From microscopic to macroscopic scale equations: mean field, hydrodynamic and graph limits

Thierry Paul, Emmanuel Trélat

From microscopic to macroscopic scale equations: mean field, hydrodynamic and graph limits

Thierry Paul Emmanuel Trélat*

Abstract

We elaborate on various ways to pass to the limit a given family of finite particle systems, either by mean field limit, deriving the Vlasov equation, or by hydrodynamic or graph limit, obtaining the Euler equation. We provide convergence estimates. We also show how to pass from Liouville to Vlasov or to Euler by taking adequate moments. Our results encompass and generalize a number of known results of the literature. As a surprising consequence of our analysis, we show that, under appropriate regularity assumptions, solutions of any quasilinear PDE can be approximated by the solutions of finite particle systems.

Contents

*Sorbonne Université, CNRS, Université Paris Cité, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France (thierry.paul@sorbonne-universite.fr, emmanuel.trelat@sorbonne-universite.fr).

1 Introduction

1.1 Setting

Multi-agent collective models have regained an increasing interest over the last years, due in particular to their connection with mean field and graph limit equations. At the microscopic scale, such models consist of considering particles or agents evolving according to the dynamics

$$\dot{\xi}_i(t) = \frac{1}{N}\sum_{j=1}^{N} G_{ij}^N(t, \xi_i(t), \xi_j(t)), \qquad i \in \{1, \dots, N\}, \tag{1}$$

for some (large) number of agents $N \in \mathbb{N}^*$ where, for every $i \in \{1, \dots, N\}$, $\xi_i(t) \in \mathbb{R}^d$ (for some $d \in \mathbb{N}^*$) stands for various parameters describing the behavior of the ith agent and $G_{ij}^N : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a mapping modeling the interaction between the ith and j^{th} agents.

Dynamics of the form (1) are used in a wide range of very different problems, ranging from the study of flocking and swarming in biology, of modeling traffic flows, to dynamics evolution in social sciences (see, e.g., [1, 3, 6, 13, 14, 15, 16, 27, 29], just to mention a few of a vast literature).

Among classes of multi-agent systems, we point out the so-called opinion systems that have the striking property of exhibiting features nowadays grouped under the common denomination of self-organization: their large-time asymptotic behavior shows consensus phenomena, namely an alignment of all values $\xi_i(t)$ to a single one. These models correspond to $G^N_{ij}(t,\xi_i,\xi_j) = \sigma_{ij}(\xi_j - \xi_i)$, i.e., their dynamics is

$$\dot{\xi}_i(t) = \frac{1}{N}\sum_{j=1}^{N} \sigma_{ij}(\xi_j(t) - \xi_i(t)) \tag{2}$$

where $(\sigma_{ij})_{1\leqslant i,j\leqslant N}$ is a N-by-N matrix whose spectral properties may cause the above-mentioned asymptotic behavior. We refer to the recent [4] for a large set of references concerning the two systems (1) and (2), where also the case of time-dependent matrices σ is treated.

The large N limit of systems (1), (2) has been extensively studied over the last years. In [33] the author shows how to pass to the continuum limit in nonlocally coupled dynamical networks by using the concept of *graph limit*. This concept has also been used recently in [21] to obtain discrete-to-continuum convergence results with error estimates in the Wasserstein distance. We also mention the recent articles [4, 7, 11]. In a nutshell, the graph limit allows one to pass to the limit from the general system of agents (1) to an integro-differential equation by interpreting the right-hand side of (1) as a Riemann sum. Then, obtaining the limit equation is seen as passing to the limit in a Riemann sum and thus obtaining a continuous integral. This is what has been done in [4, 7, 12] for the opinion propagation model $G(t,x,x',\xi,\xi') = \sigma(x,x')(\xi' - \xi)$, leading to the graph limit equation

$$\partial_t y(t,x) = \int_\Omega \sigma(x,x')(y(t,x') - y(t,x))\,dx'$$

where for instance $\sigma(i,j) = \sigma_{ij}$. This example is particularly paradigmatic of what we develop in the present paper.

Another important class of systems (1) concerns particles $\xi_i = (p_i, q_i) \in \mathbb{R}^d \times \mathbb{R}^d$, either Hamiltonian in which case G takes the form

$$G_H(t,\xi_i,\xi_j) = \begin{pmatrix} p_i \\ \nabla V(q_i - q_j) \end{pmatrix} \tag{3}$$

for some potential V, or of Cucker-Smale type in which case we have

$$G_{CS}(t,\xi_i,\xi_j) = \begin{pmatrix} p_i \\ F(|q_i - q_j|)(p_i - p_j) \end{pmatrix} \tag{4}$$

for some influence function F.

The main difference between general systems (1) the particular systems (3) and (4) is that for the latter the mapping G_{ij} does not depend on i and j, that is on the "names" of the agents. A consequence is that the associated evolution equation preserves the *indistinguishability* of the particles, a feature often consider as fundamental for the large N limit of particle systems. One of the objectives of the present article is to show how to extend the standard mean field methods to the non-indistinguishable setting, simply by endowing to the index i the status of a parameter, treated as a new state variable of zero dynamics.

The systematic study of large N limit of particle systems has a long and glorious history, starting with Hartree (see [28]) in the late 20's for quantum systems, and then Vlasov in the 40's

(see [46]) who derived the eponymous *kinetic equation* (here, we present it in a form suitable for our purposes), called *Vlasov equation*,

$$\partial_t \mu_t + \mathrm{div}_\xi(\mathcal{X}[\mu_t]\mu_t) = 0 \qquad \text{with} \qquad \mathcal{X}[\mu] = \int G(\xi, \xi')\, d\mu(\xi'). \tag{5}$$

There are two classical ways for deriving (5). A first consists of using the concept of empirical measure $\mu_t^E(\xi) = \frac{1}{N}\sum_{i=1}^N \delta(\xi - \xi_i(t))$, which is a solution for (5), and then of taking the *mean field limit*. A second consists of using marginals of the solution of the *Liouville equation* associated to the particle system, namely the equation satisfied by the pushforward of probability measures on $\mathbb{R}^{2Nd}$ under the flow generated by the particle system. In the latter case one shows that the first marginal of this pushforward, which is a probability measure on $\mathbb{R}^{2d}$, satisfies at the limit $N \to +\infty$ the Vlasov equation (5). The last step of this process, called the *hydrodynamic limit*, starts from the observation that (5) preserves the stucture $\mu_t(\xi) = \mu_t(q,p) = \nu(t,q)\delta(p - y(t,q))$ leading to the so-called *Euler* system of equations satisfied by the pair (ν, y).

One of the main steps in the developments of the present article is to highlight that, after having derived the Vlasov equation associated to (1) thanks to the trick consisting of parametrizing the status of the index i as already mentioned, the associated Euler equation (not a system anymore because the extra dynamical variables i remain at rest and thus give no kinetic part in the Euler system) coincides with the *graph limit* equation associated to (1) – a nontrivial fact, even at the conceptual level as discussed in Section 4.2 of the paper.

This article is devoted to unifying and generalizing, to some extent, the classical ways to pass to the limit in families of particle systems. The mean field limit, even for distinguishable particles, leads to the Vlasov equation. The hydrodynamic limit leads to the Euler equation. The Liouville equation is a lift of the particle system in a space of probability measures. We analyze in detail the various relationships between particle system, Vlasov, Liouville and Euler, showing how to pass from one to another and deriving, under appropriate assumptions, some convergence estimates. While some of the results are classical (or straightforward extensions of known results), most of them are new and we hope that the overall study may serve to unify different viewpoints.

The last part of our paper deals somehow with the inverse path: given a general partial differential equation (PDE), is is possible to construct explicitly an agent system of the form (1) such that the corresponding graph limit equation coincides with the given PDE we started with? Surprisingly, this happens to be true in a very general setting and this is a consequence of the analysis done in the paper.

The question of whether some classes of PDEs are a "natural" limit of particle systems is classical in fluid mechanics and certainly dates back to Euler: it is classical that the Euler fluid equation can be seen, at least formally, as the limit of evolving "particles of fluids". This has been formalized in the famous article [2] where Arnol'd interpreted the Euler equation as a geodesic equation in the space of diffeomorphisms, leading to a number of subsequent studies; we refer to [5] (see also the references therein) for a survey on how to "cook up" appropriate groups of diffeomorphisms (and thus, of particle systems) to generate classes of fluid PDEs, like Euler, Camassa-Holm, etc. We also refer to [25] for a recent survey on microscopic, mesoscopic and macroscopic scales for fluid dynamics.

But it is much less classical to show that other, more general PDEs can as well be obtained by passing to the limit in some particle systems. For transport equations, the topic has been extensively studied in [17, 18, 19]. Recently, thanks to the concept of graph limit elaborated in [33], it has been possible to show that heat-like equations can as well be obtained as limits of particle systems (see also [4, 7, 11, 21]). In [22], the authors provide a rigorous derivation from the kinetic Cucker-Smale model to the macroscopic pressureless Euler system by hydrodynamic limit, using entropy methods and deriving error estimates.

Actually, during the *Leçons Jacques-Louis Lions* given in our laboratory in the fall 2021 by Dejan Slepcev, we were intrigued by his way of deriving heat-like equations from unusual particle systems, by taking not only the limit as the number N of agents tends to $+\infty$, but also another parameter ε tends to 0, at some precise scaling (see [21]). The role of ε is to smoothen the dynamics. His striking exposition has been for us a great source of inspiration and has motivated the last part of the present article.

In this last part, we provide for a large range of quasilinear PDEs a natural and constructive way for associating an agent system to them. Shortly, as a particular case of our analysis, considering a general PDE

$$\partial_t y(t,x) = \sum_{|\alpha|\leqslant p} a_\alpha(t,x,y(t,x))\partial_x^\alpha y(t,x) = A(t,x,y(t,x))y(t,x), \tag{6}$$

we show that (6) is the graph limit of (for example) the particle system (1) with

$$G_{ij}(t,\xi,\xi') = G_\varepsilon(t,i,j,\xi,\xi') = \xi' \sum_{|\alpha|\leqslant p} a_\alpha(t,x,\xi)\partial_{x'}^\alpha \frac{e^{-\frac{(x-x')^2}{2\varepsilon}}}{(\pi\varepsilon)^{\frac{1}{2}}}$$

in the limit $N \gg \varepsilon^{-1} \to +\infty$, with some appropriate scalings. We establish convergence estimates in Wasserstein distance in general, and in L^2 norm under an additional (but general) semigroup assumption.

Structure of the article. Section 2 is devoted to studying the passage from microscopic to mesoscopic scale: in other words we show how to pass "from particle to Vlasov" by mean field limit. In Theorem 1, we establish existence, uniqueness and stability properties for the Vlasov equation (5) for distinguishable particles. We recall the classical Lagrangian and Eulerian viewpoints. For the latter, we elaborate on the Liouville equation associated with the particle system. Theorems 2 and 3 are devoted to establish Wasserstein estimates quantifying the discrepancy, as $N \to +\infty$, between the first marginal of the solution of the Liouville equation and the solution of the Vlasov equation; in other words, these results show how Vlasov can be recovered from Liouville by taking marginals and passing to the limit (propagation of chaos).

Section 3 is devoted to studying the passage from mesoscopic to macroscopic scale: in other words we show how to pass "from Vlasov to Euler" by hydrodynamic limit, mainly consisting of taking the moment of order 1. Proposition 4 in that section is concerned with the well known monokinetic approach, but we also investigate the moment of order 2, yielding some consensus results.

Section 4 is devoted to studying the passage from microscopic to macroscopic scale: in other words we show how to pass "from particle to Euler" by graph limit. Theorems 4 and 5 quantify some convergence estimates in L^∞ norm as $N \to +\infty$. We also discuss how Euler can be recovered from Liouville by taking adequate moments.

Section 5 provides a synthetic summarize of all relationships that we have unraveled. In particular, Figure 1 illustrates the various two-ways passages between particle (microscopic) systems, the Liouville (probabilistic) equation, the Vlasov (mesoscopic, mean field) equation, and the Euler (macroscopic, graph limit) equation. This section can even be read as a motivating preliminary before going ahead.

Finally, as announced, as a surprising byproduct built on the previous developments, we show in Section 6 that general quasilinear PDEs can be obtained by passing to the limit in explicit particle systems, thanks to two asymptotic parameters.

In order to state all subsequent results, we recall in Section 1.2 hereafter some notations and concepts that we use throughout, in particular the Wasserstein distance and another distance

obtained by disintegration of measures, and the concept of tagged partition that is classically used in Riemann integration theory.

We gather in Appendix A a number of useful results on the Wasserstein distance, empirical and (so-called semi-empirical measures. Appendix B is devoted to proving some of the main theorems.

1.2 General notations

Let $d \in \mathbb{N}^*$ be arbitrary. We denote by $\| \ \|$ the Euclidean norm in $\mathbb{R}^d$. Let (Ω, d_Ω) be a complete metric space.

