)\,(f^0(dx,dp)-g^0(dx,dp)) + \int_0^{t_0}\int (E_f-E_g)\cdot\nabla_p\phi\, g(t,dx,dp).$$

Of course by (4.2) and (3.15)

$$\int \phi(0,x,p)\,(f^0(dx,dp)-g^0(dx,dp)) \leq W_1(f^0,g^0)\,\exp\left(t\,\left(1+\|\nabla F\|_{L^\infty}\right)\right).$$

For any fixed t,x

$$\begin{aligned} E_f(t,x)-E_g(t,x) &= \int F(x-y)\,(f(t,dy,dp)-g(t_0,dy,dp)) \\ &\leq \|\nabla F\|_{L^\infty}\,W_1(f(t,.,.),g(t,.,.)), \end{aligned}$$

by (3.15) using $F/\|\nabla F\|_{L^\infty}$ as the test function.

Combining the last two inequalities, one finds

$$\int \bar\phi\,(f(t_0,dx,dp)-g(t_0,dx,dp)) \leq W_1(f^0,g^0)\,\exp\left(t\,\left(1+\|\nabla F\|_{L^\infty}\right)\right) + \|\nabla F\|_{L^\infty}\int_0^{t_0} W_1(f(t,.,.),g(t,.,.))\,dt,$$

which gives the desired result by a last application of Gronwall's lemma.

□

Theorem 5 of course implies the most general form of mean field limit

Corollary 1. *Assume that $F\in W^{1,\infty}$. For any $f^0\in\mathcal{P}(\Omega\times\mathbb{R}^d)$, consider any sequence of initial conditions to* (1.3) *s.t. $\mu_N^0\to f^0$ in the weak $-*$ topology of measures. Then the empirical measure μ_N converges to the unique solution f to* (1.13).

Similarly by, propagation of chaos holds

Corollary 2. *Assume that $F\in W^{1,\infty}$. For any $f^0\in\mathcal{P}(\Omega\times\mathbb{R}^d)$, consider any sequence of f^0 chaotic initial conditions to* (1.3) *as per Def. 3. Then the empirical measure μ_N converges to the unique solution f to* (1.13) *with probability one.*

Because of (3.18) and (3.19), the result is even more precise. Assuming that f^0 has compact support, with very large probability in the case of f^0 chaotic initial date or for well chosen initial μ_N^0 in the deterministic case, one has that

$$W_1(\mu_N(t,.,.), f(t,.,.)) \sim \varepsilon_N \, \exp(t\,(1+2\,\|\nabla F\|_{L^\infty})). \tag{4.3}$$

This estimate provided a polynomial convergence in N, $\varepsilon_N = N^{-1/2d}$, for any finite time. However it only guarantees that μ_N will remain close to f for times of order $\log N$. This is widely conjectured to be non optimal but for the time being, it could only be improved in some limited cases, see [32, 33].

The bound (4.3) can also be used to deal with some singular but truncated kernels, in the spirit of (1.12). It requires a scale of truncation which is too high for numerical purposes as obviously for the bound to be useful, one needs that

$$\|\nabla F_N\|_{L^\infty} \leq k\,\log N,$$

meaning that the interaction must be truncated at a $\log N$ scale, and in which case the mean field limit or propagation of chaos holds until time C_d/k. This is the basis of the approach in [12], for the mean field limit, or [65] for the propagation of chaos.

Note that Theorem 5 can also be used to extend the dynamics of Eq. (1.13) to infinite dimension. More precisely, one may define a linear operator L on $\mathcal{P}(\mathcal{P}(E))$, yielding a bounded semi-group. This operator is uniquely determined by asking that $e^{t\,L}\,\delta_{f^0}$ is δ_f with f the solution to (1.13). In that sense L generalizes (1.13) to the mixed states of [91] per Eq. (3.10); the "pure" state corresponding to chaotic initial data given by Def. 2. We refer to [74, 127].

The Lipschitz case also propagates the stronger notions of chaos of Def. 9, see again for instance [125]. For example, one can easily see how to obtain strong convergence; obviously not on the empirical measure μ_N but on the marginals $f_{N,k}$. This relies on the following observation

Lemma 3. *Assume that $F \in W^{1,\infty}$, then for any k*

$$\|f_{N,k}\|_{W^{1,1}(\Omega^k\times\mathbb{R}^{kd})} \leq \|f_N^0\|_{W^{1,1}(\Omega^N\times\mathbb{R}^{Nd})}\, e^{t\,(1+\|\nabla F\|_{L^\infty})}.$$

Proof. Just differentiate the Gibbs equation (1.15) to obtain

$$\|f_N\|_{W^{1,1}(\Omega^N\times\mathbb{R}^{Nd})} \leq \|f_N^0\|_{W^{1,1}(\Omega^N\times\mathbb{R}^{Nd})}\, e^{t\,(1+\|\nabla F\|_{L^\infty})}.$$

Then by integrating, notice that

$$\|f_{N,k}\|_{W^{1,1}(\Omega^k\times\mathbb{R}^{kd})} \leq \|f_N\|_{W^{1,1}(\Omega^N\times\mathbb{R}^{Nd})}.$$

□

It is now enough to notice that $\|f_N^0\|_{W^{1,1}(\Omega^N\times\mathbb{R}^{Nd})}$ is typically bounded. For instance if the initial data is chosen according to Def. 2 then

$$\|f_N^0\|_{W^{1,1}(\Omega^N\times\mathbb{R}^{Nd})} = \|f^0\|_{W^{1,1}(\Omega\times\mathbb{R}^d)}.$$

In that case, $\|f_{N,k}\|_{W^{1,1}(\Omega^k\times\mathbb{R}^{kd})}$ is uniformly bounded for any fixed k, implying the strong convergence of all the marginals.

4.2 Some examples of the compactness method, F continuous

Before presenting more refined estimates, we show a very simple example where the mean field limit can be obtained but without any quantitative estimates.

Proposition 8. *Assume that $F\in C_0(\Omega)$. For any $f^0\in\mathcal{P}(\Omega\times\mathbb{R}^d)$, consider any sequence of initial conditions to* (1.3) *s.t. $\mu_N^0\to f^0$ in the weak $-*$ topology of measures. Then there is an extracted subsequence of the empirical measure μ_N which converges to a solution f to* (1.13) *in the sense of distribution.*

Proof. It is a direct application of the compactness in M^1 for bounded measures for the *weak* $-*$ topology of measures. Note that μ_N is uniformly in $L^\infty(\mathbb{R}_+,\ \mathcal{P}(\Omega\times\mathbb{R}^d))$. Therefore there exists σ and $f\in L^\infty(\mathbb{R}_+,\ M^1(\Omega\times\mathbb{R}^d))$ s.t.

$$\mu_{\sigma(N)}\longrightarrow f,\quad \text{in } weak-*\ L^\infty(\mathbb{R}_+,\ M^1(\Omega\times\mathbb{R}^d)).$$

Of course a priori $f\geq 0$ but it may not be a probability measure. Nevertheless recall that $\mu_{\sigma(N)}$ solves (1.13) in the sense of distribution and $F\in C_0(\Omega)\subset L^\infty$ so that

$$E(t,x)\leq \|F\|_{L^\infty}.$$

Hence at any time t

$$\int_{|x|+|p|>R}\mu_N(t,dx,dp)\leq \int_{|x|+|p|>R-\|F\|_{L^\infty}t}\mu_N^0(dx,dp),$$

and μ_N is tight as per (3.11). This shows that f is a probability measure.

It remains to pass to the limit in (1.13) where the only difficulty is the non linear term

$$\mu_{\sigma(N)} \int_{\Omega\times\mathbb{R}^d} F(x-y)\,\mu_{\sigma(N)}(t,dy,dq)$$

Since $F\in C_0(\Omega)$ then the term $\int_{\Omega\times\mathbb{R}^d} F(x-y)\,\mu_{\sigma(N)}(t,dy,dq)$ is compact in x in $C_0(\Omega)$. Compactness in time is obtained through Aubin's lemma by remarking that, from (1.13), $\partial_t\mu_N\in L^\infty([0,\ T],\ W^{-k,1}_{loc})$ uniformly in N. This enables us to conclude that $\int_{\Omega\times\mathbb{R}^d} F(x-y)\,\mu_{\sigma(N)}(t,dy,dq)$ is compact in $C_0([0,\ T]\times\Omega)$ for any finite T and finally to pass to the limit in the nonlinear term and (1.13). □

As with all compactness method, the interest Prop. 8 is limited: What if $W_1(\mu_N,f)$ is only vanishing as $1/\log N$ for instance? But its main problem is that there is no uniqueness of Eq. (1.13) with only $F\in C_0$. This is the reason why one only obtains convergence of an extracted sequence and why Prop. 8 is useful only when coupled with some additional structure on F to provide uniqueness on (1.13).

As seen from the proof, the key point in any compactness method is to pass to the limit in the non linear term. $F\in C_0$ is the critical regularity in order to do that in the space of measures. When F is more singular, additional estimates are needed, typically to control the distances between particles.

To be more specific, assume that F satisfies (1.11) and consider a smoothing F_ε for any ε s.t. $|F_\varepsilon|\le C\,|x|^{-\alpha}$, $F_\varepsilon\in C_0(\Omega)$ and $F_\varepsilon=F$ for $|x|\ge\varepsilon$. Then for a fixed ε, one may pass to the limit in

$$\mu_{\sigma(N)} \int_{\Omega\times\mathbb{R}^d} F_\varepsilon(x-y)\,\mu_{\sigma(N)}(t,dy,dq)$$

just as in the previous proof. Since $F(x-y)$ and $F_\varepsilon(x-y)$ coincide when $|x-y|\ge\varepsilon$, to conclude one would need to show that

$$\mu_{\sigma(N)} \int_{|x-y|\le\varepsilon} (F_\varepsilon(x-y)-F(x-y))\,\mu_{\sigma(N)}(t,dy,dq)\longrightarrow 0,$$

in the sense of distribution as $\varepsilon\to 0$. Of course this convergence would be trivially implies by a uniform in N bound on

$$\sup_{t\in[0,\ T]}\int_{|x-y|\le\varepsilon}\frac{1}{|x-y|^\beta}\,\mu_{\sigma(N)}(t,dy,dq)\,\mu_N(t,dx,dp) \tag{4.4}$$

for any $\beta > \alpha$. This is reminiscent of the potential energy for $F = -\nabla V$ but in that case typically $\beta = \alpha - 1$ which is not enough. Obtaining (4.4) with $\beta > \alpha$ (or even $\beta > \alpha - 1$) seems to be an extremely difficult problem in general, maybe harder than the actual mean field limit.

Instead of (4.4), most compactness methods first try to prove some other uniform estimates on the distribution of particles such as the minimal distance between particles, the maximum number of particles in a ball of small radius (the scale ε_N for instance), see for instance [86].

4.3 The incompressible $2d$ Euler

The methods here typically apply to the more general case of (1.7) with anti-symmetric kernels $F(-x) = F(x)$. This specific structure only works for 1st order model and provides additional cancellations.

The first results were obtained for grid like initial data in [76, 75, 45]; those were actually the first results obtaining the mean field limit for any singular kernel, with practical and physical importance.

The main result of [76] compares the solution of (1.7) to the characteristics of (1.14) defined as

$$\dot{X}(t, x) = F \star \rho(t, X(t, x)), \quad X(0, x) = x,$$

where ρ solves (1.14). From this system of characteristics, one defines the vector $Y_i(t)$ by $Y_i(t) = X(t, X_i^0)$ and it is possible to show that the X_i and Y_i remain very close

Theorem 6. *For $d = 2$, assume that $F = C\, x^\perp/|x|^2$ and that $\rho^0 \in \mathcal{S}$ (the Schwartz class). Take the X_i^0 on a mesh and define $\omega_i = \rho^0(X_i^0)$. Then for $4 < p < \infty$, the solution $(X_i)_{i=1\ldots N}$ to System* (1.10) *satisfies*

$$\|(X_i - Y_i)_{i=1\ldots N}\|_p \le C(t, p)\, \varepsilon_N^2,$$

where the p norm is defined by (2.1).

The estimate is remarkable as it is second order in ε_N, whereas any MKW distance between ρ^0 and μ_N^0 is at best first order. This ε_N^2 term of course relies on the very specific choices of the X_i^0 and ω_i.

The method of the proof is delicate and too long to be presented here. Instead we later show a simplified approach, relying on the minimal distance between particles, which cannot reach the critical case $F \sim 1/|x|^{d-1}$ as here but does not require the anti-symmetry of F.

The main drawback of Th. 6 (and of its various extensions) is the strong requirement on the initial positions which does not allow to treat random initial positions as per Def. 2 of 3. By using the anti-symmetry of F, it is however possible to use instead Delort's cancellation, see [54], as was done in [144, 145] to obtain

Theorem 7. *For $d = 2$, assume that $F = C\, x^\perp/|x|^2$ and that $\rho^0 \in M^1$. Consider any sequence μ_N^0 of initial data (where the empirical measure is defined through* (3.3)*) s.t. μ_N^0 converges in the weak $-*$ topology of measures to ρ^0 and with uniformly bounded kinetic energy*

$$\sup_N \frac{-1}{4\,\pi\,N^2} \sum_i \sum_{j\neq i} \log|X_i^0 - X_j^0|\,\omega_i\,\omega_j < \infty.$$

Then there exists an extracted subsequence of μ_N converging weak $-*$ to a solution ρ to (1.14).

Proof. We only sketch the main steps. For simplicity assume that $\Omega = \Pi^d$. The first step is to pass to the limit in μ_N. Because it is defined through (3.3) it is not anymore a probability measure and one has to be more careful. However since μ_N^0 converges in the *weak* $-*$ topology of measures, its total mass $\sum_i |\omega_i|$ is uniformly bounded. But this is also the total mass of $\mu_N(t,.)$ giving

$$\sup_N \sup_t |\mu_N|(t, \Pi^d) < \infty.$$

Therefore one may extract a subsequence, still denoted by μ_N for simplicity, s.t. μ_N converges in the *weak* $-*$ topology of $L_t^\infty M_x^1$ to some finite mass measure ρ.

As usual for compactness method, the difficulty is passing to the limit in the non linear term. For any $\phi \in C_c^1$, using that $F(-x) = F(x)$, one has that

$$\begin{aligned}
&\int_{\mathbb{R}_+\times\Pi^d} \phi(x)\, F\star\mu_N\mu_N(t,dx)\,dt\\
&= \int_{\mathbb{R}_+\times\Pi^{2d}} (\nabla\phi(t,x) - \nabla\phi(t,y))\cdot F(x-y)\,\mu_N(t,dx)\,\mu_N(t,dy)\,dt.
\end{aligned}$$

This is of course the way Delort uses the anti-symmetry of the equation to get an additional cancellation.

Note that $|\nabla\phi(t,x) - \nabla\phi(t,y)| \le \|\nabla^2\phi\|_{L^\infty}\,|x-y|$ and therefore if F satisfies (1.11) with $\alpha \le 1$ then the kernel

$$L(t,x,y) = (\nabla\phi(t,x) - \nabla\phi(t,y)) \cdot F(x-y)$$

is bounded in $x,\ y$. This implies that

$$\partial_t \mu_N \in L^\infty(\mathbb{R}_+,\ W^{-2,\infty}(\Pi^d)),$$

uniformly in N and gives compactness in time of μ_N. It also shows that $\rho(t=0) = \rho^0$ as a trace in the appropriate negative Sobolev space.

Now in the case where F satisfies (1.11) with $\alpha < 1$, L is actually continuous in $x,\ y$. Using the *weak* $-\ *$ convergence of measure coupled with the compactness in time, it is then straightforward to deduce that

$$\int_{\mathbb{R}_+\times\Pi^{2d}} L(t,x,y)\,\mu_N(t,dx)\,\mu_N(t,dy)\,dt \longrightarrow \int L(t,x,y)\,\rho(t,dx)\,\rho(t,dy)\,dt.$$

This would be enough to conclude but unfortunately here F satisfies (1.11) with exactly $\alpha = 1$. This is where the delicate additional work of Delort is required and where the bound on the initial kinetic energy is used. To give an idea of how one may proceed, we now assume that each $\omega_i \ge 0$ (again the general case is more difficult).

The measure μ_N is now positive. Moreover the kinetic energy is preserved by the flow of (1.10) and this has for consequence that

$$\sup_N \sup_t \int_{\Pi^{2d}} -\log|x-y|\,\mu_N(t,dx)\,\mu_N(t,dy) < \infty. \tag{4.5}$$

Consider for any $\varepsilon > 0$ a truncation L_ε of L s.t. $L_\varepsilon = L$ if $|x-y| \ge \varepsilon$ and L_ε is a smooth function, uniformly bounded in ε (just like L). Then as before for a fixed ε

$$\int_{\mathbb{R}_+\times\Pi^{2d}} L_\varepsilon(t,x,y)\,\mu_N(t,dx)\,\mu_N(t,dy)\,dt \longrightarrow \int L_\varepsilon(t,x,y)\,\rho(t,dx)\,\rho(t,dy)\,dt.$$

On the other hand, using the uniform bounds on L and L_ε, the difference with the actual term using L is

$$\begin{aligned}\int_{\mathbb{R}_+\times\Pi^{2d}} |L_\varepsilon - L|\,\mu_N(t,dx)\,\mu_N(t,dy)\,dt &\le C\int_0^T\int_{|x-y|\le\varepsilon} \mu_N(t,dx)\,\mu_N(t,dy)\,dt\\ &\le \frac{C\,T}{-\log\varepsilon}\sup_t \int_{\Pi^{2d}} -\log|x-y|\,\mu_N(t,dx)\,\mu_N(t,dy).\end{aligned}$$

Using the bound (4.5), we deduce that this difference converges to 0 as $\varepsilon \to 0$, uniformly in N. Combining this estimate with the previous convergence with L_ε allows to conclude. □

Note that Th. 7 does not provide rate of convergence (as usual for this type of method). Just as in the case $F \in C_0$, one does not have uniqueness of the Delort solution to (1.14) at the limit and therefore it is not possible in general to identify the limit, to guarantee that the whole sequence converges or to deduce propagation of chaos.

However if one assumes that ρ^0 is smoother, C^1 for instance, then with probability asymptotically close to 1 any initial data chosen according to Def. 2 has a finite kinetic energy. Moreover in that case, there exists a classical solution to (1.14) with ρ^0 as initial data. The weak-strong uniqueness principle of incompressible Euler further implies that it is unique. Therefore one can deduce that with probability 1, μ_N converges to that unique solution.

This is still not strictly propagation of chaos in the more general sense, as it is not possible to use Def. 3 because one cannot then control the initial kinetic energy.

4.4 The control of the truncated force term

As seen for instance in the proof of Th. 5, the two crucial steps in the derivation of the mean field limit are a control on the derivative of the force field and on the difference between two force fields.