Hölder and Lipschitz mappings. Let U be a subset of $\mathbb{R}^d$. Let $p \in \mathbb{N}^*$. Given any $\alpha \in (0,1]$, we denote by $\mathscr{C}^{0,\alpha}(U, \mathbb{R}^p)$ the set of all continuous mappings $g \in \mathscr{C}^0(U, \mathbb{R}^d)$ that are *α-Hölder continuous*, meaning that

$$\mathrm{Hol}_\alpha(g) = \sup_{\substack{x,x' \in U \\ x \neq x'}} \frac{\|g(x) - g(x')\|}{\|x - x'\|^\alpha} < +\infty.$$

When U is compact, $\mathscr{C}^{0,\alpha}(U, \mathbb{R}^p)$ is a Banach space endowed with the norm

$$\|g\|_{\mathscr{C}^{0,\alpha}(U,\mathbb{R}^p)} = \max_{x \in U} \|g(x)\| + \mathrm{Hol}_\alpha(g).$$

When $\alpha = 1$, we speak of a Lipschitz mapping and we denote $\mathrm{Lip}(g) = \mathrm{Hol}_1(g)$. We do not consider exponents $\alpha > 1$ since any $g \in \mathscr{C}^{0,\alpha}(U, \mathbb{R}^p)$ with $\alpha > 1$ must be constant if U has a nonempty interior. When $p = 1$ and $\alpha = 1$, we denote $\mathrm{Lip}(U) = \mathscr{C}^{0,1}(U, \mathbb{R})$.

We define similarly $\mathscr{C}^{0,\alpha}(U, \mathbb{R}^p)$ when U is a subset of Ω, replacing $\|x - x'\|$ by $d_\Omega(x, x')$.

Probability Radon measures. Given any $d \in \mathbb{N}^*$, we denote by $\mathcal{P}(\mathbb{R}^d)$ the set of nonnegative probability Radon measures on $\mathbb{R}^d$. We also consider $\mathcal{P}_c(\mathbb{R}^d)$, $\mathcal{P}^{ac}(\mathbb{R}^d)$, where the subscript c means "with compact support" and the superscript ac means "absolutely continuous with respect to the Lebesgue measure", and for every $p \geqslant 1$ the set $\mathcal{P}_p(\mathbb{R}^d)$ stands for the set of all $\mu \in \mathcal{P}(\mathbb{R}^d)$ that have a finite moment of order p, i.e., $\int_{\mathbb{R}^d} \|x\|^p \, d\mu(x) < +\infty$. Given any Borel mapping $\phi : \mathbb{R}^d \to \mathbb{R}^p$ and given any $\mu \in \mathcal{P}(\mathbb{R}^d)$, the image (or pushforward) of μ under ϕ is $\phi_* \mu = \mu \circ \phi^{-1}$. Given any $\mu \in \mathcal{P}_c(\mathbb{R}^d)$, we define $\|\mathrm{supp}(\mu)\|_\infty = \max\{\|x\| \ \mid \ x \in \mathrm{supp}(\mu)\}$.

We denote by $\mathscr{C}^0(\mathbb{R}^d)$ (resp., $\mathscr{C}^\infty(\mathbb{R}^d)$) the set of continuous (resp., smooth) functions on $\mathbb{R}^d$ and by $\mathscr{C}^0_c(\mathbb{R}^d)$ (resp., $\mathscr{C}^\infty_c(\mathbb{R}^d)$) the set of continuous (resp., smooth) functions of compact support on $\mathbb{R}^d$. We recall that the topological dual $(\mathscr{C}^0_c(\mathbb{R}^d))'$ (resp., $(\mathscr{C}^0(\mathbb{R}^d))'$) is the set of all Radon measures on $\mathbb{R}^d$ (resp., with compact support). Endowed with the total variation norm $\| \ \|_{TV}$ which is the dual norm, it is a Banach space.

We use the same notations for Radon measures on $\Omega \times \mathbb{R}^d$. Denoting $(x, \xi) \in \Omega \times \mathbb{R}^d$, given any $\mu \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ of marginal $\nu \in \mathcal{P}_c(\Omega)$ on Ω, we define

$$\|\mathrm{supp}(\mu)\|_\infty = \min_{x_0 \in \mathrm{supp}(\nu)} \max\{d_\Omega(x_0, x) + \|\xi\| \ \mid \ (x, \xi) \in \mathrm{supp}(\mu)\}.$$

Wasserstein distance. Given any $p \geqslant 1$, the Wasserstein distance $W_p(\mu_1, \mu_2)$ of order p between two probability measures $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^d)$ is defined as the infimum of the Monge-Kantorovich cost $\int_{\mathbb{R}^{2d}} \|x - y\|^p \, d\Pi(x, y)$ over the set of probability measures $\Pi \in \mathcal{P}(\mathbb{R}^{2d})$ coupling μ_1 with μ_2, i.e., whose marginals on the two copies of $\mathbb{R}^d$ are μ_1 and μ_2:

$$W_p(\mu_1, \mu_2) = \inf \left\{ \int_{\mathbb{R}^{2d}} \|x - y\|^p \, d\Pi(x, y) \ \mid \ \Pi \in \mathcal{P}(\mathbb{R}^{2d}), \ (\pi_1)_* \Pi = \mu_1, \ (\pi_2)_* \Pi = \mu_2 \right\} \tag{7}$$

where $\pi_1 : \mathbb{R}^{2d} \to \mathbb{R}^d$ and $\pi_2 : \mathbb{R}^{2d} \to \mathbb{R}^d$ are the canonical projections defined by $\pi_1(x,y) = x$ and $\pi_2(x,y) = y$ for all $(x,y) \in \mathbb{R}^{2d}$. Then, W_p is a distance on $\mathcal{P}_p(\mathbb{R}^d)$, which metrizes the weak convergence in $\mathcal{P}_p(\mathbb{R}^d)$ in the following sense: given $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ and given a sequence $(\mu_k)_{k\in\mathbb{N}^*}$ in $\mathcal{P}_p(\mathbb{R}^d)$, we have $W_p(\mu_k,\mu) \to 0$ as $k \to +\infty$ if and only if $\int_{\mathbb{R}^d} f\,d\mu_k \to \int_{\mathbb{R}^d} f\,d\mu$ for every continuous bounded function f on $\mathbb{R}^d$ and $\int_{\mathbb{R}^d} \|x\|^p\,d\mu_k(x) \to \int_{\mathbb{R}^d} \|x\|^p\,d\mu(x)$ as $k \to +\infty$ (see [41, Chap. 5.2] or [45, Theorem 6.9]), if and only if $\int_{\mathbb{R}^d} f\,d\mu_k \to \int_{\mathbb{R}^d} f\,d\mu$ for every continuous function f on $\mathbb{R}^d$ such that $|f(x)| \leqslant C(1+\|x_0-x\|^p)$ for every $x \in \mathbb{R}^d$, for some $C>0$ and some $x_0 \in \mathbb{R}^d$ (see [44, Theorem 7.12]). It can be noted that, given any compact subset $\Omega \subset \mathbb{R}^d$, we have

$$W_1(\mu_1,\mu_2) \leqslant W_p(\mu_1,\mu_2) \leqslant \mathrm{diam}(\Omega)^{1-1/p} W_1(\mu_1,\mu_2)^{1/p} \tag{8}$$

for all probability measures μ_1 and μ_2 of compact support contained in Ω (see [41, Chap. 5]).

For $p=1$, the duality formula for the Kantorovich-Rubinstein distance (see, e.g., [45]) gives the equivalent definition

$$W_1(\mu_1,\mu_2) = \sup\left\{ \int_{\mathbb{R}^d} f\,d(\mu_1-\mu_2) \;\;|\;\; f \in \mathrm{Lip}(\mathbb{R}^d),\ \mathrm{Lip}(f) \leqslant 1 \right\}, \tag{9}$$

valid for all $\mu_1,\mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$.

There are many other distances (for instance, the bounded Lipschitz distance) but W_1 is the most useful in the study that we develop here, due in particular to the fact that, in the definition of the mean field $\mathcal{X}[\mu]$ that we are going to consider, the "potential" G may be linear in ξ at infinity.

Disintegration. In this paper, we are going to consider measures on $\Omega \times \mathbb{R}^d$, where (Ω, d_Ω) is a complete metric space. Denoting by $\pi : \Omega \times \mathbb{R}^d \to \Omega$ the canonical projection, given any $\mu \in \mathcal{P}(\Omega \times \mathbb{R}^d)$, we will always denote by ν the nonnegative probability Radon measure on Ω defined as the image (pushforward) of μ under π,

$$\nu = \pi_*\mu = \mu \circ \pi^{-1}, \tag{10}$$

that is also the marginal of μ on Ω. Note that, since π is continuous, $\mathrm{supp}(\nu) = \overline{\pi(\mathrm{supp}(\mu))}$. By disintegration of μ with respect to ν, there exists a family $(\mu_x)_{x\in\Omega}$ of probability Radon measures on $\mathbb{R}^d$ (uniquely defined ν-almost everywhere) such that $\mu = \int_\Omega \mu_x\,d\nu(x)$, i.e.,

$$\int_{\Omega\times\mathbb{R}^d} h(x,\xi)\,d\mu(x,\xi) = \int_\Omega \int_{\mathbb{R}^d} h(x,\xi)\,d\mu_x(\xi)\,d\nu(x)$$

for every Borel measurable function $h : \Omega\times\mathbb{R}^d \to [0,+\infty)$ (see, e.e., [10]). Moreover, we set $\mu_x = 0$ whenever $x \in \Omega \setminus \mathrm{supp}(\nu)$.

When Ω is a smooth manifold, if $\mu \in \mathcal{P}^{ac}(\Omega \times \mathbb{R}^d)$ with a density $f \in L^1(\Omega \times \mathbb{R}^d)$, i.e., $\frac{d\mu}{dx\,d\xi}(x,\xi) = f(x,\xi)$, then ν is absolutely continuous, of density $\frac{d\nu}{dx}(x) = \int_{\mathbb{R}^d} f(x,\xi)\,d\xi$, and for ν-almost every $x \in \Omega$ the probability measure μ_x has the density $\frac{d\mu_x}{d\xi}(\xi) = \frac{f(x,\xi)}{\int_{\mathbb{R}^d} f(x,\xi')\,d\xi'}$.

Given any $\mu^1,\mu^2 \in \mathcal{P}_1(\Omega \times \mathbb{R}^d)$ having the same marginal ν on Ω, we define

$$L^1_\nu W_1(\mu^1,\mu^2) = \int_\Omega W_1(\mu^1_x,\mu^2_x)\,d\nu(x). \tag{11}$$

Obviously, $L^1_\nu W_1$ is a distance on the subset denoted $\mathcal{P}^\nu_1(\Omega\times\mathbb{R}^d)$ of elements of $\mathcal{P}_1(\Omega\times\mathbb{R}^d)$ having the same marginal ν, and we have $W_1(\mu^1,\mu^2) \leqslant L^1_\nu W_1(\mu^1,\mu^2)$ for all $\mu_1,\mu_2 \in \mathcal{P}^\nu_1(\Omega \times \mathbb{R}^d)$.[1]

[1] Indeed, $\int_{\Omega\times\mathbb{R}^d} f\,d(\mu^1-\mu^2) = \int_\Omega \int_{\mathbb{R}^d} f(x,\xi)\,d(\mu^1_x-\mu^2_x)\,d\nu(x) \leqslant \int_\Omega \mathrm{Lip}(f(x,\cdot))\,W_1(\mu^1_x,\mu^2_x)\,d\nu(x)$ for every $f \in \mathrm{Lip}(\Omega\times\mathbb{R}^d)$, and if $\mathrm{Lip}(f) \leqslant 1$ then $\mathrm{Lip}(f(x,\cdot)) \leqslant 1$ for every $x \in \Omega$. Then, take the supremum over all f.

Tagged partitions. Let $\nu \in \mathcal{P}(\Omega)$. Given any $N \in \mathbb{N}^*$, we say that $(\mathcal{A}^N, X^N)$ is a *tagged partition* of Ω associated with ν if $\mathcal{A}^N = (\Omega_1, \dots, \Omega_N)$ is a N-tuple of disjoint subsets $\Omega_i \subset \Omega$ such that

$$\Omega = \bigcup_{i=1}^{N} \Omega_i \qquad \text{with} \qquad \nu(\Omega_i) = \frac{1}{N} \quad \text{and} \quad \operatorname{diam}(\Omega_i) \leqslant \frac{C_\Omega}{N} \qquad \forall i \in \{1, \dots, N\}, \tag{12}$$

for some constant $C_\Omega > 0$ not depending on N, and $X^N = (x_1, \dots, x_N)$ is a N-tuple of points $x_i \in \Omega_i$ (we do not put any superscript N to Ω_i and x_i for readability). Here, $\operatorname{diam}(\Omega_i)$ is the supremum of all $d_\Omega(x, x')$ over all possible $x, x' \in \Omega_i$.

Tagged partitions always exist, for any $N \in \mathbb{N}^*$, when Ω is a compact finite-dimensional smooth manifold having a boundary or not and ν is a Lebesgue measure on Ω. A typical example is when $\Omega = [0, 1]$: in this case we can always choose $\Omega_i = [a_i, a_{i+1})$ for some subdivision $0 = a_1 < a_2 < \cdots < a_{N+1} = 1$ satisfying (12); when $d\nu(x) = dx$, a natural choice is $a_i = \frac{i-1}{N}$, and $x_i = a_i$ or $\frac{a_i + a_{i+1}}{2}$, for every $i \in \{1, \dots, N\}$ (and then $C_\Omega = 1$ in this case). When Ω is a compact domain of $\mathbb{R}^n$, a family of tagged partitions is obtained by considering a family of meshes, as done in numerical analysis.