At the continuous level, those bounds are easy to obtain. Indeed assume that F satisfies (1.11) with $\alpha \leq d-1$. Then ∇F is locally a Calderon-Zygmund operator, implying that if $f \in L^p$ compactly supported in $B(0,R) \subset \Omega \times \mathbb{R}^d$ then

$$\left\| \nabla F \star_x \int f(t,.,dp) \right\|_{L^p} \leq C_p \, \|f\|_{L^p} \, R^d. \tag{4.6}$$

If $\alpha < d-1$ then it is not even necessary to use Calderon-Zygmund theory and traditional convolution estimates are enough. For instance if $f \in L^1 \cap L^p$ with $1/p < 1 - d/(\alpha+1)$

$$\left\| \nabla F \star_x \int f(t,.,dp) \right\|_{L^\infty} \leq C_p \, (\|f\|_{L^1} + \|f\|_{L^p}). \tag{4.7}$$

It is natural to wonder whether such estimates can be mimicked at the discrete level.

The answer is partially positive in the sense that those techniques can indeed be adapted but only until the scale ε_N both for first and second order system. Other approaches are required to go below that scale (with a more precise control on the distribution of particles).

Among several possible ways, we present here a result based on the MKW distance between μ_N and f, in line with the more recent contributions for instance in [37, 84, 85, 87].

Proposition 9. *For any $\alpha < d-1$, any $\varepsilon > 0$, any $\rho(x) \in L^1 \cap L^\infty(\Omega)$ and any measure $\mu \in \mathcal{P}(\Omega)$*

$$\left\| \int_\Omega \frac{\mu(dy)}{(\varepsilon + |x-y|)^{\alpha+1}} \right\|_{L^\infty} \leq C_d \left(\max\left(1,\ \frac{W_\infty(\rho,\mu)}{\varepsilon}\right)\right)^{\alpha+1} (\|\rho\|_{L^1} + \|\rho\|_{L^\infty}).$$

Proof. As ρ is absolutely continuous with respect to the Lebesgue measure, there exists an optimal map T_x from it to μ.

In particular since $T_x \# \rho = \mu$

$$\begin{aligned}
&\int_\Omega \frac{1}{(\varepsilon + |x-y|)^{\alpha+1}} \mu(dy) = \int_\Omega \frac{1}{(\varepsilon + |x - T_x(y)|)^{\alpha+1}} \rho(y)\, dy \\
&\leq \left(\max\left(1,\ \frac{W_\infty(\rho,\mu)}{\varepsilon}\right)\right)^{\alpha+1} \int_\Omega \frac{\rho(y)\, dy}{(W_\infty(\rho,\mu) + |x - T_x(y)|)^{\alpha+1}} \\
&\leq \left(\max\left(1,\ \frac{W_\infty(\rho,\mu)}{\varepsilon}\right)\right)^{\alpha+1} \int_\Omega \frac{\rho(y)\, dy}{(W_\infty(\rho,\mu) - |y - T_x(y)| + |x-y|)^{\alpha+1}}.
\end{aligned}$$

As T_x is an optimal map then on the support of $\rho(y)$, $|y - T_x(y)| \leq W_\infty(\rho,\mu)$. Therefore

$$\int_\Omega \frac{1}{(\varepsilon + |x-y|)^{\alpha+1}} \mu(dy) \leq \left(\max\left(1,\ \frac{W_\infty(\rho,\mu)}{\varepsilon}\right)\right)^{\alpha+1} \int_\Omega \frac{\rho(y)\, dy}{(|x-y|)^{\alpha+1}},$$

while by the usual convolution estimates, since $\alpha + 1 < d$,

$$\left\| \int_\Omega \frac{1}{(|x-y|)^{\alpha+1}} \rho(y)\, dy \right\|_{L^\infty} \leq C_d\, (\|\rho\|_{L^1} + \|\rho\|_{L^\infty}),$$

which concludes. $\square$

Prop. 9 is naturally formulated with only the space variable x but can easily be extended to the phase space case. Indeed one obviously has

$$W_\infty\left(\int f(t,.,p)\,dp,\ \int \mu_N(t,.,dp)\right) \leq W_\infty(f,\mu_N). \tag{4.8}$$

It is also possible to replace the L^∞ norm of ρ by some L^p norm where $1/p < 1-(\alpha+1)/d$. But apart from those simple improvements, Prop. 9 is rather optimal in its main limitations.

Note that by Prop. 5, one expect $W_\infty(\rho,\mu)$ to be typically of order ε_N. Therefore the minimal scale ε for which Prop. 5 should guarantee a bound of order 1 is $\varepsilon \sim \varepsilon_N$. This is obviously the best that one can do: Just take μ an empirical measure for a distribution of positions on a mesh or grid of size ε_N. Then the maximum of the term $1/(\varepsilon+|x|)^{\alpha+1}\star\mu_N$ is of order $(\varepsilon_N/\varepsilon)^{\alpha+1}$ if $\varepsilon \leq \varepsilon_N$ by evaluating at x equal one of the particle's position.

The question of whether the W_∞ MKW distance is necessary or not is more delicate. It cannot be replaced in general by the W_1 distance. For instance consider again an empirical measures for positions on a grid of size ε_N except for M particles, with $M << N$, which occupy all the same position at $x=0$. Then

$$\int_\Omega \frac{1}{(\varepsilon_N+|y|)^{\alpha+1}}\,\mu_N(dy) \geq \frac{M}{N}\,\varepsilon_N^{-\alpha-1},$$

while $W_1(\rho,\mu_N)$ is of order $\varepsilon_N + M/N$ and cannot control the previous right-hand side if $\alpha > 0$. However this scaling suggests that W_∞ could possibly be replaced by some W_p with p large enough.

4.5 The mean field limit for truncated kernels F

Some results proving the mean field limit for (1.3) for truncated kernels as per (1.12) are found in [66, 65, 157, 87], with various techniques leading in turn to various limitations on α and the truncation ε_N. We give as an example a simplified result based on Prop. 9 or similar estimates, for which the proof can be sketched easily.

Note that the (X_i, V_i) solves (1.3) with a truncated force kernel F_N or alternatively μ_N solves (1.13) with F_N. But the conjectured limit f solves (1.13) with the "real" force kernel F. Because of that, we need a more precise version of well posedness for (1.13) (and (1.14) in the next subsection) than Prop. 1, namely

Proposition 10. *Assume that F satisfies* (1.11) *with $\alpha < d-1$. There exists $T > 0$ s.t. for any $f^0 \in L^1 \cap L^\infty(\Omega \times \mathbb{R}^d)$ and any $\rho^0 \in L^1 \cap L^\infty(\Omega)$, there exists a constant C_T for which the two solutions, f or ρ to* (1.13) *or* (1.14) *with F, and f_N or ρ_N to* (1.13) *or* (1.14) *with F_N defined from F through* (1.12)*, satisfy for any $t < T$*

$$W_1(f_N, f) \le C_T\,(N^{-m} + W_1(f_N^0, f^0)), \quad W_1(\rho_N, \rho) \le C_T\,(N^{-m} + W_1(f_N^0, f^0).$$

The constant C_T depends only on T, p , the constants in (1.11), (1.12) and the L^1 and L^∞ norm of f^0. Prop. 10 can be proved for any $T < \infty$ in many important physical situations; in particular in dimension 2 or 3, see for instance [120].

Theorem 8. *Under the assumptions of Prop. 10; assume moreover that F satisfies* (1.11) *with $\alpha < d-1$ and that F_N satisfies* (1.12) *with $m < 1/2d$,* i.e. *for a truncation $\varepsilon >> \varepsilon_N$. Consider any $\gamma < 1$ with $m < \gamma/2d$, any $f^0 \in L^1 \cap L^\infty$ with compact support and any sequence of initial data μ_N^0 s.t.*

$$\sup_N \frac{W_1(\mu_N^0, f^0)}{\varepsilon_N} < \infty, \quad \sup_N \|\phi_{N^{-\gamma/2d}} \star \mu_N^0\|_{L^\infty} < \infty.$$

Then there exists a constant C_T s.t. for any $t \le T$, the empirical measure solving (1.13) *with F_N satisfies*

$$W_1(\mu_N, f) \le C_T\,(N^{-m} + W_1(\mu_N^0, f^0)).$$

As before C_T depends only on T, p , the constants in (1.11), (1.12) and the L^1 and L^∞ norm of f^0. ϕ is any smooth, positive kernel with compact support.

Proof. There are three small scales here: $\varepsilon_N = N^{-1/2d}$, $\varepsilon = N^{-m}$ and the intermediary scale $\eta = N^{-\gamma/2d}$. By the choices of m and γ, $\varepsilon_N << \eta << \varepsilon$.

First introduce the intermediary f_N as the solution to (1.13) with F_N as force kernel but $f_N^0 = \phi_\eta \star f^0$ as an initial data. f_N is a strong solution to (1.13) as it is in L^∞ by the assumption of the theorem. Moreover

$$W_1(f_N^0, f^0) \le W_1(\mu_N^0, f^0) + W_\infty(f_N^0, \mu_N^0) \le C\,\varepsilon_N + C\,\eta,$$

by the assumptions of the theorem and by Prop. 6. By Prop. 10

$$W_1(f_N, f) \leq C_T\,(N^{-m} + C\,\varepsilon_N + C\,\eta) \leq \tilde{C}_T\,N^{-m}. \tag{4.9}$$

The main step is therefore to compare f_N with μ_N in the W_∞ distance. Both solve the same equation, (1.13) with F_N, which, because of the cut-off, is well posed for measures for any fixed N. Estimating the W_∞ distance is more complicated than for the W_1 (there is no known equivalent of (3.15)). One does not try to find the optimal map at time t but instead constructs one map (non optimal) at t based on one initial optimal map and the characteristics.