The concept of tagged partition is classically used in Riemann (and more generally, Henstock-Kurzweil) integration theory. We refer to [24] for (much more) general results. A real-valued function f on Ω, of compact support, is said to be ν-Riemann integrable if it is bounded, ν-measurable, and for any family $(\mathcal{A}^N, X^N)_{N \in \mathbb{N}^*}$ of tagged partitions, with $\mathcal{A}^N = (\Omega_1, \dots, \Omega_N)$ and $X^N = (x_1, \dots, x_N)$, we have

$$\sum_{i=1}^{N} \int_{\Omega_i} |f(x) - f(x_i)| \, d\nu(x) = \mathrm{o}(1) \tag{13}$$

and thus

$$\int_\Omega f \, d\nu = \frac{1}{N} \sum_{i=1}^{N} f(x_i) + \mathrm{o}(1) \tag{14}$$

as $N \to +\infty$. A function f of compact support on Ω is ν-Riemann integrable if and only if f is bounded and continuous ν-almost everywhere on Ω.

2 From microscopic to mesoscopic scale

("from particle to Vlasov", mean field limit)

2.1 Particle system

Let $d \in \mathbb{N}^*$ be fixed. At the microscopic level, given any $N \in \mathbb{N}^*$, we consider a system of N interacting particles $\xi_i(t) \in \mathbb{R}^d$, called the *particle system*, of dynamics

$$\boxed{\dot{\xi}_i(t) = \frac{1}{N} \sum_{j=1}^{N} G_{ij}^N(t, \xi_i(t), \xi_j(t)), \qquad i = 1, \dots, N} \tag{15}$$

where $G_{ij}^N : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ stands for the interaction between the particles i and j. The usual case, widely treated in the existing literature, is when the interaction mapping is the same for all pairs of particles and moreover does not depend on N, i.e., $G_{ij}^N = G$. We show here that there is no difficulty to treat the more general situation where the interactions depend on the agents. Above, in (15), G_{ij}^N depends on i, j, N.

Throughout the paper, we make the following assumptions $(\mathbf{G_1})$ and $(\mathbf{G_2})$:

($\mathbf{G_1}$) There exist a complete metric space (Ω, d_Ω) and a *continuous* mapping

$$\begin{array}{rcl} G : \mathbb{R} \times \Omega \times \Omega \times \mathbb{R}^d \times \mathbb{R}^d & \to & \mathbb{R}^d \\ (t, x, x', \xi, \xi') & \mapsto & G(t, x, x', \xi, \xi'), \end{array}$$

locally Lipschitz with respect to (ξ, ξ') (uniformly with respect to (t, x, x') on any compact), and, for every $N \in \mathbb{N}^*$, there exist distinct points $x_1, \ldots, x_N$ in Ω such that

$$G(t, x_i, x_j, \xi, \xi') = G_{ij}^N(t, \xi, \xi') \qquad \forall t \in \mathbb{R} \qquad \forall \xi, \xi' \in \mathbb{R}^d \qquad \forall i, j \in \{1, \ldots, N\}.$$

Under Assumption ($\mathbf{G_1}$), the particle system (15) is equivalently written as

$$\boxed{\begin{array}{l} \dot{x}_i(t) = 0 \\ \dot{\xi}_i(t) = \dfrac{1}{N} \displaystyle\sum_{j=1}^N G(t, x_i, x_j, \xi_i(t), \xi_j(t)), \qquad i = 1, \ldots, N \end{array}} \tag{16}$$

The variables $x_i \in \Omega$ are parameters, and a usual way to treat parameters in differential equations is to treat them as state variables whose dynamics are zero, whence the dynamics $\dot{x}_i(t) = 0$ above.

($\mathbf{G_2}$) Given any initial condition, the system (16) has a unique global solution on $\mathbb{R}$, i.e., there is no blow-up in finite time.

Comments on Assumptions ($\mathbf{G_1}$) and ($\mathbf{G_2}$). Assumption ($\mathbf{G_1}$) is a kind a continuous interpolation. The continuity assumption already encodes the idea of the existence of a limit system as $N \to +\infty$. Indeed, if G were not required to be continuous, then completely different systems (15) could be considered as N varies and then obviously no limit for large N could exist.

The "limit" mapping G is assumed to be locally Lipschitz, which ensures the application of the Cauchy-Lipschitz theorem and thus the local existence and uniqueness of solutions of (16). Note that Assumption ($\mathbf{G_1}$) implies that the Lipschitz constants of the mappings G_{ij}^N are uniformly bounded (with respect to i, j, N) on any compact. Assumption ($\mathbf{G_2}$) requires a bit more, by assuming that the maximal domain of definition of the solutions of (16) is equal to the whole real line. For example, this is true if G is globally Lipschitz with respect to (ξ, ξ'). But Assumption ($\mathbf{G_2}$) is weaker and allows for superlinearities provided that there is no blow-up in finite time. This is the case for many classes of Hamiltonian systems, or when the particle system (16) is stable in the sense of Lyapunov. All in all, this non blow-up assumption implies that, given a compact set of initial conditions and given any $T > 0$, the set of all corresponding states, solutions of (16), is contained in a compact set on $[0, T]$.

In Assumption ($\mathbf{G_1}$), the complete metric space Ω used for the parameters x_i is arbitrary. For instance we can take $\Omega = [0, 1]$, but we allow for more general sets, in view of deriving on Ω some interesting classes of PDEs.

The choice of the possible values of the x_i is not imposed in Assumption ($\mathbf{G_1}$). If one wishes moreover to fix some precise points x_i, such as the natural one $x_i = \frac{i}{N}$ when $\Omega = [0, 1]$, often used in numerical analysis, then there are compatibility conditions on the mappings G_{ij}^N.

Given any $X = (x_1, \ldots, x_N) \in \Omega^N$, the system (16) can also be written in the form

$$\boxed{\dot{\Xi}(t) = Y(t, X, \Xi(t))} \tag{17}$$

where $\Xi(t) = (\xi_1(t), \ldots, \xi_N(t))$ and

$$Y(t, X, \cdot) = (Y_1(t, X, \cdot), \ldots, Y_N(t, X, \cdot))^\top \tag{18}$$

is a time-dependent vector field defined on $(\mathbb{R}^d)^N$, depending on the parameter X, with

$$Y_i(t, X, \Xi) = \frac{1}{N} \sum_{j=1}^{N} G(t, x_i, x_j, \xi_i, \xi_j) \qquad \forall i \in \{1, \dots, N\}$$

where $\Xi = (\xi_1, \dots, \xi_N)$. We denote by $(\Phi(t, X, \cdot))_{t \in \mathbb{R}}$ the flow of diffeomorphisms of $\mathbb{R}^{dN}$ generated by the time-dependent vector field $Y(t, X, \cdot)$: this flow is parametrized by $X \in \Omega^N$. We have $\Phi(t, X, \Xi(0)) = \Xi(t)$ where $t \mapsto \Xi(t) = (\xi_1(t), \dots, \xi_N(t))$ is solution of (16), or equivalently, of (17), with $X = (x_1, \dots, x_N)$. We also denote $\Phi_{t,X} = \Phi(t, X, \cdot)$ the diffeomorphism of $\mathbb{R}^{dN}$ depending on (t, X), and $\Phi_t = \Phi(t, \cdot, \cdot)$ the diffeomorphism of $\Omega^N \times \mathbb{R}^{dN}$ depending on t and letting X invariant.

Some examples covered by this general framework are in order.

Example 1. *Consider the general linear Hegselmann and Krause first-order consensus system (see [29]), modeling for instance the propagation of opinions (studied in [12]), of dynamics*

$$\dot{\xi}_i(t) = \frac{1}{N} \sum_{j=1}^{N} \sigma_{ij}^N (\xi_j(t) - \xi_i(t)), \qquad i = 1, \dots, N, \tag{19}$$

with $\sigma_{ij}^N \geqslant 0$ (interaction coefficients, not necessarily symmetric). Assumption $(\mathbf{G_1})$ requires that there exists Ω and a continuous function σ on Ω^2 such that, for every $N \in \mathbb{N}^*$, there exist distinct points $x_1, \dots, x_N$ in Ω such that $\sigma(x_i, x_j) = \sigma_{ij}^N$. We have then $G(t, x, x', \xi, \xi') = \sigma(x, x')(\xi' - \xi)$ for all $(t, x, x', \xi, \xi') \in \mathbb{R} \times \Omega^2 \times (\mathbb{R}^d)^2$, and Assumption $(\mathbf{G_2})$ is satisfied.

Example 2. *Consider the Cucker and Smale model (see [16]) where, setting $d = 2r$ and $\xi = (q, p)^\top \in \mathbb{R}^r \times \mathbb{R}^r$, the dynamics is written as*

$$\dot{q}_i(t) = p_i(t), \qquad \dot{p}_i(t) = \frac{1}{N} \sum_{j=1}^{N} a(\|q_i(t) - q_j(t)\|)(p_j(t) - p_i(t)), \qquad i = 1, \dots, N,$$

for some potential function a on $\mathbb{R}$ of class C^1. Assumptions $(\mathbf{G_1})$ and $(\mathbf{G_2})$ are satisfied with $G = (G_q, G_p)^\top$ where $G_q(t, x, x', \xi, \xi') = p$ and $G_p(t, x, x', \xi, \xi') = a(\|q - q'\|)(\xi' - \xi)$.

Many variants of that model are covered by our framework, for instance the potential may depend on i and j, and we could add other terms to the dynamics of p_i, standing for instance for self-propulsion and attraction-repulsion forces (see, e.g., [13]).

Example 3. *Still with $d = 2r$ and $\xi = (q, p)^\top \in \mathbb{R}^r \times \mathbb{R}^r$, given any $N \in \mathbb{N}^*$, consider the Hamiltonian*

$$H^N(q_1, p_1, \dots, q_N, p_N) = \sum_{i=1}^{N} h_i^N(q_i, p_i) + \frac{1}{N} \sum_{i,j=1}^{N} h_{ij}^N(q_i, p_i, q_j, p_j)$$

for some C^1 functions h_i^N and h_{ij}^N. The corresponding Hamiltonian system of N particles, given by $\dot{q}_i = \frac{\partial H}{\partial p_i}$, $\dot{p}_i = -\frac{\partial H}{\partial q_i}$ for $i = 1, \dots, N$, can be written as (16) with

$$G_{ij}^N(t, \xi, \xi') = \begin{pmatrix} \partial_2 h_i^N(q, p) + \delta_{ij}\, \partial_2 h_{ij}^N(q, p, q', p') + \partial_4 h_{ij}^N(q, p, q', p') \\ -\partial_1 h_i^N(q, p) - \delta_{ij}\, \partial_1 h_{ij}^N(q, p, q', p') - \partial_3 h_{ij}^N(q, p, q', p') \end{pmatrix}$$

where ∂_k denotes the partial derivative with respect to the kth-variable, and where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Assumptions $(\mathbf{G_1})$ *and* $(\mathbf{G_2})$ *require at least that the Hamiltonians* h_i^N *and* h_{ij}^N *be uniformly (wrt* i, j, N*) locally Lipschitz. Actually, most of classical Hamiltonian systems of* N *particles are written as above with Hamiltonians not depending on* i, j, N*. Note anyway that the above Hamiltonian* H^N *involves sums of "single" and of "pairwise" Hamiltonians, but not of "triplewise" or more.*

When N becomes larger and larger, we want to pass to the limit in some sense and replace the set of particles with a nonnegative Radon measure. In this objective, two classical viewpoints are the Lagrangian and the Eulerian one.

The *Lagrangian* viewpoint consists of keeping the trajectories of (16), taking the mean field limit by embedding trajectories with an empirical measure on $\Omega \times \mathbb{R}^d$ to solutions of the *Vlasov equation* (or *continuity equation*) in $\Omega \times \mathbb{R}^d$. This is done in Section 2.2.

The *Eulerian* viewpoint consists of using the flow of diffeomorphisms of Y to propagate an initial measure on $\Omega^N \times \mathbb{R}^{dN}$, thus obtaining the *Liouville equation* in $\Omega^N \times \mathbb{R}^{dN}$. This is done in Section 2.3.

Hereafter, we elaborate in detail on these procedures and we then show in Section 2.4 how to recover the Vlasov equation from the Liouville equation by taking adequate marginals, obtaining convergence estimates in Wasserstein distance.

2.2 Lagrangian viewpoint: mean field limit and Vlasov equation

Within the Lagrangian viewpoint, the N particles at time t are embedded as Dirac masses to the space of Radon measures, and their corresponding average, the empirical measure, converges by the so-called mean field limit procedure, as $N \to +\infty$, to a probability Radon measure $\mu(t)$ on $\Omega \times \mathbb{R}^d$ satisfying the Vlasov equation. When $\mu(t)$ is absolutely continuous with respect to a Lebesgue measure, its density $\frac{d\mu(t)}{d\xi} = f(t, x, \xi)$ is the density of particles at time t.

In the existing literature, it is often said that, in order to pass to the mean field limit, it is necessary that the particles be indistinguishable, and that the interaction mapping G must be the same for all pairs of particles. Hereafter, we show that such an assumption is absolutely unnecessary and that there is no difficulty of considering interactions depending on agents.

2.2.1 Mean field

Given any $\mu \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$, we define ν by (10) (the marginal of μ on Ω), and we define the *mean field*, also called *interaction kernel*, as the non-local time-dependent one-parameter (the parameter is $x \in \Omega$) vector field on $\mathbb{R}^d$ given by

$$\boxed{\begin{aligned}\mathcal{X}[\mu](t, x, \xi) &= \int_{\Omega \times \mathbb{R}^d} G(t, x, x', \xi, \xi')\, d\mu(x', \xi') \\ &= \int_\Omega \int_{\mathbb{R}^d} G(t, x, x', \xi, \xi')\, d\mu_{x'}(\xi')\, d\nu(x') \qquad \forall (t, x, \xi) \in \mathbb{R} \times \Omega \times \mathbb{R}^d\end{aligned}} \tag{20}$$

Example 4. *In Example 1 (opinion model), the mean field is*

$$\mathcal{X}[\mu](t, x, \xi) = \int_{\Omega \times \mathbb{R}^d} \sigma(x, x')(\xi' - \xi)\, d\mu(x', \xi') \qquad \forall (t, x, \xi) \in \mathbb{R} \times \Omega \times \mathbb{R}^d.$$

2.2.2 Vlasov equation

We consider the *Vlasov (or continuity) equation*

$$\boxed{\partial_t \mu + \mathrm{div}_\xi(\mathcal{X}[\mu]\mu) = 0} \tag{21}$$

where the divergence[2] acts only with respect to ξ. It is a nonlocal transport equation because the velocity field $\mathcal{X}[\mu]$ defined by (20) is nonlocal.