Hence define $Z(t,s,x,p) = (X,P)(t,s,x,p)$ by

$$\begin{aligned}
&\dot{X}(t,s,x,p) = v(P), \quad X(t=s,s,x,p) = x,\\
&\dot{P}(t,s,x,p) = \int_{\mathbb{R}^d} F_N(X-y)\,f_N(t,y,q)\,dy\,dq, \quad P(t=s,s,x,p) = p.
\end{aligned}$$

Denote by Z_N the characteristics associated to (1.3), that is $Z_N(t,x,v) = (X_i(t), P_i(t))$ if $X_i^0 = x$ and $P_i^0 = p$. Denote by T_0 one initial optimal map, $T_0\# f_N^0 = \mu_N^0$, which exists per Prop. 3. Define a map at time t by $T_t = Z_N \circ T_0 \circ Z(0,t,.,.)$.

There is no reason why T_t should be an optimal map but it satisfies $T_t\# f_N = \mu_N$ and thus

$$W_\infty(f_N(t,.,.), \mu_N(t,.,.)) \leq \sup_{\mathrm{Supp}\, f_N} |T_t - Id|.$$

Using now (1.3) and the system of characteristics for Z, one may now estimate, in a manner similar to the calculation in the proof of Th. 5

$$\begin{aligned}
\frac{d}{dt} W_\infty(f_N(t,.,.), \mu_N(t,.,.)) &\leq (1 + \|\nabla E_N\|_{L^\infty} + \|E_{f_N}\|_{L^\infty})\, W_\infty(f_N, \mu_N)\\
&+ \left\| \int_{\Omega\times\mathbb{R}^d} F_N(.-y)\,(f_N(t,y,q)\,dy\,dq - \mu_N(t,dy,dq)) \right\|_{L^\infty}.
\end{aligned}$$

The interested reader can find a more detailed and precise calculation in [87] for instance. The gradient of E_{f_N} is bounded by (4.7). For the gradient of E_N, by (1.11)

$$\|\nabla E_N\|_{L^\infty} \leq \left\| \frac{1}{(\varepsilon + |.|)^{\alpha+1}} \star_x \int_{\mathbb{R}^d} \mu_N(t,.,dq) \right\|_{L^\infty},$$

with $\varepsilon = N^{-m}$. By Prop. 9 and (4.8),

$$\|\nabla E_N\|_{L^\infty} \leq C \max\left(1, \left(\frac{W_\infty(f_N, \mu_N)}{\varepsilon}\right)^{\alpha+1}\right),$$

where the constant C depends on the L^1 and L^∞ norms of f_N and hence f^0.

As for the last term, fixing any x and denoting $T_t = (T_t^x, T_t^p)$

$$\int_{\Omega\times\mathbb{R}^d} F_N(x-y)\,(f_N(t,y,q)\,dy\,dq - \mu_N(t,dy,dq))$$
$$= \int_{\Omega\times\mathbb{R}^d} (F_N(x-y) - F_N(x - T_t^x(y)))\, f_N(t,y,q)\,dy\,dq.$$

By (1.12)

$$|F_N(x-y) - F_N(x-T_t^x(y))| \leq \left(\frac{C\,|y - T_t^x(y)|}{(\varepsilon + |x-y|)^{\alpha+1}} + \frac{C\,|y - T_t^x(y)|}{(\varepsilon + |x - T_t^x(y)|)^{\alpha+1}}\right)$$
$$\leq C\, W_\infty(f_N, \mu_N)\left(\frac{1}{(\varepsilon + |x-y|)^{\alpha+1}} + \frac{1}{(\varepsilon + |x - T_t^x(y)|)^{\alpha+1}}\right).$$

Therefore

$$\left|\int_{\Omega\times\mathbb{R}^d} F_N(x-y)\,(f_N(t,y,q)\,dy\,dq - \mu_N(t,dy,dq))\right| \leq C\, W_\infty(f_N,\mu_N)$$
$$\int_{\Omega\times\mathbb{R}^d}\left(\frac{f_N(t,y,q)\,dy\,dq}{(\varepsilon + |x-y|)^{\alpha+1}} + \frac{\mu_N(t,dy,dq)}{(\varepsilon + |x - T_t^x(y)|)^{\alpha+1}}\right).$$

Thus again by Prop. 9 and (4.8)

$$\left|\int_{\Omega\times\mathbb{R}^d} F_N(x-y)\,(f_N(t,y,q)\,dy\,dq - \mu_N(t,dy,dq))\right|$$
$$\leq C \max\left(1, \left(\frac{W_\infty(f_N,\mu_N)}{\varepsilon}\right)^{\alpha+1}\right) W_\infty(f_N,\mu_N).$$

Combining all those estimates

$$\frac{d}{dt} W_\infty(f_N,\mu_N) \leq C\left(1 + \left(\frac{W_\infty(f_N,\mu_N)}{\varepsilon}\right)^{\alpha+1}\right) W_\infty(f_N,\mu_N),$$

with a constant C independent of N.

By Prop. 6, initially $W_\infty(\mu_N^0, f_N^0)$ is of order η. By the definition of $\varepsilon = N^{-m} >> \eta$, the previous inequality yields a bound on $W_\infty(\mu_N, f_N)$ with a blow-up (due to the super linearity) but at a time $T_N \to \infty$ as $N \to \infty$. Therefore given any $T < \infty$, for N large enough then for any $t \leq T$

$$W_\infty(f_N, \mu_N) \leq C_T\, W_\infty(f_N^0, \mu_N^0),$$

where the constant C_T depends on the time, the dimension, the L^1 and L^∞ bounds of f^0 and the various constants in (1.11) and (1.12) but does not depend on N. Combining this inequality with (4.9) concludes. $\square$

The main drawback in Th. 8 is the limit on m which forces F_N to be truncated at too large a scale. The phase space scale ε_N is not very natural from a physical point of view and one would rather have instead the physical space scale $\varepsilon_N^2 = N^{-1/d}$ as the critical scale here. This loss occurs when one combines (4.8) with Prop. 9.

Indeed with a "good" distribution of particles, $W_\infty(f_N, \mu_N)$ could be of order ε_N while $W_\infty\left(\int f_N dq, \int \mu_N(dp)\right)$ would be of order ε_N^2. But this strongly depends on the precise distribution of particles which cannot be controlled with estimates as simple as the ones presented above. The results in [66, 157] have ε_N^2 as the critical scale while [87] for instance actually allow for a truncation ε which could be much lower than ε_N^2 but have to study more precisely the trajectories of the particles.

The only estimates in Th. 8 on the initial data are that $W_1(f^0, \mu_N^0)$ be of order ε_N and that $\|\phi_{N^{-\gamma/2d}} \star \mu_N^0\|_{L^\infty}$ be of order 1 and by (3.18)-(3.19) and by Prop. 7, these estimates are satisfied with probability asymptotically 1 for an initial data given by Def. 2. Therefore one has a "weak" propagation of chaos, weak in the sense that Def. 2 has to be used instead of Def. 3.

Corollary 4. *We put ourselves in the framework of Prop. 10. Assume that F_N satisfies* (1.12) *with $m < 1/2d$,* i.e. *for a truncation $\varepsilon >> \varepsilon_N$. Consider any $f^0 \in L^1 \cap L^\infty$ with compact support and any sequence of initial data μ_N^0 obtained from f^0 through Def. 2. Then the empirical measure μ_N, solving* (1.13) *with F_N, converges weak $-*$ to the unique solution f to* (1.13) *with F and initial data f^0.*

Many improvements that can be made to Th. 8 are however not compatible with random initial data of this sort, as in [66, 157].

4.6 The mean field limits for 1st order system with control on the minimal distance

The main criticism leveled at Th. 8 vanishes for first order systems like (1.7). Indeed in that case, the estimate (4.8) is not used, there is no phase space scale and $\varepsilon_N = N^{-1/d}$. Therefore the equivalent of Th. 8 for (1.7) would only require $N^{-m} >> N^{-1/d}$ or $m < 1/d$, obtaining the right physical scale for the critical truncation parameter.

However in that case, it is possible to completely remove any need for a truncation by considering the minimal distance between particles

$$d_N(t) = \min_{i\neq j} |X_i(t) - X_j(t)|.$$

If $d_N(t=0)$ is of order ε_N then it is possible to show that it remains of order ε_N. The idea is simply to use the previous estimates with a truncation lower than d_N which therefore does not change the dynamics.

The fact that the minimal distance can play a crucial role for the mean field limit has long been recognized. That is one the reasons why it is easier to perform the limit for particles initially on a mesh or grid, as in [76, 75]. It was used to control by itself the distribution of particles in [99] and was shown to be propagated. Its combination with Wasserstein distances was implemented in [84].

Theorem 9. *Under the assumptions of Prop. 10; assume moreover that F satisfies* (1.11) *with $\alpha < d-1$. Consider any $\rho^0 \in L^1 \cap L^\infty$ with compact support and any sequence of initial data μ_N^0 s.t.*

$$\sup_N \frac{W_1(\mu_N^0, \rho^0)}{\varepsilon_N} < \infty, \quad \inf_N \frac{d_N(0)}{\varepsilon_N} > 0.$$

Then there exists a constant C_T s.t. for any $t \leq T$, the empirical measure solving (1.14) *with F satisfies*

$$W_1(\mu_N, f) \leq C_T\,(\varepsilon_N + W_1(\mu_N^0, f^0)).$$

Proof. The proof mostly follows the one of Th. 8. One defines $\delta_N = \inf_{t\leq T} d_N(t)/2$ and F_N as $F_N = F$ for $|x| \geq \delta_N$ and F_N satisfies the estimates of (1.12) at that scale δ_N with the convention $F_N(0) = 0$. Because of this choice, the particles X_i solve the system (1.7) with either F or F_N.