Remark 1. Given any solution $t \mapsto \mu(t)$ of the Vlasov equation (21), the total mass $\mu(t)(\Omega \times \mathbb{R}^d)$ is constant with respect to t, i.e., $\mu(t)$ is a probability measure for every $t \in \mathbb{R}$. Also, the marginal $\nu = \pi_* \mu(t)$ does not depend on t, because the Vlasov equation can be written as $\partial_t \mu + L_{\mathcal{X}[\mu]}\mu = 0$ with the Lie derivative acting with respect to the variable ξ, and we have $\pi_* L_{\mathcal{X}[\mu]} = 0$.

Disintegrating $\mu_t = \mu(t)$ as $\mu_t = \int_\Omega \mu_{t,x}\, d\nu(x)$ with respect to its marginal $\nu = \pi_*\mu_t$ on Ω (which does not depend on t by Remark 1), by uniqueness ν-almost everywhere of the disintegration, (21) is equivalent to

$$\partial_t \mu_{t,x} + \mathrm{div}_\xi(\mathcal{X}[\mu_t](t,x,\cdot)\mu_{t,x}) = 0 \tag{22}$$

for ν-almost every $x \in \Omega$. Note that the time evolution of $\mu_{t,x} = \varphi_{\mu_0}(t,x,\cdot)_*\mu_{0,x}$ depends on the whole μ_0 and not only on $\mu_{0,x}$, insofar $\mathcal{X}[\mu_t]$ involves an integral over all possible $x' \in \Omega$.

Theorem 1. *[Existence, uniqueness and stability properties for the Vlasov equation (21)]*

(A) *Given any $\mu_0 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$, there exists a unique solution $t \mapsto \mu(t)$ of the Vlasov equation (21) in $\mathscr{C}^0(\mathbb{R}, \mathcal{P}_c(\Omega \times \mathbb{R}^d))$, such that $\mu(0) = \mu_0$, locally Lipschitz with respect to t for the distance $L^1_\nu W_1$, and we have*

$$\mu_t = \mu(t) = \varphi_{\mu_0}(t,\cdot,\cdot)_*\mu_0,$$

which means that $\mu_{t,x} = \varphi_{\mu_0}(t,x,\cdot)_*\mu_{0,x}$ for every $t \in \mathbb{R}$ and ν-almost every $x \in \Omega$, and where $t \mapsto \varphi_{\mu_0}(t,x,\cdot)$ is the unique solution of

$$\partial_t \varphi_{\mu_0}(t,x,\cdot) = \mathcal{X}[\mu(t)](t,x,\cdot) \circ \varphi_{\mu_0}(t,x,\cdot)$$

such that $\varphi_{\mu_0}(0,x,\cdot) = \mathrm{id}_{\mathbb{R}^d}$ for ν-almost every $x \in \Omega$. Moreover, if $\mu_0 \in \mathcal{P}_c^{ac}(\Omega \times \mathbb{R}^d)$ then $\mu(t) \in \mathcal{P}_c^{ac}(\Omega \times \mathbb{R}^d)$ for every $t \in \mathbb{R}$. Furthermore:

($\mathbf{A_1}$) *Any solution $\mu(\cdot)$ of (21) depend continuously on its initial condition $\mu(0) \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ for the weak topology. Equivalently, given any $\mu(0) \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$, given any sequence of measures $\mu^k(0) \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$, for $k \in \mathbb{N}^*$, if $\mu^k(0)$ converges weakly to $\mu(0)$ (equivalently, $W_1(\mu^k(0), \mu(0)) \to 0$) as $k \to +\infty$, then $\mu^k(t)$ converges weakly to $\mu(t)$ (equivalently, $W_1(\mu^k(t), \mu(t)) \to 0$) as $k \to +\infty$, uniformly on any compact interval of times.*

($\mathbf{A_2}$) *Setting*

$$c(\mu^1(t), \mu^2(t)) = \exp\Bigg(2\int_0^t \Big(\|G(s,\cdot,\cdot,\cdot,\cdot)_{|S(s)\times S(s)}\|_{L^\infty} + \max_{x,x' \in \mathrm{supp}(\nu)} \mathrm{Lip}(G(s,x,x',\cdot,\cdot)_{|S_x(s)\times S_{x'}(s)})\Big)\, ds\Bigg) \tag{23}$$

where $S(s) = \mathrm{supp}(\mu^1(s)) \cup \mathrm{supp}(\mu^2(s))$ and $S_x(s) = \{\xi \in \mathbb{R}^d \mid (x,\xi) \in S(s)\}$ for any $x \in \mathrm{supp}(\nu)$ (note that $S(s)$, $S_x(s)$ and $\mathrm{supp}(\nu)$ are compact and that $\mathrm{supp}(\mu^i(s)) = \phi_{\mu_0}(s, \mathrm{supp}(\mu^i(0)))$ for $i = 1,2$), we have

$$L^1_\nu W_1(\mu^1(t), \mu^2(t)) \leqslant c(\mu^1(t), \mu^2(t))\, L^1_\nu W_1(\mu^1(0), \mu^2(0)) \qquad \forall t \geqslant 0 \tag{24}$$

for all locally Lipschitz solutions $\mu^1(\cdot)$ and $\mu^2(\cdot)$ of (21) such that $\mu^1(0), \mu^2(0) \in \mathcal{P}_c^\nu(\Omega \times \mathbb{R}^d)$ have the same marginal ν on Ω.

[2] Recall that $\mathrm{div}(\mathcal{X}\mu) = L_{\mathcal{X}}\mu$ (Lie derivative of the measure μ) is the measure defined by $\langle L_{\mathcal{X}}\mu, f\rangle = -\langle \mu, L_{\mathcal{X}} f\rangle = -\int_{\mathbb{R}^d} \mathcal{X}.\nabla f\, d\mu$ for every $f \in \mathscr{C}_c^\infty(\mathbb{R}^d)$.

(**B**) *Assuming that G is locally Lipschitz with respect to (x, x', ξ, ξ') (uniformly with respect to t on any compact) and setting*

$$C(\mu^1(t), \mu^2(t)) = \exp\left(2\int_0^t \|G(s,\cdot,\cdot,\cdot,\cdot)_{|S(s)\times S(s)}\|_{\mathscr{C}^{0,1}}\, ds\right) \tag{25}$$

where $S(s)$ is defined as in (23), we have

$$W_1(\mu^1(t), \mu^2(t)) \leqslant C(\mu^1(t), \mu^2(t))\, W_1(\mu^1(0), \mu^2(0)) \qquad \forall t \geqslant 0 \tag{26}$$

for all locally Lipschitz solutions $\mu^1(\cdot)$ and $\mu^2(\cdot)$ of (21) such that $\mu^1(0), \mu^2(0) \in \mathcal{P}_c(\Omega \times I\!R^d)$.

Theorem 1 is proved in Appendix B.1. The statement (**B**) of Theorem 1 is a slight extension, with parameter x, of [40, Theorem 2.3] (see also [37, 38, 39]) where it is assumed that G is globally Lipschitz, by using the more general Assumption (**G$_2$**) that there is no blow-up in finite time and the fact that $\Phi(t)$ maps a compact set to a compact set. The statement (**A**) seems to be new. Note that, in (**A$_2$**), the initial measures $\mu_1(0)$ and $\mu_2(0)$ are required to have the same marginal (and thus, equivalently, $\mu^1(t)$ and $\mu^2(t)$ have the same marginal for any t). At the contrary, in (**A$_1$**) and in (**B**), the measures under consideration are ont assumed to have the same marginal. In (**A$_1$**), the weak convergence $\mu^k(0) \rightharpoonup \mu(0)$ implies the weak convergence $\nu^k \rightharpoonup \nu$ of marginals but it is wrong in general that $\mu^k_x(0) \rightharpoonup \mu_x(0)$ for $x \in \Omega$.

In the statement (**B**), the assumption that G is locally Lipschitz with respect to (x, x', ξ, ξ') is much stronger than Assumption (**G$_1$**): in Example 1 (resp., Example 2) it requires σ (resp., a) to be locally Lipschitz. In general, requiring that G be locally Lipschitz with respect to (x, x') is not a natural assumption for the particle system (16). Note that, under this stronger assumption, the unique solution $\mu(\cdot)$ in (**A**) is locally Lipschitz with respect to t for the Wasserstein distance W_1.

Particular case: the indistinguishable case. We speak of the "indistinguishable case" whenever G does not depend on (x, x'). This is the classical case that has been much studied in the existing literature. The mean field is then

$$\mathcal{X}[\bar\mu](t,\xi) = \int_{\mathbb{R}^d} G(t,\xi,\xi')\, d\mu(\xi) \qquad \forall (t,\xi) \in \mathbb{R}\times\mathbb{R}^d$$

where $\bar\mu \in \mathcal{P}(\mathbb{R}^d)$ is denoted with an upper bar, to avoid any confusion with measures $\mu \in \mathcal{P}(\Omega \times \mathbb{R}^d)$.

We have the following corollary of Theorem 1, already well known in the existing literature.

Corollary 1. *Assume that G does not depend on (x, x'). Given any $\bar\mu_0 \in \mathcal{P}_c(I\!R^d)$, there exists a unique solution $t \mapsto \bar\mu(t)$ of the Vlasov equation (21) (without dependence on x) in $\mathscr{C}^0(I\!R, \mathcal{P}_c(\times I\!R^d))$, locally Lipschitz with respect to t for the distance W_1, such that $\bar\mu(0) = \bar\mu_0$, and we have*

$$\bar\mu(t) = \varphi_{\mu_0}(t,\cdot)_*\bar\mu_0$$

where $t \mapsto \varphi_{\mu_0}(t,\cdot)$ is the unique solution of $\partial_t\varphi_{\mu_0}(t,\cdot) = \mathcal{X}[\mu(t)](t,\cdot)\circ\varphi_{\mu_0}(t,\cdot)$ such that $\varphi_{\mu_0}(0,\cdot) = \mathrm{id}_{I\!R^d}$. Moreover, if $\bar\mu_0 \in \mathcal{P}^{ac}_c(I\!R^d)$ then $\bar\mu(t) \in \mathcal{P}^{ac}_c(I\!R^d)$ for every $t \in I\!R$.

Furthermore, we have

$$W_1(\bar\mu^1(t), \bar\mu^2(t)) \leqslant C(\bar\mu^1(t), \bar\mu^2(t))\, W_1(\bar\mu^1(0), \bar\mu^2(0)) \qquad \forall t \geqslant 0 \tag{27}$$

for all locally Lipschitz solutions $\bar\mu^1(\cdot)$ and $\bar\mu^2(\cdot)$ of (21) such that $\bar\mu^1(0), \bar\mu^2(0) \in \mathcal{P}_c(I\!R^d)$. Here, $C(\bar\mu^1(t), \bar\mu^2(t))$ defined by (25) coincides with $c(\bar\mu^1(t), \bar\mu^2(t))$ defined by (23).

Proof. Let $\bar{\nu}$ be an arbitrary probability measure on Ω. Given any $\bar{\mu} \in \mathcal{P}(\mathbb{R}^d)$, we define $\mu \in \mathcal{P}(\Omega \times \mathbb{R}^d)$ by $\mu = \bar{\nu} \otimes \bar{\mu}$. This means that the marginal of μ on Ω is $\bar{\nu}$ and that the disintegration of $\mu = \int_\Omega \mu_x \, d\bar{\nu}(x)$ with respect to $\bar{\nu}$ is given by $\mu_{\bar{x}} = \bar{\mu}$ on $\mathrm{supp}(\bar{\nu})$ (and we can take $\mu_x = 0$ if $x \in \Omega \setminus \mathrm{supp}(\bar{\nu})$).

This embedding allows us to recover Corollary 1 as a consequence of Theorem 1. Indeed, obviously, $\bar{\mu}(\cdot)$ is solution of the Vlasov equation (21) without dependence on x if and only if $\mu(\cdot) = \bar{\nu} \otimes \bar{\mu}(\cdot)$ is solution of the Vlasov equation (21). This gives the first part of the corollary.

To obtain (27), it suffices to note that $W_1(\bar{\mu}^1, \bar{\mu}^2) = W_1(\bar{\nu} \otimes \bar{\mu}^1, \bar{\nu} \otimes \bar{\mu}^2)$, by Lemma 11 in Appendix A.1. Then, (27) follows from (24) or from (26). □

2.2.3 Relationship between the particle system and the Vlasov equation

The relationship between the particle system (16) and the Vlasov equation (21) is given by the result below. Here and throughout, δ_ξ is the Dirac measure at ξ.

Let $N \in \mathbb{N}^*$ be arbitrary. Given any $X = (x_1, \ldots, x_N) \in \Omega^N$ and any $\Xi = (\xi_1, \ldots, \xi_N) \in \mathbb{R}^{dN}$, we define the *empirical measure* $\mu^E_{(X,\Xi)} \in \mathcal{P}(\Omega \times \mathbb{R}^d)$ by

$$\mu^E_{(X,\Xi)} = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \delta_{\xi_i}. \tag{28}$$

The disintegration of $\mu^E_{(X,\Xi)}$ with respect to its marginal $\nu^E_X = \pi_* \mu^E_{(X,\Xi)} = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ on Ω (that is itself an empirical measure) gives the family of conditional measures defined by $\mu^E_x = \delta_{x_i}$ if $x = x_i$ and 0 otherwise.

Proposition 1. *Let $N \in \mathbb{N}^*$ be arbitrary. The mapping $t \mapsto \Xi(t) = (\xi_1(t), \ldots, \xi_N(t)) \in \mathbb{R}^{dN}$, with $X = (x_1, \ldots, x_N) \in \Omega^N$, is a solution of the particle system (17) if and only if the mapping $t \mapsto \mu^E_{(X,\Xi(t))}$ is a locally Lipschitz solution of the Vlasov equation (21).*

By the way, we obviously have

$$\mu^E_{(X,\Xi(t))} = \varphi_{\mu^E_{(X,\Xi(0))}}(t, \cdot, \cdot)_* \, \mu^E_{(X,\Xi(0))},$$

i.e., the propagation of an empirical measure under the flow of the Vlasov equation is the empirical measure of the corresponding solution of the particle system.

Indeed, we have $\mathcal{X}[\mu^E_{(X,\Xi(t))}(t)](t, x, \xi) = \frac{1}{N} \sum_{j=1}^N G(t, x, x_j, \xi, \xi_j(t))$ and then (17) is equivalent to $\dot{\xi}_i(t) = \mathcal{X}[\mu^E_{(X,\Xi(t))}(t)](t, x_i, \xi_i(t))$ for $i = 1, \ldots, N$. Note, in passing, that $\mathcal{X}[\mu^E_{(X,\Xi)}](t, x_i, \xi_i) = Y_i(t, X, \Xi)$ for $i = 1, \ldots, N$.

Proof. The Vlasov equation (21) is written as $\partial_t \mu + L_{\mathcal{X}[\mu]} \mu = 0$ with the Lie derivative acting with respect to the variable ξ. Hence, setting $X(t) = (x_1(t), \ldots, x_N(t))$ and $\Xi(t) = (\xi_1(t), \ldots, \xi_N(t))$, the mapping $t \mapsto \mu^E_{(X(t),\Xi(t))}$ is a locally Lipschitz solution of the Vlasov equation (21) if and only if, for any $g \in \mathscr{C}^\infty_c(\Omega \times \mathbb{R}^d)$, we have $\langle \partial_t \mu^E_{(X,\Xi)} + L_{\mathcal{X}[\mu^E_{(X,\Xi)}]} \mu^E_{(X,\Xi)}, g \rangle = 0$, i.e.,

$$0 = \frac{1}{N} \sum_{i=1}^N \left(\frac{d}{dt} g(x_i(t), \xi_i(t)) - \partial_\xi g(x_i(t), \xi_i(t)) . \frac{1}{N} \sum_{j=1}^N G(t, x_i(t), x_j(t), \xi_i(t), \xi_j(t)) \right)$$

which gives immediately (16), and conversely. □

This shows that there is no difficulty to consider the mean field limit of a system of interacting particles in which the interactions depend on the agents, provided that there is a limit function G satisfying Assumption $(\mathbf{G_1})$: one just has to define additional variables x_i, for $i = 1, \dots, N$, and couple the dynamics of the ξ_i to the inertial equation $\dot{x}_i = 0$. This idea originates from [12, Section 5.2]. It seems that, in spite of its simplicity, such a generalization has not been considered in the literature.

Therefore, as in the classical situation where the interaction mapping G does not depend on the agents, any solution of the system of particles (16) can be embedded to a solution of the Vlasov equation (21) by considering an empirical measure.

As a consequence of the statements $(\mathbf{A_1})$ and $(\mathbf{B})$ of Theorem 1 and of Proposition 1, we have the following corollary.

Corollary 2. *Let $\mu_0 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ and let $t \mapsto \mu(t) = \varphi_{\mu_0}(t, \cdot, \cdot)_* \mu_0$ be the solution of the Vlasov equation (21) such that $\mu(0) = \mu_0$. Besides, for every $N \in \mathbb{N}^*$, let $(X, \Xi_0) \in \Omega^N \times \mathbb{R}^{dN}$ (we do not put any superscript N to (X, Ξ_0) to keep a better readability) be such that the empirical measure $\mu^E_{(X,\Xi_0)}$ converges weakly (equivalently, in Wasserstein distance W_1) to μ_0 as $N \to +\infty$ (see Appendix A.2). For every $N \in \mathbb{N}^*$, let $t \mapsto \Xi(t)$ be the solution of the particle system (17), with parameter X, such that $\Xi(0) = \Xi_0$ (we do not put any superscript N).*

Then, the empirical measure $\mu^E_{(X,\Xi(t))}$ converges weakly (equivalently, in Wasserstein distance W_1) to $\mu(t)$ as $N \to +\infty$, uniformly with respect to t on any compact interval.

If moreover G is locally Lipschitz with respect to (x, x', ξ, ξ') (uniformly with respect to t on any compact), then $W_1(\mu(t), \mu^E_{(X,\Xi(t))}) \leqslant C(\mu(t), \mu^E_{(X,\Xi(t))})\, W_1(\mu_0, \mu^E_{(X,\Xi_0)})$ for every $t \geqslant 0$ (with C defined by (25)).

Remark 2. Alternatively, we may also consider *semi-empirical measures*: taking

$$(\mu_0)^{SE}_X = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \mu_{0,x_i},$$

the unique solution of the Vlasov equation (21), of initial condition $(\mu_0)^{SE}_X$, is the semi-empirical measure

$$(\mu_t)^{SE}_X = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \mu_{t,x_i}.$$

This is due to the fact that the marginal of $(\mu_0)^{SE}_X$ on Ω is the empirical measure $\nu = \nu^E_X = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, and (21) is equivalent to (22). In other words, the propagation of a semi-empirical measure under the flow of the Vlasov equation is a semi-empirical measure.

This also means that, considering a solution $\mu(t) = \varphi_{\mu_0}(t, \cdot, \cdot)_* \mu_0$ of the Vlasov equation (21), the operation of taking the semi-empirical measure approximation commutes with the operation of propagating under the Vlasov flow.

2.3 Eulerian viewpoint: Liouville equation

The Eulerian viewpoint consists of propagating, for any parameter $X \in \Omega^N$, an initial probability measure in $\mathbb{R}^{dN}$ under the flow of diffeomorphisms $\Phi(t, X, \cdot)$ of $\mathbb{R}^{dN}$ generated by the time-dependent vector field $Y(t, X, \cdot)$ defined by (18).

We consider the *(N-body) Liouville equation* associated with the time-dependent vector field Y defined by (18), depending on the parameter $X \in \Omega^N$, given by

$$\boxed{\partial_t \rho + \mathrm{div}_\Xi(Y\rho) = 0} \tag{29}$$

This is a usual (local) transport equation on $\mathbb{R}^{dN}$, parametrized by $X \in \Omega^N$, where the divergence is considered with respect to $\Xi = (\xi_1, \ldots, \xi_N)$, and we thus have the following standard result. Here, it is understood that $\rho(t)$ is a probability Radon measure on $\Omega^N \times \mathbb{R}^{dN}$.

Proposition 2. *Given any $\rho_0 \in \mathcal{P}(\Omega^N \times \mathbb{R}^{dN})$, there exists a unique solution $t \mapsto \rho(t)$ of the Liouville equation (29) in $\mathscr{C}^0(\mathbb{R}, \mathcal{P}(\Omega^N \times \mathbb{R}^{dN}))$, locally Lipschitz with respect to t for the distance $L^1_\theta W_1$, such that $\rho(0) = \rho_0$, given by*

$$\rho(t) = \Phi(t)_* \rho_0,$$

i.e., $\rho(t)$ is the image (pushforward) of ρ_0 under the particle flow.

Let us make precise some notations and in particular the disintegration procedure. Given any measure $\rho \in \mathcal{P}(\Omega^N \times \mathbb{R}^{dN})$, denoting by $\pi^{\otimes N} : \Omega^N \times \mathbb{R}^{dN} \to \Omega^N$ the canonical projection, we will always denote by θ the probability Radon measure on Ω^N given by $\theta = (\pi^{\otimes N})_* \rho$ (image of ρ under $\pi^{\otimes N}$), that is the marginal of ρ on Ω^N. By disintegration of ρ with respect to θ, there exists a family $(\rho_X)_{X \in \Omega^N}$ of probability Radon measures on $\mathbb{R}^{dN}$ such that $\rho = \int_{\Omega^N} \rho_X \, d\theta(X)$.

With these notations, $\rho_t = \rho(t) = \Phi(t)_* \rho_0$ is disintegrated as $\rho_t = \int_{\Omega^N} \rho_{t,X} \, d\theta(X)$ with respect to its marginal $\theta = (\pi^{\otimes N})_* \rho(t)$ on Ω^N. The marginal θ does not depend on t because (29) can be written as $\partial_t \rho + L_Y \rho = 0$, with the Lie derivative acting with respect to the variable ξ, and we have $(\pi^{\otimes N})_* L_Y = 0$. Moreover, we have

$$\rho_{t,X} = (\Phi_{t,X})_* \rho_{0,X}$$

for every $t \in \mathbb{R}$ and for θ-almost every $X \in \Omega^N$.

Remark 3. If $\rho_0 = \delta_X \otimes \delta_{\Xi_0}$ for some $(X, \Xi_0) \in \Omega^N \times \mathbb{R}^{dN}$ then $\rho(t) = \delta_X \otimes \delta_{\Xi(t)}$ where $t \mapsto \Xi(t)$ is the solution of the particle system (17) with parameter X such that $\Xi(0) = \Xi_0$. In other words, the solutions of the particle system are naturally embedded as Dirac measures solutions of the Liouville system.

Hence, in some sense, the Liouville equation contains all possible solutions of the particle system. But it contains more: considering the particle system (17), instead of taking a *deterministic* initial condition $\Xi(0) = \Xi_0 \in \mathbb{R}^{dN}$, one may want to take a *distribution* of initial conditions, for instance one may want to consider all possible initial conditions that are distributed around Ξ_0 according to a Gaussian law, in order to take into account noise or uncertainties in the initial conditions. In such a way, the Liouville equation (29) has a *probabilistic* interpretation with respect to the particle system (16).

If the probability measure $\rho(t)$ on $\Omega^N \times \mathbb{R}^{dN}$ has a density f, then $f(t, X, \Xi)$ represents the density of particles having the positions $X = (x_1, \ldots, x_N) \in \Omega^N$ and respective momenta $\Xi = (\xi_1, \ldots, \xi_N) \in \mathbb{R}^{dN}$. This is in contrast with the mean field procedure that consists of taking the large N limit of the average over all particles but one. In the next section we show how to recover Vlasov from Liouville by taking marginals.

2.4 Recovering Vlasov from Liouville by taking marginals

Compared with $\mu(t)$ that is a probability measure on $\Omega \times \mathbb{R}^d$, here, $\rho(t)$ is a probability measure on $\Omega^N \times \mathbb{R}^{dN}$. It is thus tempting to search for a relationship between $\mu(t)$ and $\rho(t)$ by taking marginals of $\rho(t)$. This is what has been done in [31] or in [26] in the different context of quantum mechanics. Adapted to the present situation, the method developed in [26], which provides an explicit rate of convergence, consists of proving that the marginals of the solutions $\rho(t)$ of (29) are close, in Wasserstein topology, to solutions $\mu(t)$ of the Vlasov equation (21), as established hereafter.

As we are going to see, this can be done by taking adequate initial conditions ρ_0 for the Liouville equation (29). We have to perform a symmetrization under permutations for the initial condition ρ_0 and also for the corresponding solution $\rho(t)$, not only with respect to Ξ but also with respect to the parameter variable X. This creates several difficulties. Note that the symmetrization is not preserved by the flow, so we have to consider the symmetrization $\rho(t)^s$ at any time t.

Given any $\rho \in \mathcal{P}(\Omega^N \times \mathbb{R}^{dN})$, we define the measure $\rho^s \in \mathcal{P}(\Omega^N \times \mathbb{R}^{dN})$, called the symmetrization under permutations of ρ, by

$$\int_{(\Omega\times\mathbb{R}^d)^N} f(X,\Xi)\, d\rho^s(X,\Xi) = \frac{1}{N!}\sum_{\sigma\in\mathfrak{S}_N}\int_{\Omega^N\times\mathbb{R}^{dN}} f(X_\sigma,\Xi_\sigma)\, d\rho(X,\Xi) \tag{30}$$

for every $f \in \mathscr{C}_c^\infty(\Omega^N \times \mathbb{R}^{dN})$, where $X_\sigma = (x_{\sigma(1)},\dots,x_{\sigma(N)})$ and $\Xi_\sigma = (\xi_{\sigma(1)},\dots,\xi_{\sigma(N)})$ for all $X \in \Omega^N$ and $\Xi \in \mathbb{R}^{dN}$, and where $\mathfrak{S}_N$ is the group of permutations of N elements.

Now, given any $n \in \{1,\dots,N\}$, we denote by $\rho^s_{N:n}$ the nth-order marginal of ρ^s (not to be confused with the symmetrization under permutations of the marginal, which we do not use), which is, by definition, the image of ρ^s under the projection of $\Omega^N \times \mathbb{R}^{dN}$ on the product $\Omega^n \times \mathbb{R}^{dn}$ of the n first copies of Ω with the n first copies of $\mathbb{R}^d$.