One uses again $\rho_N^0 = \phi_{\varepsilon_N} \star \rho^0$ and ρ_N the solution to (1.14) with F_N. Using Prop. 10, one deduces that

$$W_1(\rho_N, \rho) \leq C_T\, \varepsilon_N. \tag{4.10}$$

With the same calculations as in the proof of Th. 8, one may show that

$$\frac{d}{dt} W_\infty(\rho_N, \mu_N) \leq W_\infty(\rho_N, \mu_N) \left\| \int_\Omega \frac{\rho_N(t,y)\, dy + \mu_N(t,dy)}{(\delta_N + |x-y|)^{\alpha+1}} \right\|_{L^\infty}.$$

Since the X_i solve (1.7) with F_N, one also deduces that

$$\frac{d}{dt} d_N(t) \geq -d_N(t)\, \|\nabla E_N\|_{L^\infty} \geq -C\, d_N(t) \left\| \int_\Omega \frac{\mu_N(t,dy)}{(\delta_N + |x-y|)^{\alpha+1}} \right\|_{L^\infty}.$$

The assumption that $d_N(0) \geq \varepsilon_N / C$ automatically guarantees that $\|\rho_N^0\|_{L^\infty} \leq C$. Prop. 1 implies that this L^∞ bound is propagated in time thus yielding a bound on

$$\left\| \int_\Omega \frac{\rho_N(t,y)\, dy}{(\delta_N + |x-y|)^{\alpha+1}} \right\|_{L^\infty}.$$

As for the other term, note that for $\beta > \alpha$

$$(\delta_N + |x-y|)^{\alpha+1} \geq \frac{1}{C} \left(\delta_N^{(\alpha+1)/(\beta+1)} + |x-y| \right)^{\beta+1}.$$

Therefore choosing $\beta > \alpha$ but $\beta < d-1$ and denoting $\nu = (\alpha+1)/(\beta+1) < 1$, one has by Prop. 9

$$\left\| \int_\Omega \frac{\mu_N(t,dy)}{(\delta_N + |x-y|)^{\alpha+1}} \right\|_{L^\infty} \leq C_T \left(1 + \frac{W_\infty(\rho_N, \mu_N)}{\delta_N^\nu} \right)^{\beta+1}.$$

Let us finally normalize the W_∞ distance and d_N, δ_N so as to work with quantities of order 1

$$\tilde{W}_\infty(t) = \frac{W_\infty(\rho_N, \mu_N)}{\varepsilon_N}, \quad \tilde{d}_N(t) = \frac{d_N(t)}{\varepsilon_N}, \quad \tilde{\delta}_N = \frac{\delta_N}{\varepsilon_N}.$$

Combining all the estimates together we obtain the differential inequalities

$$\frac{d}{dt} \tilde{W}_\infty \leq \tilde{C}_T\, \tilde{W}_\infty \left(1 + \varepsilon_N^{1-\nu} \frac{\tilde{W}_\infty}{\tilde{\delta}_N^\nu} \right)^{\beta+1},$$

$$\frac{d}{dt} \tilde{d}_N \geq -\tilde{C}_T\, \tilde{d}_N \left(1 + \varepsilon_N^{1-\nu} \frac{\tilde{W}_\infty}{\tilde{\delta}_N^\nu} \right)^{\beta+1},$$

where we recall that $\tilde{\delta}_N = \inf_{t\leq T} \tilde{d}_N(t)$. The system is hence super linear and by the Gronwall lemma may blow up at a time T_N. Nevertheless one has that $T_N \to \infty$ thanks to the term $\varepsilon_N^{1-\nu}$ which vanishes as $N \to \infty$ because $\nu < 1$. The theorem then follows by combining those bounds with (4.10). □

The one drawback of this approach is the requirement that $d_N(0) \sim \varepsilon_N$. As noted before this is not compatible with random initial data chosen according to Defs. 2 or 3 and therefore it is not possible to deduce any propagation of chaos from Th. 9.

The reason why the proof succeeds is that two very close particles cannot get much closer: If they are close then their velocities $\dot{X}_i$ and $\dot{X}_j$ are close as well provided some regularity is proved on the force field E_N. This regularity on E_N precisely relies on the control on the distance between particles.

4.7 Mean field limit and propagation of chaos for (1.3) with weakly singular force terms

The previous approach cannot be carried over to second order models: Even if two particles are very close in the physical space, $|X_i - X_j|$ small, they can get closer because their relative velocity, $v(P_i) - v(P_j)$, has no reason to be small. As a matter of fact, collisions are possible in (1.3) even for free transport, $F = 0$.

The $d = 1$ case is well understood, being somewhat simpler as the force $F(x) = \text{sign}(x)$ for the Poisson kernel is "only" discontinuous. The first mean field limit in that case, and propagation of chaos as a corollary, was obtained in [151], and re-discovered as a particular case of semi-geostrophic equations in [49]; see also a simpler proof in [85] using a weak-strong stability inequality.

In higher dimensions, the only results available so far for (1.3) for singular kernels F without truncation are [86, 87]. The main result from [87] is for instance

Theorem 10. *Assume that $\Omega = \mathbb{R}^d$, $v(p) = p$, $d \geq 2$ and that the interaction force F satisfies* (1.11) *for $\alpha < 1$. Choose any $0 < \gamma < 1$.*

Assume that $f^0 \in L^\infty(\mathbb{R}^{2d})$ is non-negative, and has compact support and total mass one, and denote by f the unique non-negative, global, bounded, and compactly supported solution f of the Vlasov equation (1.13)*, as per Prop. 1.*

Assume that the initial conditions $(X_i^0, P_i^0)_{i=1\ldots N}$ are such that for each N, there exists a global solution to the N particle system (1.3)*, and that the initial empirical distributions μ_N^0 of the particles satisfy*

i) For a constant C_∞ independent of N,

$$\sup_{z\in\mathbb{R}^{2d}} N^\gamma \mu_N^0\Big(B_{2d}\big(z, N^{-\frac{\gamma}{2d}}\big)\Big) \leq C_\infty, \quad \textit{and} \quad \|f_0\|_\infty \leq C_\infty;$$

ii) For some $R_0 > 0$, $\forall N \in \mathbb{N}$, $\quad supp\, \mu_N^0 \subset B_{2d}(0, R_0)$;

iii) for some $r \in (0, r^*)$ where $r^* := \frac{d-1}{1+\alpha}$,

$$\inf_{i\neq j} |(X_i^0, V_i^0) - (X_j^0, V_j^0)| \geq N^{-\gamma(1+r)/2d}.$$

Then for any $T > 0$ given by Prop. 1, there exist constants $C_0(R_0, C_\infty, F, T)$ and $C_1(R_0, C_\infty, F, \gamma, r, T)$ such that for $N \geq C_1$, the following estimate holds

$$\forall\, t \in [0,\ T], \quad W_1(\mu_N(t), f(t)) \leq e^{C_0 t}\Big(W_1(\mu_N^0, f^0) + 2\, N^{-\frac{\gamma}{2d}}\Big). \tag{4.11}$$

From the discussion in the third section and in particular (3.18)-(3.19) and Prop. 7, one can check that the assumptions $[i]--[iii]$ are generic for chaotic initial data and it is possible to deduce a weak propagation of chaos

Corollary 5. *Assume that $d \geq 3$ and that F satisfies* (1.11) *with $\alpha < 1$. There exist a positive real number $\gamma^* \in (0,1)$ depending only on (d, α) and a function $s^* : \gamma \in (\gamma^*, 1) \to s^*_\gamma \in (0, \infty)$ s.t.:*

- For any non negative initial data $f^0 \in L^\infty(\mathbb{R}^{2d})$ with compact support and total mass one, denoting by f the unique global, non-negative bounded, and compactly supported solution f to the Vlasov equation (1.13)*, see Prop. 1;*

- For each $N \in \mathbb{N}^*$, denoting by μ_N the empirical measure corresponding to the solution to (1.3) *with initial positions $(X_i^0, V_i^0)_{i\leq N}$ chosen randomly according to the probability $(f^0)^{\otimes N}$ per Def. 2;*

Then, for all $T > 0$, any

$$\gamma^* < \gamma < 1 \quad \textit{and} \quad 0 < s < s^*_\gamma,$$

there exists three positive constants $C_0(T, f, F)$, $C_1(\gamma, s, T, f, F)$ and $C_2(f^0, \gamma)$ such that for $N \geq C_1$

$$\mathbb{P}\left(\exists\, t \in [0, T],\ W_1(\mu_N(t), f(t)) \geq 3\, e^{C_0 t}\, N^{-\frac{\gamma}{2d}}\right) \leq \frac{C_2}{N^s}. \tag{4.12}$$

The constants C_1 and C_2 blow up when γ or s approach their maximum value.

The main limitation for Th. 10 and Corollary 5 is the condition $\alpha < 1$. The kernel F is then sometimes called weakly singular since, if $F = -\nabla V$, then the potential V is continuous. It is unfortunately fair to say that the mean field limit for $\alpha \geq 1$ is mostly not understood at all.

The proof of Th. 10 is intricate and we do not try to present it here. Instead we attempt to explain where and why the condition $\alpha < 1$ is useful. From Prop. 9, it is used at the discrete level of the problem, *i.e.* when two particles are very close $|X_i - X_j| \leq \varepsilon_N$.

At this scale, one does not try to compare anymore the discrete and continuous forces as in the proof of Ths 8 and 9. Instead the goal is only to show that the contribution from such close pairs of particles is small. The first difficulty is that one may have collisions (or near collisions) and the force term is hence not even bounded pointwise in time. This difficulty is solved by averaging the force over a small time interval $[t,\ t+\varepsilon]$ with $\varepsilon >> \varepsilon_N$ well chosen.