In this section, we establish two ways for recovering Vlasov from Liouville by taking marginals.

Let $\mu_0 \in \mathcal{P}_c(\Omega\times\mathbb{R}^d)$, disintegrated as $\mu_0 = \int_\Omega \mu_{0,x}\, d\nu(x)$ with respect to its marginal $\nu = \pi_*\mu_0$ on Ω. We consider the unique solution $t \mapsto \mu(t) = \varphi_{\mu_0}(t,\cdot,\cdot)_*\mu_0$ of the Vlasov equation (21) such that $\mu(0) = \mu_0$, given by Theorem 1. Recall that $\mu_{t,x} = \varphi_{\mu_0}(t,x,\cdot)_*\mu_{0,x}$ for ν-almost every $x \in \Omega$. When G is locally Lipschitz with respect to (x,x',ξ,ξ'), we set

$$C(\mu(t)) = \exp\left(2\int_0^t \|G(s,\cdot,\cdot,\cdot,\cdot)_{|S(s)\times S(s)}\|_{\mathscr{C}^{0,1}}\, ds\right) \tag{31}$$

where $S(s) = \operatorname{supp}(\mu(s))$.

Hereafter, we propose two possible choices of $\rho_0 \in \mathcal{P}_c((\Omega\times\mathbb{R}^d)^N)$, generating by Proposition 2 the solution $\rho(t) = \Phi(t)_*\rho_0$ of the Liouville equation (29) from which we recover $\mu(t)$ solution of the Vlasov equation (21) by taking marginals.

2.4.1 First way, with ρ_0 Dirac

Given any fixed $N \in \mathbb{N}^*$, let $(X,\Xi_0) \in (\operatorname{supp}(\mu_0))^N \subset \Omega^N \times \mathbb{R}^{dN}$; we do not put any superscript N to (X,Ξ_0) to keep a better readability, but typically we may want that the empirical measure $\mu^E_{(X,\Xi_0)}$ converges to μ_0 in Wasserstein distance as $N \to +\infty$ (see Appendix A.2 for such conditions). Let $t \mapsto \Xi(t)$ be the solution of the particle system (17) such that $\Xi(0) = \Xi_0$. If $\mu^E_{(X,\Xi_0)}$ converges to μ_0 then, by Corollary 2, the empirical measure $\mu^E_{(X,\Xi(t))}$ converges to $\mu(t)$ in Wasserstein distance as $N \to +\infty$. Note that $(X,\Xi(t)) \in \operatorname{supp}(\mu(t)) = \varphi_{\mu_0}(t,\operatorname{supp}(\mu_0))$.

Defining $\rho_0 \in \mathcal{P}(\Omega^N \times \mathbb{R}^{dN})$ as the Dirac measure $\rho_0 = \delta_X \otimes \delta_{\Xi_0}$, by Remark 3, the unique solution of the Liouville equation (29) such that $\rho(0) = \rho_0$, is given by the Dirac measure

$$\rho(t) = \Phi(t)_*\rho_0 = \delta_X \otimes \delta_{\Xi(t)}.$$

It is then easy to see that $\rho(t)^s_{N:1} = \mu^E_{(X,\Xi(t))}$ (see the proof of the theorem below). Therefore, if $\mu^E_{(X,\Xi_0)}$ converges weakly to μ_0 then $\rho(t)^s_{N:1}$ converges weakly to $\mu(t)$ as $N \to +\infty$. The convergence is less obvious for the marginals of order $n \geqslant 2$. We have the following general result, for any given $N \in \mathbb{N}^*$, independently of any sequence.

Theorem 2. *With the above notations, we have the following statements.*

(**A**) *If $\mu^E_{(X,\Xi_0)}$ converges weakly (equivalently, in Wasserstein distance W_1) to μ_0 as $N \to +\infty$, then, for every $n \in \mathbb{N}^*$, $\rho(t)^s_{N:n}$ converges weakly (equivalently, in Wasserstein distance W_1) to $\mu(t)^{\otimes n}$ as $N \to +\infty$, uniformly with respect to t on any compact interval.*

(**B**) *Assuming that G is locally Lipschitz with respect to (x, x', ξ, ξ') (uniformly with respect to t on any compact), for every $t \geqslant 0$, we have*

$$W_1\left(\rho(t)^s_{N:1}, \mu(t)\right) \leqslant C(\mu(t))\, W_1\left(\mu^E_{(X,\Xi_0)}, \mu_0\right) \tag{32}$$

with $\rho(t)^s_{N:1} = \mu^E_{(X,\Xi(t))}$, where $C(\mu(t))$ is defined by (31), and, for every $n \in \{2, \ldots, N\}$,

$$W_1\left(\rho(t)^s_{N:n}, \mu(t)^{\otimes n}\right) \leqslant 2\left(e^{\frac{n^2}{2N}} - 1\right) \max\left(1, \|\mathrm{supp}(\mu(t))\|_\infty\right) + n\, C(\mu(t))\, W_1\left(\mu^E_{(X,\Xi_0)}, \mu_0\right). \tag{33}$$

Theorem 2 is proved in Appendix B.2.

As alluded above, to obtain an interesting convergence result from this theorem, let us assume that $W_1\left(\mu^E_{(X,\Xi_0)}, \mu_0\right) \leqslant \frac{C_1}{N}$ for some $C_1 > 0$ not depending on N. The latter requirement is satisfied if the sequence (X, Ξ_0) "adequately approximates" μ_0 on $\mathrm{supp}(\mu_0)$ as $N \to +\infty$ (see Appendix A.2 for such results).

Then, noting that $e^{\frac{n^2}{2N}} - 1 \sim \frac{n^2}{2N}$ when $n \ll \sqrt{N}$, it follows from (33) that $\rho(t)^s_{N:n}$ converges to $\mu(t)^{\otimes n}$ in Wassertein distance as $N \to +\infty$, uniformly with respect to t on compact intervals, with a rate of convergence $\frac{n^2}{N}$. This convergence result is thus relevant only when $n \ll \sqrt{N}$: it does not bring any meaningful information when $n \simeq N$.

Note that, in Theorem 2, we do not assume that ν be absolutely continuous with respect to a Lebesgue measure on Ω.

Particular case: the indistinguishable case. We have the following corollary of Theorem 2.

Corollary 3. *Assume that G does not depend on (x, x'). Let $\bar{\mu}_0 \in \mathcal{P}_c(\mathbb{R}^d)$ and let $t \mapsto \bar{\mu}(t)$ be the unique solution of the Vlasov equation (21) (without dependence on x) such that $\bar{\mu}(0) = \bar{\mu}_0$ (see Corollary 1). Besides, let $\bar{\rho}_0 = \delta_{\Xi_0}$ and let $t \mapsto \bar{\rho}(t) = \delta_{\Xi(t)}$ be the unique solution of the Liouville equation (29) (without dependence on X) such that $\bar{\rho}(0) = \bar{\rho}_0$. Then*

$$W_1\left(\bar{\rho}(t)^s_{N:1}, \bar{\mu}(t)\right) \leqslant C(\bar{\mu}(t))\, W_1\left(\bar{\mu}^E_{\Xi_0}, \bar{\mu}_0\right) \tag{34}$$

with $\bar{\rho}(t)^s_{N:1} = \mu^E_{\Xi(t)}$, where the empirical measure $\bar{\mu}^E_{\Xi}$ is defined by $\bar{\mu}^E_{\Xi} = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_i}$, and where $C(\bar{\mu}(t))$ is defined by (31) (without dependence on x, x'), and, for every $n \in \{2, \ldots, N\}$,

$$W_1\left(\bar{\rho}(t)^s_{N:n}, \bar{\mu}(t)^{\otimes n}\right) \leqslant 2\left(e^{\frac{n^2}{2N}} - 1\right) \max\left(1, \|\mathrm{supp}(\bar{\mu}(t))\|_\infty\right) + n\, C(\bar{\mu}(t))\, W_1\left(\bar{\mu}^E_{\Xi_0}, \bar{\mu}_0\right). \tag{35}$$

Proof. Following the proof of Corollary 1 and choosing $\bar{\nu} = \delta_{\bar{x}}$ for some arbitrary $\bar{x} \in \Omega$, in the indistinguishable case $\bar{\mu}(\cdot)$ is solution of the Vlasov equation (21) (without dependence on x) if and only if $\mu(\cdot) = \delta_{\bar{x}} \otimes \bar{\mu}(\cdot)$ is solution of the Vlasov equation (21). We now define $X = (\bar{x}, \ldots, \bar{x}) \in \Omega^N$, and we take $\rho_0 = \delta_X \otimes \delta_{\Xi_0}$ as initial condition for the Liouville equation in Theorem 2, so that $\rho(t) = \delta_X \otimes \bar{\rho}(t)$ where $\bar{\rho}(t) = \delta_{\Xi(t)}$. With these choices, we obviously have $\rho(t)^s_{N:n} = \delta_{\bar{x}} \otimes \bar{\rho}(t)^s_{N:n}$, and then (34) and (35) straightforwardly follows from (32) and (33), by applying Lemma 11 in Appendix A.1. □

2.4.2 Second way, with ρ_0 "semi-Dirac"

In this section, we assume that, for every $N \in \mathbb{N}^*$, there exists a tagged partition $(\mathcal{A}, X)$ of Ω associated with ν (satisfying (12), see Section 1.2; see also Appendix A.3), with $X = (x_1, \dots, x_N) \in \Omega^N$ (we do not write any superscript N in x_i). We set $\delta_X = \delta_{x_1} \otimes \cdots \delta_{x_N}$ and $\rho_{0,X} = \mu_{0,x_1} \otimes \cdots \otimes \mu_{0,x_N}$. Defining $\rho_0 \in \mathcal{P}(\Omega^N \times \mathbb{R}^{dN})$ as the "semi-Dirac" measure $\rho_0 = \delta_X \otimes \rho_{0,X}$, we consider the unique solution of the Liouville equation (29) such that $\rho(0) = \rho_0$, given by the "semi-Dirac" measure

$$\rho(t) = (\Phi_t)_* \rho_0 = \delta_X \otimes (\Phi_{t,X})_* \rho_{0,X} = \delta_X \otimes \rho_{t,X}$$

because the marginal $\theta = (\pi^{\otimes N})_* \rho_t$ of $\rho_t = \rho(t)$ on Ω is $\theta = \delta_X$, and $\rho_{t,X} = (\Phi_{t,X})_* \rho_{0,X}$ for every $t \in \mathbb{R}$.

As a preliminary remark, we claim that, at $t = 0$, we have

$$(\rho_0)^s_{N:1} = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \otimes \mu_{0,x_i} = (\mu_0)^{SE}_X \tag{36}$$

(semi-empirical measure), which converges weakly to μ_0 as $N \to +\infty$ under slight assumptions on μ_0, by Lemma 20 in Appendix A.3.2; more generally, $(\rho_0)^s_{N:n}$ converges weakly to $\mu_0^{\otimes n}$ (in the proof of the theorem hereafter, we give an explicit expression for $(\rho_0)^s_{N:n}$, using (123) in Appendix A.4). In the theorem below, we establish that this convergence is propagated in time.

Theorem 3. *Assume that there exists a family of tagged partitions as settled above.*

(**A**) *Assume that $x \mapsto \mu_{0,x}$ is ν-almost everywhere continuous for the Wasserstein distance W_1. Then, for every $n \in \mathbb{N}^*$, $\rho(t)^s_{N:n}$ converges weakly to $\mu(t)^{\otimes n}$ (equivalently, $W_1\left(\rho(t)^s_{N:n}, \mu(t)^{\otimes n}\right) \to 0$) as $N \to +\infty$, uniformly with respect to t on compact intervals.*

(**B**) *Assume that G is locally Lipschitz with respect to (x, x', ξ, ξ') (uniformly with respect to t on any compact). We also assume that $x \mapsto \mu_{0,x}$ is Lipschitz for the Wasserstein distance W_1, i.e., that there exists $L > 0$ such that $W_1(\mu_{0,x}, \mu_{0,y}) \leqslant L\, d_\Omega(x, y)$ for ν-almost all $x, y \in \Omega$. Then*

$$W_1\left(\rho(t)^s_{N:1}, \mu(t)\right) \leqslant \frac{1}{N}(L+1)\, C(\mu(t)) \tag{37}$$

where $C(\mu(t))$ is defined by (31), and, for every $n \in \{2, \dots, N\}$,

$$W_1\left(\rho(t)^s_{N:n}, \mu(t)^{\otimes n}\right) \leqslant \left(e^{\frac{n^2}{2N}} - 1\right)(1 + 2\|\mathrm{supp}(\mu(t))\|_\infty) + \frac{1}{N}(L+1)\, C(\mu(t)). \tag{38}$$

Theorem 3 is proved in Appendix B.3.

As above, noting that $e^{\frac{n^2}{2N}} - 1 \sim \frac{n^2}{2N}$ when $n \ll \sqrt{N}$, it follows from Theorem 3 that $\rho(t)^s_{N:n}$ converges to $\mu(t)^{\otimes n}$ in Wassertein distance as $N \to +\infty$, uniformly with respect to t on compact intervals, with a rate of convergence $\frac{n^2}{N}$.

Particular case: the indistinguishable case. We have the following corollary of Theorem 3.