Consider now two close particles $j \neq i$ at t, and neglect the variation of velocities on $[t, t+\varepsilon]$. Because of (1.11), with $\alpha < 1$, we have

$$\int_t^{t+\varepsilon} |F(X_i(s) - X_j(s))|\, ds \sim \int_t^{t+\varepsilon} \frac{ds}{|\delta + (s-s_0)(V_i - V_j)|^\alpha} \lesssim \frac{\varepsilon^{1-\alpha}}{|V_i - V_j|^{-\alpha}}$$

where δ is the minimal distance between the two particles on the time interval $[t, t+\varepsilon]$, which is reached at time denoted s_0.

Obviously this estimate is only possible if the integral in time is bounded, independently how small δ may be; thus the requirement $\alpha < 1$. The full contribution is then obtained after a careful summation on all the particles j of the domain, using the W_∞ distance.

Let us add that this condition $\alpha < 1$ also appears in the classical calculation of the angle deviation between two particles undergoing a near collision. If $\alpha < 1$, the deviation in velocity due to a collision (another particle coming very close) with a sufficiently large relative velocity cannot be too large: for instance, two particles with sufficiently large relative velocity will never bounce back even if they exactly collides at some time. So one does not expect any fast variation in the velocities of the particles (the difficulty is of course to prove this rigorously). The only "bad events" are the collisions with very small relative velocities, which are controlled by a lower bound on the distance in $\mathbb{R}^{2d}$ between particles.

In contrast when $\alpha > 1$, a particle coming very close to another one can change its velocity over a very short time interval (even if their relative velocity remains of order 1): For instance the two particles can bounce back.

Acknowledgements. The author had many discussions about this subject over the years with J. Barré, J. Carrillo and especially M. Hauray. Those discussions have proved invaluable when writing those notes.

References

[1] S. J. Aarseth. *Gravitational N-Body Simulations*. Cambridge, UK: Cambridge University Press, 2010, Jan. 2010.

[2] L. Ambrosio. Transport equation and Cauchy problem for BV vector fields. *Invent. Math.*, 158(2):227–260, 2004.

[3] L. Ambrosio. Transport equation and Cauchy problem for non-smooth vector fields. In *Calculus of variations and nonlinear partial differential equations*, volume 1927 of *Lecture Notes in Math.*, pages 1–41. Springer, Berlin, 2008.

[4] H. Andréasson, M. Kunze, and G. Rein. Global existence for the spherically symmetric Einstein-Vlasov system with outgoing matter. *Comm. Partial Differential Equations*, 33(4-6):656–668, 2008.

[5] A. A. Arsen′ev. Existence in the large of a weak solution of Vlasov's system of equations. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 15:136–147, 276, 1975.

[6] R. Balescu. *Equilibrium and nonequilibrium statistical mechanics.* Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1975.

[7] C. Bardos and P. Degond. Global existence for the Vlasov-Poisson equation in 3 space variables with small initial data. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 2(2):101–118, 1985.

[8] C. Bardos, L. Erdös, F. Golse, N. Mauser, and H.-T. Yau. Derivation of the schrdinger-poisson equation from the quantum n-body problem. *C. R. Math. Acad. Sci. Paris*, 334:515520, 2002.

[9] C. Bardos, F. Golse, A. D. Gottlieb, and N. J. Mauser. Mean field dynamics of fermions and the time-dependent Hartree-Fock equation. *J. Math. Pures Appl. (9)*, 82(6):665–683, 2003.

[10] J. Barré, M. Hauray, and P. E. Jabin. Stability of trajectories for N-particle dynamics with a singular potential. *Journal of Statistical Mechanics: Theory and Experiment*, 7, July 2010.

[11] J. Barré and P. E. Jabin. Free transport limit for N-particles dynamics with singular and short range potential. *J. Stat. Phys.*, 131(6):1085–1101, 2008.

[12] J. Batt. N-particle approximation to the nonlinear Vlasov-Poisson system. In *Proceedings of the Third World Congress of Nonlinear Analysts, Part 3 (Catania, 2000)*, volume 47, pages 1445–1456, 2001.

[13] J. Batt and G. Rein. Global classical solutions of the periodic Vlasov-Poisson system in three dimensions. *C. R. Acad. Sci. Paris Sér. I Math.*, 313(6):411–416, 1991.

[14] J. Bedrossian and N. Masmoudi. Asymptotic stability for the Couette flow in the 2D Euler equations. *Appl. Math. Res. Express. AMRX*, (1):157–175, 2014.

[15] L. Berlyand, P. Jabin, and M. Potomkin. Complexity reduction in many particle systems with random initial data. *Submitted to J. Uncertainty Quantification.*

[16] A. L. Bertozzi, J. A. Carrillo, and T. Laurent. Blow-up in multidimensional aggregation equations with mildly singular interaction kernels. *Nonlinearity*, 22(3):683–710, 2009.

[17] A. L. Bertozzi, T. Laurent, and J. Rosado. L^p theory for the multidimensional aggregation equation. *Comm. Pure Appl. Math.*, 64(1):45–83, 2011.

[18] C. Birdsall and A. Langdon. *Plasma physics via computer simulation.* Series in plasma physics. Adam Hilger, 1991.

[19] N. N. Bogoliubov. Kinetic equations. *Journal of Experimental and Theoretical Physics (in Russian)*, 16 (8):691–702, 1946.

[20] N. N. Bogoliubov. Kinetic equations. *Journal of Physics USSR*, 10 (3):265–274, 1946.

[21] E. Boissard. *Problèmes d'interaction discret-continu et distances de Wasserstein*. PhD thesis, Université de Toulouse III, 2011.

[22] E. Boissard. Simple bounds for convergence of empirical and occupation measures in 1-Wasserstein distance. *Electron. J. Probab.*, 16:no. 83, 2296–2333, 2011.

[23] F. Bolley, J. A. Cañizo, and J. A. Carrillo. Stochastic mean-field limit: non-Lipschitz forces and swarming. *Math. Models Methods Appl. Sci.*, 21(11):2179–2210, 2011.

[24] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields*, 137(3-4):541–593, 2007.

[25] M. Born and H. S. Green. A general kinetic theory of liquids i. the molecular distribution functions. *Proc. Roy. Soc. A*, 188:10–18, 1946.

[26] F. Bouchut. Renormalized solutions to the Vlasov equation with coefficients of bounded variation. *Arch. Ration. Mech. Anal.*, 157(1):75–90, 2001.

[27] F. Bouchut, F. Golse, and C. Pallard. Classical solutions and the Glassey-Strauss theorem for the 3D Vlasov-Maxwell system. *Arch. Ration. Mech. Anal.*, 170(1):1–15, 2003.

[28] F. Bouchut, F. Golse, and M. Pulvirenti. *Kinetic Equations and Asymptotic Theory*. L. Desvillettes and B. Perthame eds, Gauthier-Villars, Paris, 2000.

[29] W. Braun and K. Hepp. The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. *Comm. Math. Phys.*, 56(2):101–113, 1977.

[30] E. Caglioti, P.-L. Lions, C. Marchioro, and M. Pulvirenti. A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description. *Comm. Math. Phys.*, 143(3):501–525, 1992.

[31] E. Caglioti, P.-L. Lions, C. Marchioro, and M. Pulvirenti. A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description. II. *Comm. Math. Phys.*, 174(2):229–260, 1995.

[32] E. Caglioti and F. Rousset. Long time behavior of particle systems in the mean field limit. *Commun. Math. Sci.*, (suppl. 1):11–19, 2007.

[33] E. Caglioti and F. Rousset. Quasi-stationary states for particle systems in the mean-field limit. *J. Stat. Phys.*, 129(2):241–263, 2007.

[34] E. Caglioti and F. Rousset. Long time estimates in the mean field limit. *Arch. Ration. Mech. Anal.*, 190(3):517–547, 2008.

[35] V. Calvez and L. Corrias. The parabolic-parabolic Keller-Segel model in $\mathbb{R}^2$. *Commun. Math. Sci.*, 6(2):417–447, 2008.

[36] E. A. Carlen, M. C. Carvalho, J. Le Roux, M. Loss, and C. Villani. Entropy and chaos in the Kac model. *Kinet. Relat. Models*, 3(1):85–122, 2010.

[37] J. Carrillo, Y.-P. Choi, and M. Hauray. The derivation of swarming models: Mean-field limit and wasserstein distances. In *Collective Dynamics from Bacteria to Crowds*, volume 553 of *CISM International Centre for Mechanical Sciences*, pages 1–46. Springer Vienna, 2014.

[38] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepčev. Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Math. J.*, 156(2):229–271, 2011.

[39] J. A. Carrillo, S. Lisini, and E. Mainini. Gradient flows for non-smooth interaction potentials. *Nonlinear Anal.*, 100:122–147, 2014.

[40] C. Cercignani, R. Illner, and M. Pulvirenti. *The Mathematical Theory of Dilute Gases*. Springer-Verlag, New York, 1994.

[41] T. Champion, L. De Pascale, and P. Juutinen. The ∞-Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.*, 40(1):1–20, 2008.

[42] P.-H. Chavanis. Hamiltonian and Brownian systems with long-range interactions. V. Stochastic kinetic equations and theory of fluctuations. *Phys. A.*, 387(23):5716–5740, 2008.

[43] J.-Y. Chemin. *Perfect incompressible fluids*, volume 14 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press, Oxford University Press, New York, 1998. Translated from the 1995 French original by Isabelle Gallagher and Dragos Iftimie.

[44] L. Corrias, B. Perthame, and H. Zaag. Global solutions of some chemotaxis and angiogenesis systems in high space dimensions. *Milan J. Math.*, 72:1–28, 2004.