Corollary 4. *Assume that G does not depend on (x, x'). Let $\bar{\mu}_0 \in \mathcal{P}_c(\mathbb{R}^d)$ and let $t \mapsto \bar{\mu}(t)$ be the unique solution of the Vlasov equation (21) (without dependence on x) such that $\bar{\mu}(0) = \bar{\mu}_0$ (see Corollary 1). Besides, let $\bar{\rho}_0 = \bar{\mu}_0^{\otimes N}$ and let $t \mapsto \bar{\rho}(t)$ be the unique solution of the Liouville equation (29) (without dependence on X) such that $\bar{\rho}(0) = \bar{\rho}_0$. Then*

$$W_1\left(\bar{\rho}(t)_{N:1}, \bar{\mu}(t)\right) \leqslant \frac{C(\bar{\mu}(t)) + 1}{N} \tag{39}$$

where $C(\bar{\mu}(t))$ is defined by (31) (without dependence on x, x'), and, for every $n \in \{2, \dots, N\}$,

$$W_1\left(\bar{\rho}(t)_{N:n}, \bar{\mu}(t)^{\otimes n}\right) \leqslant \left(e^{\frac{n^2}{2N}} - 1\right)(3 + 2\|\mathrm{supp}(\bar{\mu}(t))\|_\infty) + \frac{1}{N}C(\bar{\mu}(t)). \tag{40}$$

Corollary 4 is proved in Appendix B.4, using the proof of Corollary 1 with a choice of $\bar{\nu}$ different from the once done in the proof of Corollary 3.

Remark 4. Corollary 4, applied to the Hamiltonian case

$$G(t, (q_i, p_i), (q_j, p_j)) = \begin{pmatrix} p_i \\ \nabla V(q_i - q_j) \end{pmatrix}$$

improves the rate of convergence in W_1 distance obtained in [26, Theorem 3.1] from $\frac{1}{\sqrt{N}}$ to $\frac{1}{N}$. It can also be applied to more general Hamiltonian systems, for example in the presence of a magnetic field associated to a vector potential $A : \mathbb{R}^d \to \mathbb{R}^d$, where

$$G(t, (q_i, p_i), (q_j, p_j)) = \begin{pmatrix} p_i \\ -\nabla\left(V(q_i - q_j) + (p_i - A(q_i))^2\right) \end{pmatrix}.$$

Corollary 4 also applies to Cucker-Smale systems, for which we have

$$G(t, (q_i, p_i), (q_j, p_j)) = \begin{pmatrix} p_i \\ F(|q_i - q_j|)(p_i - p_j) \end{pmatrix},$$

and to its generalizations introduced in [34]; this improves in the same way the case $p = 1$ treated in [34, Remark 3.5].

2.4.3 Further comments

On the strategy. Given any $\mu_0 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$, we have given two possible ways for getting an approximation of $\mu(t)$, which converges to $\mu(t)$ as $N \to +\infty$. In the first way, ρ_0 Dirac gives (by taking the first marginal of the symmetrization) an approximation of $\mu(t)$ that is an empirical measure, while in the second way, ρ_0 semi-Dirac gives an approximation that is a semi-empirical measure.

Higher-order Wasserstein estimates. All estimates that we have derived can be obtained as well for the Wasserstein distance W_p, by using the inequality (8).

Extension of the support of the initial data. Theorems 2 and 3 can be extended to the slightly more general case where $\mu_0 \in \mathcal{P}_1(\Omega \times \mathbb{R}^d)$, i.e., the probability measure μ_0 is not of compact support but has a finite first moment, provided that we add the following assumption: the particle system (16) is stable in the sense of Lyapunov, meaning that, for every $R > 0$, there exists $R_0 > 0$ such that, taking any initial condition $(X, \Xi_0) \in \Omega^N \times \mathbb{R}^{dN}$ such that $\|\Xi_0\| \leqslant R_0$, the unique solution $t \mapsto \Xi(t)$ of (16), of parameter X, such that $\Xi(0) = \Xi_0$, satisfies $\|\Xi(t)\| \leqslant R$ for every $t \in \mathbb{R}$. The proof is done by approximating the measures considered in the proofs by measures of sufficiently large compact support. Due to the assumption on the particle flow, there is no mass escaping to infinity. Moreover, interestingly, in this case the estimates (32) and (33) of Theorem 2, and (37) and (38) of Theorem 3 are uniform with respect to $t \in \mathbb{R}$ because $C(\mu(t))$ and $\|\mathrm{supp}(\mu(t))\|_\infty$ are uniformly bounded with respect to t.

3 From mesoscopic to macroscopic scale

("from Vlasov to Euler", hydrodynamic limit)

Given any $\mu \in \mathcal{P}(\Omega \times \mathbb{R}^d)$, disintegrated as $\mu = \int_\Omega \mu_x \, d\nu(x)$, the three macroscopic quantities that are usually considered in the *hydrodynamic limit* procedure are the three first moments of the measure μ with respect to ξ, leading to define, for ν-almost every $x \in \Omega$:

- the total mass $\rho(x) \geqslant 0$ of μ_x by
$$\rho(x) = \int_{\mathbb{R}^d} d\mu_x(\xi) = \mu_x(\mathbb{R}^d),$$
(moment of order 0) which is here assumed to be equal to 1 for ν-almost every $x \in \Omega$;
- the "speed" $y(x) \in \mathbb{R}^d$ by
$$\rho(x)y(x) = \int_{\mathbb{R}^d} \xi \, d\mu_x(\xi),$$
(moment of order 1) which is also the expectation of any random law of probability distribution μ_x;
- and the "temperature" $T(x) \geqslant 0$ by
$$\frac{1}{2}\rho(x)\|y(x)\|^2 + \frac{d}{2}\rho(x)T(x) = \frac{1}{2}\int_{\mathbb{R}^d} \|\xi\|^2 \, d\mu_x(\xi)$$
or equivalently by
$$d\rho(x)T(x) = \int_{\mathbb{R}^d} \|\xi - y(x)\|^2 \, d\mu_x(\xi),$$
(moment of order 2) which is a variance.

Consider the mean field $\mathcal{X}[\mu]$ defined by (20) for every $\mu \in \mathcal{P}(\Omega \times \mathbb{R}^d)$. Let μ be a fixed locally Lipschitz solution of the Vlasov equation (21). According to Remark 1, $\nu(t) = \nu$ does not depend on t.

Following the hydrodynamic limit procedure recalled above (see also, e,g. [13, 22, 34]), for every $t \in \mathbb{R}$ and for ν-almost every $x \in \Omega$, we define $\rho(t,x)$, $y(t,x)$ and $T(t,x)$, that are the three first moments of $\mu(t)$. It is not useful, here, to consider the quantity $\rho(t,x)$, which does not depend on t and is equal to 1 for ν-almost every $x \in \Omega$ and 0 otherwise.

3.1 Moment of order 1: Euler equation

Given any solution $\mu(\cdot)$ of the Vlasov equation (21), we define

$$\boxed{y(t,x) = \int_{\mathbb{R}^d} \xi \, d\mu_{t,x}(\xi)} \tag{41}$$

for ν-almost every $x \in \Omega$, and $y(t,x) = 0$ for every $x \in \Omega \setminus \mathrm{supp}(\nu)$. Integrating by parts, we obtain

$$\partial_t y(t,x) = \left\langle \mu_{t,x}, L_{\mathcal{X}[\mu](x,\cdot)}(\xi \mapsto \xi) \right\rangle = \int_{\mathbb{R}^d} \mathcal{X}[\mu](x,\xi) \, d\mu_{t,x}(\xi)$$

which is a kind of "mean" mean field, since the mean field is now averaged under $\mu_{t,x}$. Hence

$$\partial_t y(t,x) = \int_{\mathbb{R}^d} \int_{\Omega \times \mathbb{R}^d} G(t,x,x',\xi,\xi') \, d\mu_t(x',\xi') \, d\mu_{t,x}(\xi). \tag{42}$$

It is remarkable that, for some classes of functions G, we obtain a "closed" equation in y:

3.1.1 Opinion propagation model: linear Euler equation

Proposition 3. *In the opinion propagation model of Example 1, we have* $G(t,x,x',\xi,\xi') = \sigma(x,x')(\xi'-\xi)$ *and*

$$\boxed{\partial_t y(t,x) = (Ay(t))(x)} \tag{43}$$

(Euler equation), where A *is the bounded operator on* $L^2_\nu(\Omega, I\!R^d)$ *defined by*

$$(Ay)(x) = \int_\Omega \sigma(x,x')(y(x')-y(x))\, d\nu(x') \qquad \forall y \in L^2_\nu(\Omega, I\!R^d) \tag{44}$$

and ν *is the marginal of* $\mu(t)$ *on* Ω *(not depending on* t*).*

Proof. Using the disintegration of the measure, we infer from (42) that

$$\partial_t y(t,x) = \underbrace{\int_{\mathbb{R}^d} d\mu_{t,x}(\xi)}_{=1} \int_\Omega \sigma(x,x') \underbrace{\int_{\mathbb{R}^d} \xi'\, d\mu_{t,x'}(\xi')}_{=y(t,x')}\, d\nu(x') - \underbrace{\int_{\mathbb{R}^d} \xi\, d\mu_{t,x}(\xi)}_{=y(t,x)} \int_\Omega \sigma(x,x') \underbrace{\int_{\mathbb{R}^d} d\mu_{t,x'}(\xi')}_{=1}\, d\nu(x')$$

and the result follows. □

Remark 5. If $\mu(0) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \delta_{\xi_i(0)} = \mu^E_{(X,\Xi(0))}$ as in Proposition 1, then $\mu(t) = \mu^E_{(X,\Xi(t))}$ (empirical measure), whose marginal on Ω is $\nu = \nu^E_X = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and whose disintegration with respect to ν is $\mu_{t,x} = \delta_{\xi_i}$ if $x = x_i$ for $i \in \{1,\dots,N\}$ and 0 otherwise. In this case, in the context of Proposition 3, we have then $y(t,x) = \xi_i(t)$ if $x = x_i$ for $i \in \{1,\dots,N\}$ and 0 otherwise, and the differential equation (43) exactly coincides with the particle system (19).

3.1.2 Open issue: how to obtain a closed equation?

In Section 3.1.1, the operator A is linear. An open question is to characterize the functions G so that, for any solution μ of (21), the function y defined by (41) satisfies an "Euler equation", possibly nonlinear, $\partial_t y(t,\cdot) = A(y(t,\cdot))$. We face here with the classical problem in kinetic theory of considering the three first moments of a solution μ of the Vlasov equation, and searching how to close the moment system since a priori the equations depend on higher-order moments. Suitable closure assumptions are not known so far, in general (see [13] for interesting comments, see also Section 3.3 further). This is why it is usual to consider a monokinetic ansatz for μ, as explained in the following section.

3.1.3 The ν-monokinetic case: general nonlinear Euler equation

Let us consider specific solutions μ of the Vlasov equation (21), that are ν-*monokinetic*, meaning that μ is delta-valued in the ξ variable and has the marginal ν on Ω. Given any $\nu \in \mathcal{P}(\Omega)$ and any measurable function $y : \Omega \to \mathbb{R}^d$, we define the ν-monokinetic measure μ^ν_y on $\Omega \times \mathbb{R}^d$ by

$$\mu^\nu_y = \nu \otimes \delta_{y(\cdot)}. \tag{45}$$

Proposition 4. *Let* $\nu \in \mathcal{P}_c(\Omega)$*. The mapping* $t \mapsto \mu(t) = \mu^\nu_{y(t,\cdot)} \in \mathcal{P}_c(\Omega \times I\!R^d)$*, of marginal* ν *on* Ω*, is a (*ν*-monokinetic) locally Lipschitz solution of the Vlasov equation* (21) *with the general*

mean field (20) if and only if the mapping $t \mapsto y(t,\cdot) \in L^\infty_\nu(\Omega, \mathbb{R}^d)$, given by (41), with $y(0,\cdot)$ of compact (essential) support, is a locally Lipschitz solution of the (nonlinear) Euler equation

$$\boxed{\partial_t y(t,\cdot) = A(t, y(t,\cdot))} \tag{46}$$

where $A : \mathbb{R} \times L^\infty_\nu(\Omega, \mathbb{R}^d) \to L^\infty_\nu(\Omega, \mathbb{R}^d)$ is the nonlinear operator (depending on ν) defined by

$$\boxed{(A(t,y))(x) = \int_\Omega G(t, x, x', y(x), y(x'))\, d\nu(x')} \tag{47}$$

for every $t \in \mathbb{R}$ and for every $y \in L^\infty_\nu(\Omega, \mathbb{R}^d)$ of compact (essential) support.

Proof. When $\mu_t = \mu^\nu_{y(t,\cdot)}$, (20) gives $\mathcal{X}[\mu_t](t,x,\xi) = \int_\Omega G(t,x,x',\xi,y(t,x'))\, d\nu(x')$. The computation is then straightforward, and we can note that $(A(t,y))(x) = \mathcal{X}[\mu^\nu_y](t,x,y(x))$.

Note also that $\mu(0)$ is of compact support if and only if $y(0,\cdot)$ is of compact (essential) support. Under this assumption $\mu(t)$ is of compact support and $y(t,\cdot)$ is of compact (essential) support. □

When μ_t is not of the form $\mu^\nu_{y(t,\cdot)}$, $t \mapsto y(t,\cdot)$ fails in general to satisfy a "closed" equation (i.e., $\partial_t y(t,\cdot)$ may not be expressible only in function of $y(t,\cdot)$). Instead, there may be a full hierarchy of equations coupling all the moments of $\mu_{t,x}$ (see further). Anyway, when convergence to consensus holds, we expect that any solution μ of (21) is asymptotically of the form $\mu^\nu_{y(t,\cdot)}$.