[45] G.-H. Cottet, J. Goodman, and T. Y. Hou. Convergence of the grid-free point vortex method for the three-dimensional Euler equations. *SIAM J. Numer. Anal.*, 28(2):291–307, 1991.

[46] G. Crippa and C. De Lellis. Estimates and regularity results for the DiPerna-Lions flow. *J. Reine Angew. Math.*, 616:15–46, 2008.

[47] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Trans. Automat. Control*, 52:852–862, 2007.

[48] F. Cucker and S. Smale. On the mathematics of emergence. *Japan J. Math.*, 2:197–227, 2007.

[49] M. Cullen, W. Gangbo, and G. Pisante. The semigeostrophic equations discretized in reference and dual variables. *Arch. Ration. Mech. Anal.*, 185(2):341–363, 2007.

[50] C. De Lellis. Notes on hyperbolic systems of conservation laws and transport equations. In *Handbook of differential equations: evolutionary equations. Vol. III*, Handb. Differ. Equ., pages 277–382. Elsevier/North-Holland, Amsterdam, 2007.

[51] P. Degond, A. Frouvelle, and J.-G. Liu. Macroscopic limits and phase transition in a system of self-propelled particles. *J. Nonlinear Sci.*, 23(3):427–456, 2013.

[52] P. Degond and F.-J. Mustieles. A deterministic approximation of diffusion equations using particles. *SIAM J. Sci. Statist. Comput.*, 11(2):293–310, 1990.

[53] W. Dehnen. A Very Fast and Momentum-conserving Tree Code. *The Astrophysical Journal*, 536:L39–L42, June 2000.

[54] J.-M. Delort. Existence de nappes de tourbillon en dimension deux. *J. Amer. Math. Soc.*, 4(3):553–586, 1991.

[55] L. Desvillettes, F. Golse, and V. Ricci. The mean-field limit for solid particles in a Navier-Stokes flow. *J. Stat. Phys.*, 131(5):941–967, 2008.

[56] M. Di Francesco, P. A. Markowich, J.-F. Pietschmann, and M.-T. Wolfram. On the Hughes' model for pedestrian flow: the one-dimensional case. *J. Differential Equations*, 250(3):1334–1362, 2011.

[57] R. J. DiPerna and P.-L. Lions. Global weak solutions of Vlasov-Maxwell systems. *Comm. Pure Appl. Math.*, 42(6):729–757, 1989.

[58] R. J. DiPerna and P.-L. Lions. Ordinary differential equations. *Invent. Math*, 98:511–547, 1989.

[59] V. Dobrić and J. E. Yukich. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.*, 8(1):97–118, 1995.

[60] R. L. Dobrušin. Vlasov equations. *Funktsional. Anal. i Prilozhen.*, 13(2):48–58, 96, 1979.

[61] L. Erdős, B. Schlein, and H.-T. Yau. Derivation of the cubic non-linear Schrödinger equation from quantum dynamics of many-body systems. *Invent. Math.*, 167(3):515–614, 2007.

[62] L. Erdős, B. Schlein, and H.-T. Yau. Derivation of the Gross-Pitaevskii equation for the dynamics of Bose-Einstein condensate. *Ann. of Math. (2)*, 172(1):291–370, 2010.

[63] N. Fournier, M. Hauray, and S. Mischler. Propagation of chaos for the 2d viscous vortex model. to appear in JEMS. arXiv:1212.1437, 2012.

[64] I. Gallagher, L. Saint-Raymond, and B. Texier. From newton to boltzmann : hard spheres and short-range potentials. Zurich Advanced Lectures in Mathematics Series, 2014.

[65] K. Ganguly, J. T. Lee, and H. D. Victory, Jr. On simulation methods for Vlasov-Poisson systems with particles initially asymptotically distributed. *SIAM J. Numer. Anal.*, 28(6):1574–1609, 1991.

[66] K. Ganguly and H. D. Victory, Jr. On the convergence of particle methods for multidimensional Vlasov-Poisson systems. *SIAM J. Numer. Anal.*, 26(2):249–288, 1989.

[67] F. Gao. Moderate deviations and large deviations for kernel density estimators. *J. Theoret. Probab.*, 16(2):401–418, 2003.

[68] I. Gasser, P.-E. Jabin, and B. Perthame. Regularity and propagation of moments in some nonlinear Vlasov systems. *Proc. Roy. Soc. Edinburgh Sect. A*, 130(6):1259–1273, 2000.

[69] J. W. Gibbs. On the Fundamental Formulae of Dynamics. *Amer. J. Math.*, 2(1):49–64, 1879.

[70] J. W. Gibbs. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundation of thermodynamics.* Dover publications, Inc., New York, 1960.

[71] R. T. Glassey. *The Cauchy problem in kinetic theory.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[72] R. T. Glassey and J. Schaeffer. The relativistic Vlasov-Maxwell system in two space dimensions. I, II. *Arch. Rational Mech. Anal.*, 141(4):331–354, 355–374, 1998.

[73] F. Golse. On the dynamics of large particle systems in the mean field limit. arXiv:1301.5494, 2013.

[74] F. Golse, C. Mouhot, and V. Ricci. Empirical measures and Vlasov hierarchies. *Kinet. Relat. Models*, 6(4):919–943, 2013.

[75] J. Goodman and T. Y. Hou. New stability estimates for the 2-D vortex method. *Comm. Pure Appl. Math.*, 44(8-9):1015–1031, 1991.

[76] J. Goodman, T. Y. Hou, and J. Lowengrub. Convergence of the point vortex method for the 2-D Euler equations. *Comm. Pure Appl. Math.*, 43(3):415–430, 1990.

[77] H. Grad. On the kinetic theory of rarefied gases. *Comm. on Pure and Appl. Math.*, 2:331–407, 1949.

[78] L. Greengard and V. Rokhlin. A fast algorithm for particle simulation. *Journal of Computational Physics*, 73:325–348, 1987.

[79] L. Greengard and V. Rokhlin. Rapid evaluation of potential fields in three dimensions. *Lecture Notes in Mathematics*, 1360:121–141, 1988.

[80] Y. N. Grigoryev, V. A. Vshivkov, and M. P. Fedoruk. *Numerical "Particle-in-Cell" Methods: Theory and Applications*. De Gruyter, 2002.

[81] O. Guéant, J.-M. Lasry, and P.-L. Lions. Mean field games and applications. In *Paris-Princeton Lectures on Mathematical Finance 2010*, Lecture Notes in Math., page 205266. 2011.

[82] V. Gyrya, L. Berlyand, I. Aranson, and D. A. Karpeev. A model of hydrodynamics interaction between swimming bacteria. *Bulletin of Mathematical Biology*, 72:148–173, 2010.

[83] M. Hauray. On Liouville transport equation with force field in BV_{loc}. *Comm. Partial Differential Equations*, 29(1-2):207–217, 2004.

[84] M. Hauray. Wasserstein distances for vortices approximation of Euler-type equations. *Math. Models Methods Appl. Sci.*, 19(8):1357–1384, 2009.

[85] M. Hauray. Mean field limit for the one dimensional vlasov-poisson equation. Séminaire Laurent Schwartz, École Polytechnique, 2013. arXiv:1309.2531.

[86] M. Hauray and P.-E. Jabin. N-particles approximation of the Vlasov equations with singular potential. *Arch. Ration. Mech. Anal.*, 183(3):489–524, 2007.

[87] M. Hauray and P.-E. Jabin. Particles approximations of vlasov equations with singular forces : Propagation of chaos. *To appear Ann. Sci. Ec. Norm. Super.*, 2014.

[88] M. Hauray and S. Mischler. On kac's chaos and related problems. *J. Funct. Anal.*, 266(10):6055–6157, 2014.

[89] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artifical Societies and Social Simulation (JASSS)*, 5, no. 3, 2002.

[90] M. A. Herrero and J. J. L. Velázquez. Chemotactic collapse for the Keller-Segel model. *J. Math. Biol.*, 35(2):177–194, 1996.

[91] E. Hewitt and L. Savage. Symmetric measures on cartesian products. *Trans. Amer. Math. Soc.*, 80:470–501, 1955.

[92] A. Honig, B. Niethammer, and F. Otto. On first-order corrections to the lsw theory i: infinite systems. *Journal of Statistical Physics*, 119:61–122, 2005.

[93] A. Honig, B. Niethammer, and F. Otto. On first-order corrections to the lsw theory ii: finite systems. *Journal of Statistical Physics*, 119:123–164, 2005.

[94] E. Horst. Global strong solutions of Vlasov's equation—necessary and sufficient conditions for their existence. In *Partial differential equations (Warsaw, 1984)*, volume 19 of *Banach Center Publ.*, pages 143–153. PWN, Warsaw, 1987.

[95] E. Horst. On the asymptotic growth of the solutions of the Vlasov-Poisson system. *Math. Methods Appl. Sci.*, 16(2):75–86, 1993.

[96] T. Y. Hou and J. Lowengrub. Convergence of the point vortex method for the 3-D Euler equations. *Comm. Pure Appl. Math.*, 43(8):965–981, 1990.

[97] T. Y. Hou, J. Lowengrub, and M. J. Shelley. The convergence of an exact desingularization for vortex methods. *SIAM J. Sci. Comput.*, 14(1):1–18, 1993.

[98] R. Illner and M. Pulvirenti. Global validity of the boltzmann equation for two- and three-dimensional rare gas in vacuum. *Comm. Math. Phys.*, 121:143–146, 1989.

[99] P. Jabin and F. Otto. Identification of the dilute regime in particle sedimentation. *Comm. Math. Phys.*, 250:415–432, 2004.

[100] P. Jabin and B. Perthame. Notes on mathematical problems on the dynamics of dispersed particles interacting through a fluid. In *Modelling in applied sciences, a kinetic theory approach*, Model. Simul. Sci. Eng. Technol., pages 111–147. Birkhauser Boston, 2000.

[101] W. Jäger and S. Luckhaus. On explosions of solutions to a system of partial differential equations modelling chemotaxis. *Trans. Amer. Math. Soc.*, 329(2):819–824, 1992.

[102] J. H. Jeans. On the theory of star-streaming and the structure of the universe. *Monthly Notices of the Royal Astronomical Society*, 76:70–84, 1915.

[103] V. I. Judovič. Non-stationary flows of an ideal incompressible fluid. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 3:1032–1066, 1963.

[104] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pages 171–197, Berkeley and Los Angeles, 1956. University of California Press.

[105] E. F. Keller and L. A. Segel. Initiation of slime mold aggregation viewed as an instability. *J. Theor. Biol.*, 26:399–415, 1970.

[106] M. K.-H. Kiessling. On the equilibrium statistical mechanics of isothermal classical self-gravitating matter. *J. Statist. Phys.*, 55(1-2):203–257, 1989.

[107] M. K.-H. Kiessling. Statistical mechanics of classical particles with logarithmic interactions. *Comm. Pure Appl. Math.*, 46(1):27–56, 1993.

[108] M. K.-H. Kiessling and H. Spohn. A note on the eigenvalue density of random matrices. *Comm. Math. Phys.*, 199(3):683–695, 1999.

[109] J. G. Kirkwood. The statistical mechanical theory of transport processes i. general theory. *The Journal of Chemical Physics*, 14 (3):180, 1946.

[110] J. G. Kirkwood. The statistical mechanical theory of transport processes i. transport in gases. *The Journal of Chemical Physics*, 15 (1):72, 1947.

[111] U. Krause. A discrete nonlinear and non-autonomous model of consensus formation. *Communications in difference equations*, pages 227–236, 2000.

[112] C. Lancellotti. On the fluctuations about the Vlasov limit for N-particle systems with mean-field interactions. *J. Stat. Phys.*, 136(4):643–665, 2009.

[113] L. Landau. On the vibrations of the electronic plasma. *Akad. Nauk SSSR. Zhurnal Eksper. Teoret. Fiz.*, 16:574–586, 1946.

[114] O. E. Lanford, III. Time evolution of large classical systems. In *Dynamical systems, theory and applications (Recontres, Battelle Res. Inst., Seattle, Wash., 1974)*, pages 1–111. Lecture Notes in Phys., Vol. 38. Springer, Berlin, 1975.

[115] J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.*, 2 (1):229–260, 2007.

[116] M. Lemou, F. Méhats, and P. Raphaël. Uniqueness of the critical mass blow up solution for the four dimensional gravitational Vlasov-Poisson system. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 24(5):825–833, 2007.

[117] A. Lenard. On Bogoliubov's kinetic equation for a spatially homogeneous plasma. *Ann. Physics*, 10:390–400, 1960.

[118] P.-L. Lions and S. Mas-Gallic. Une méthode particulaire déterministe pour des équations diffusives non linéaires. *C. R. Math. Acad. Sci. Paris*, 332:369–376.

[119] P.-L. Lions and B. Perthame. Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system. *Invent. Math.*, 105(2):415–430, 1991.

[120] G. Loeper. Uniqueness of the solution to the Vlasov-Poisson system with bounded density. *J. Math. Pures Appl. (9)*, 86(1):68–79, 2006.

[121] C. Marchioro and M. Pulvirenti. *Mathematical theory of incompressible nonviscous fluids*, volume 96 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1994.

[122] R. J. McCann. Stable rotating binary stars and fluid in a tube. *Houston J. Math.*, 32(2):603–631, 2006.

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

[124] J. Messer and H. Spohn. Statistical mechanics of the isothermal Lane-Emden equation. *J. Statist. Phys.*, 29(3):561–578, 1982.

[125] S. Mischler. Sur le programme de Kac concernant les limites de champ moyen. In *Seminaire: Equations aux Dérivées Partielles. 2009–2010*, Sémin. Équ. Dériv. Partielles, pages Exp. No. XXXIII, 19. École Polytech., Palaiseau, 2012.

[126] S. Mischler and C. Mouhot. Kac's Program in Kinetic Theory. *Invent. Math.*, 193(1):1–147, 2013.

[127] S. Mischler, C. Mouhot, and B. Wennberg. A new approach to quantitative chaos propagation for drift, diffusion and jump process. *Arxiv*, 2013.

[128] S. Motsch and E. Tadmor. A new model for self-organized dynamics and its flocking behavior. *J. Stat. Phys.*, 144 (5):923–947, 2011.

[129] C. Mouhot and C. Villani. On Landau damping. *Acta Math.*, 207(1):29–201, 2011.

[130] T. Nagai. Blow-up of radially symmetric solutions to a chemotaxis system. *Adv. Math. Sci. Appl.*, 5(2):581–601, 1995.

[131] H. Neunzert and J. Wick. The convergence of simulation methods in plasma physics. In *Mathematical methods of plasmaphysics (Oberwolfach, 1979)*, volume 20 of *Methoden Verfahren Math. Phys.*, pages 271–286. Lang, Frankfurt, 1980.

[132] H. Osada. Propagation of chaos for the two-dimensional Navier-Stokes equation. In *Probabilistic methods in mathematical physics (Katata/Kyoto, 1985)*, pages 303–334. Academic Press, Boston, MA, 1987.

[133] H. G. Othmer and A. Stevens. Aggregation, blowup, and collapse: the ABCs of taxis in reinforced random walks. *SIAM J. Appl. Math.*, 57(4):1044–1081, 1997.

[134] C. Pallard. Moment propagation for weak solutions to the Vlasov-Poisson system. *Comm. Partial Differential Equations*, 37(7):1273–1285, 2012.

[135] C. S. Patlak. Random walk with persistence and external bias. *Bull. Math. Biophys.*, 15:311–338, 1953.

[136] B. Perthame. *Transport equations in biology.* Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007.

[137] K. Pfaffelmoser. Global classical solutions of the Vlasov-Poisson system in three dimensions for general initial data. *J. Differential Equations*, 95(2):281–303, 1992.

[138] F. Planchon. An extension of the Beale-Kato-Majda criterion for the Euler equations. *Comm. Math. Phys.*, 232(2):319–326, 2003.

[139] M. Rascle and C. Ziti. Finite time blow-up in some models of chemotaxis. *J. Math. Biol.*, 33(4):388–414, 1995.

[140] N. Rougerie and S. Serfaty. Higher dimensional coulomb gases and renormalized energy functionals. arXiv:1307.2805, 2013.

[141] D. G. Saari. Improbability of collisions in Newtonian gravitational systems. II. *Trans. Amer. Math. Soc.*, 181:351–368, 1973.

[142] D. G. Saari. A global existence theorem for the four-body problem of Newtonian mechanics. *J. Differential Equations*, 26(1):80–111, 1977.

[143] J. Schaeffer. Global existence of smooth solutions to the Vlasov-Poisson system in three dimensions. *Comm. Partial Differential Equations*, 16(8-9):1313–1335, 1991.

[144] S. Schochet. The weak vorticity formulation of the 2-D Euler equations and concentration-cancellation. *Comm. Partial Differential Equations*, 20(5-6):1077–1104, 1995.

[145] S. Schochet. The point-vortex method for periodic weak solutions of the 2-D Euler equations. *Comm. Pure Appl. Math.*, 49(9):911–965, 1996.

[146] Y. Sone. *Molecular Gas Dynamics. Theory, Techniques, and Applications.* Birkhuser, Boston, 2007.

[147] H. Spohn. On the vlasov hierarchy. *Math. Methods Appl. Sci.*, 3 no. 4:445–455, 1981.

[148] H. Spohn. *Large scale dynamics of interacting particles.* Springer Verlag, New York, 1991.

[149] A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[150] C. M. Topaz, A. L. Bertozzi, and M. A. Lewis. A nonlocal continuum model for biological aggregation. *Bull. Math. Biol.*, 68(7):1601–1623, 2006.

[151] M. Trocheris. On the derivation of the one-dimensional Vlasov equation. *Transport Theory Statist. Phys.*, 15(5):597–628, 1986.

[152] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyā*, 19:23–26, 1958.

[153] T. Vicsek, A. Czirk, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75(6):1226–1229, 1995.

[154] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003.

[155] A. A. Vlasov. On vibration properties of electron gas. *J. Exp. Theor. Phys. (in Russian)*, 8 (3):291, 1938.

[156] A. A. Vlasov. The vibrational properties of an electron gas. *Sov. Phys. Usp.*, 10:721–733, 1968.

[157] S. Wollman. On the approximation of the Vlasov-Poisson system by particle methods. *SIAM J. Numer. Anal.*, 37(4):1369–1398 (electronic), 2000.

[158] H. Xia, H. Wang, and Z. Xuan. Opinion dynamics: A multidisciplinary review and perspective on future research. *International Journal of Knowledge and Systems Science (IJKSS)*, 2(4):72–91, 2011.

[159] Z. Xia. The existence of noncollision singularities in Newtonian systems. *Ann. of Math. (2)*, 135(3):411–468, 1992.

[160] V. I. Yudovich. Uniqueness theorem for the basic nonstationary problem in the dynamics of an ideal incompressible fluid. *Math. Res. Lett.*, 2(1):27–38, 1995.

[161] J. Yvon. La théorie statistique des fluides et l'équation d'état (in french). *Actual. Sci. Indust.*, 203, 1935.