Relationship between the particle system (16) and the Euler equation (46). If the mapping $t \mapsto \Xi(t) = (\xi_1(t), \dots, \xi_N(t))$ is a locally Lipschitz solution of (16), then the mapping $t \mapsto y(t)$, where $y(t,x)$ is defined as the moment of order 1 (i.e., by (41)) of the empirical measure $\mu(t) = \mu^E_{(X,\Xi(t))}$ (whose marginal is $\nu^E_X = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$), is a locally Lipschitz solution of (46) where A is associated with ν^E_X. Note that, in this embedding from (16) into (46), we have $y(t,x) = \xi_i(t)$ if $x = x_i$ for $i \in \{1, \dots, N\}$ and 0 otherwise, and ν is purely atomic.

Conversely, if $\nu = \nu^E_X = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and if the mapping $t \mapsto y(t)$ is a locally Lipschitz solution of (46), then the mapping $t \mapsto \Xi(t) = (\xi_1(t), \dots, \xi_N(t))$, with $\xi_i(t) = y(t,x_i)$ for $i \in \{1, \dots, N\}$, is a locally Lipschitz solution of (16). Note that, however, we may have $y(t,x) \neq 0$ for $x \notin \{x_1, \dots, x_N\}$. This is a kind of projection.

This general equivalence works because, when $\nu = \nu^E_X = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and $y(t,x_i) = \xi_i(t)$, the ν-monokinetic measure $\mu^\nu_{y(t,\cdot)}$ coincides with the empirical measure $\mu^E_{(X,\Xi(t))}$. Indeed, we have

$$\mu^\nu_{y(t,\cdot)} = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \delta_{y(t,\cdot)} = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \delta_{y(t,x_i)} = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \delta_{\xi_i(t)} = \mu^E_{(X,\Xi(t))}.$$

3.2 Moment of order 2

We define

$$\boxed{T(t,x) = \frac{1}{d}\int_{\mathbb{R}^d} \|\xi - y(t,x)\|^2\, d\mu_{t,x}(\xi) \qquad \forall x \in \Omega}$$

Note that $T(t,x) = 0$ for ν-almost every $x \in \Omega \setminus \mathrm{supp}(\nu)$. We have

$$\partial_t T(t,x) = \frac{2}{d}\int_{\mathbb{R}^d} \langle \xi - y(t,x), \mathcal{X}[\mu](x,\xi)\rangle\, d\mu_{t,x}(\xi).$$

Proposition 5. *In the opinion propagation model of Example 1, we have* $G(t,x,x',\xi,\xi') = \sigma(x,x')(\xi'-\xi)$ *and*

$$\boxed{\partial_t T(t,x) = -2S(x)T(t,x)}$$

where $S(x) = \int_\Omega \sigma(x,x')\,d\nu(x')$ *for* ν*-almost every* $x \in \Omega$*. Hence* $t \mapsto T(t,x) = T(0,x)e^{-2tS(x)}$ *decreases exponentially to* 0 *as* $t \to +\infty$ *for* ν*-almost every* $x \in \Omega$ *such that* $S(x) > 0$*.*

Proof. We use the disintegration of μ, the fact that $\int_{\mathbb{R}^d}(\xi - y(t,x))\,d\mu_{t,x}(\xi) = 0$ by definition, and the above expression of $\partial_t T(t,x)$. The computation is then straightforward. □

Remark 6. Note that, in [12], convergence to consensus is proved under the assumptions that $d\nu(x) = dx$, that $S(x) \geqslant \delta > 0$ for almost every $x \in \Omega$ and that the (infinite-dimensional) graph associated with σ be strongly connected. Proposition 5 slightly generalizes this result by relaxing the assumption on S to the assumption that $S(x) > 0$ for ν-almost every $x \in \Omega$.

For general functions G, the question to know whether T is the solution of some "closed" equation is open.

In the ν-monokinetic case, i.e., assuming that μ is of the form (45) and is a locally Lipschitz solution of (21), we have $T(t,x) = 0$. This is expected since $T(t,x)$ is the variance and thus measures the distance to the average $y(t,x)$.

3.3 Generalization: coupled equations of moments

More generally, assuming $d = 1$ to simplify, let us set, formally,

$$G(t,x,x',\xi,\xi') = G(t,x,x',y(t,x),y(t,x')) + \sum_{i+j\geqslant 1} g_{ij}(t,x,x')(\xi - y(t,x))^i(\xi' - y(t,x'))^j$$

where $y(t,x) = \int_{\mathbb{R}} \xi\,d\mu_{t,x}(\xi)$ is the moment of order 1 of $\mu_{t,x}$ (recall that the moment of order 0 is $y_0(t,x) = \int_{\mathbb{R}} d\mu_{t,x}(\xi) = 1$). Defining the central moment of order i by

$$y_i(t,x) = \int_{\mathbb{R}} (\xi - y(t,x))^i\,d\mu_{t,x}(\xi) \qquad \forall i \in \mathbb{N}$$

(note that $y_0(t,x) = 1$ and $y_1(t,x) = 0$), we have

$$\begin{aligned}\mathcal{X}[\mu_t](t,x,\xi) &= \int_{\Omega\times\mathbb{R}} G(t,x,x',\xi,\xi')\,d\mu_t(x',\xi') \\ &= \int_\Omega G(t,x,x',y(t,x),y(t,x'))\,d\nu(x') \\ &\qquad + \sum_{i+j\geqslant 1}(\xi - y(t,x))^i \int_\Omega g_{ij}(t,x,x')y_j(t,x')\,d\nu(x')\end{aligned}$$

and thus, using (47),

$$\boxed{\mathcal{X}[\mu_t](x,\xi) = (A(t,y(t)))(x) + \sum_{i+j\geqslant 1}(\xi - y(t,x))^i \int_\Omega g_{ij}(t,x,x')y_j(t,x')\,d\nu(x')}$$

It is interesting to see that, in the above formal expansion of $\mathcal{X}[\mu_t](x,\xi)$ using the centered moments, the first term is $(A(t,y(t)))(x)$.

Therefore, we have

$$\begin{aligned}\partial_t y(t,x) &= \int_{\mathbb{R}} \mathcal{X}[\mu_t](x,\xi)\, d\mu_{t,x}(\xi) \\ &= (A(t,y(t)))(x) + \sum_{i+j\geqslant 1} \left(\int_\Omega g_{ij}(t,x,x') y_j(t,x')\, d\nu(x') \right) y_i(t,x)\end{aligned}$$

(actually since $y_1 = 0$ the above sum can be taken over all pairs (i,j) such that $i+j \geqslant 2$) and, for every $k \in \mathbb{N} \setminus \{0,1\}$,

$$\begin{aligned}\partial_t y_k(t,x) &= \langle \mu_{t,x}, L_{\mathcal{X}[\mu_t]}.(\xi \mapsto (\xi - y(t,x))^k)\rangle - \langle \mu_{t,x}, k(\xi - y(t,x))^{k-1} \partial_t y(t,x)\rangle \\ &= k \int_{\mathbb{R}} (\xi - y(t,x))^{k-1} \left(\mathcal{X}[\mu_t](x,\xi) - \partial_t y(t,x)\right) d\mu_{t,x}(\xi) \\ &= k \int_{\mathbb{R}} (\xi - y(t,x))^{k-1} \left(\mathcal{X}[\mu_t](x,\xi) - \int_{\mathbb{R}} \mathcal{X}[\mu_t](x,\xi')\, d\mu_{t,x}(\xi') \right) d\mu_{t,x}(\xi) \\ &= k \sum_{i+j\geqslant 1} \left(\int_\Omega g_{ij}(t,x,x') y_j(t,x')\, d\nu(x') \right) \int_{\mathbb{R}} (\xi - y(t,x))^{k-1} \left((\xi - y(t,x))^i - y_i(t,x) \right) d\mu_{t,x}(\xi) \\ &= k \sum_{i+j\geqslant 1} \left(\int_\Omega g_{ij}(t,x,x') y_j(t,x')\, d\nu(x') \right) \left(y_{k-1+i}(t,x) - y_{k-1}(t,x) y_i(t,x) \right)\end{aligned}$$

(actually since $y_1 = 0$ the pair $(i=0, j=1)$ does not occur in the above sum). In full generality, all equations of moments are coupled and we have no closed system.

For instance, in the opinion propagation model $G(t,x,x',\xi,\xi') = \sigma(x,x')(\xi' - \xi)$ given by Example 1, i.e., when $g_{ij} = 0$ if $i+j \geqslant 2$ and $g_{01} = -g_{10} = \sigma$, we recover the facts that the equation in y is closed and that $\partial_t y_2(t,x) = -2S(x) y_2(t,x)$.

How to close the hierarchy of equations satisfied by all the moments $y_i(t,x)$, for $i \in \mathbb{N}^*$, will be the subject of future investigations.

Remark 7. In the opinion propagation model of Example 1, $G(t,x,x',\xi,\xi') = \sigma(x,x')(\xi' - \xi)$, a straightforward computation shows that

$$\frac{1}{k} \partial_t y_k(t,x) = -S(x) y_k(t,x) \qquad \forall k \in \mathbb{N} \setminus \{0,1\},$$

thus generalizing the case $k = 2$ studied in Proposition 5. Therefore, $y_k(t,x) = y_k(0,x) e^{-tS(x)}$.

4 From microscopic to macroscopic scale

("from particle to Euler", graph limit)

We have seen in the previous section that the passage from micro to macro is very general and that, instead of considering Riemann sums, one can observe that it results from the coincidence of the empirical measure with the ν-monokinetic measure.

Anyway, in this section we are going to explore the point of view of Riemann sums, in order to derive error estimates mainly resulting from the discrepancy between an integral and a Riemann sum, building on the concept of graph limit introduced in [33].

For second-order systems like the celebrated Cucker-Smale system, the authors of [27] considered the three scales (micro, meso and macroscopic). Recently, in [22], the authors provide

a rigorous derivation from the kinetic Cucker-Smale model to the macroscopic pressureless Euler system by hydrodynamic limit, using entropy methods and deriving error estimates.

By combining the various possibilities to derive the macroscopic quantities, we show how to obtain explicit error estimates for the direct derivation of the graph limit (i.e., the macroscopic model) from the microscopic model by first taking the mean field limit and obtaining the (kinetic) Vlasov equation, and then by taking the hydrodynamic limit. The price to pay is that we obtain estimates in weak topology (Wasserstein distance) instead of estimates in L^2 or L^∞ norm as provided by the graph limit methods, but the gain is to have an explicit $\mathrm{O}(1/N)$ rate of convergence.

We have seen in Section 3 that it is not always possible to pass from the mesoscopic to the macroscopic scale, because the equation obtained for the moment of order 1 may not be closed. However, we have seen in Proposition 3 that, for the opinion propagation model, the equation in y is closed and is linear.

4.1 Convergence estimates for the graph limit

Throughout this section, we assume that Ω is compact. Let $\nu \in \mathcal{P}(\Omega)$. We consider the general nonlinear Euler equation (46), with the nonlinear operator A defined by (47).

We also assume that there exists a family $(\mathcal{A}^N, X^N)_{N\in\mathbb{N}^*}$ of tagged partitions associated with ν (satisfying (12), see Section 1.2), with $\mathcal{A}^N = (\Omega_1, \dots, \Omega_N)$ and $X^N = (x_1, \dots, x_N)$ (we do not write any superscript N in Ω_i and x_i for readability). We have the following two propositions.

Theorem 4. *With the above assumptions and notations, let $y^0 \in L^\infty(\Omega, \mathbb{R}^d)$. On the one part, we consider the unique solution $t \mapsto y(t,\cdot) \in L^\infty(\Omega, \mathbb{R}^d)$ (well defined for every $t \in \mathbb{R}$) of the (nonlinear) Euler equation (46) such that $y(0,\cdot) = y^0(\cdot)$.*

On the other part, for any $N \in \mathbb{N}^*$, we consider the unique solution $t \mapsto \Xi(t) = (\xi_1(t), \dots, \xi_N(t)) \in \mathbb{R}^{dN}$ (well defined for every $t \in \mathbb{R}$) of the particle system (16) such that $\xi_i(0) = y^0(x_i)$ for every $i \in \{1, \dots, N\}$, and we set

$$y_{\Xi(t)}(x) = \sum_{i=1}^{N} \xi_i(t) \mathbb{1}_{\Omega_i}(x) \qquad \forall (t,x) \in \mathbb{R} \times \Omega \tag{48}$$

where $\mathbb{1}_{\Omega_i}$ is the characteristic function of Ω_i, defined by $\mathbb{1}_{\Omega_i}(x) = 1$ if $x \in \Omega_i$ and 0 otherwise.

- *Assume that y^0 is bounded and continuous ν-almost everywhere on Ω (thus, ν-Riemann integrable). Then, for every $t \geqslant 0$, $y(t,\cdot)$ is bounded and continuous ν-almost everywhere on Ω, with the same continuity set as y^0, and*

$$\|y(t,\cdot) - y_{\Xi(t)}(\cdot)\|_{L^\infty(\Omega,\mathbb{R}^d)} = \mathrm{o}(1) \tag{49}$$

  as $N \to +\infty$, where the remainder term $\mathrm{o}(1)$ is uniform with respect to t on any compact interval. In particular,

$$\max_{i\in\{1,\dots,N\}} \|y(t,x_i) - \xi_i(t)\| = \mathrm{o}(1). \tag{50}$$

- *Assume that there exists $\alpha \in (0,1]$ such that $y^0 \in \mathscr{C}^{0,\alpha}(\Omega, \mathbb{R}^d)$ and G is locally α-Hölder continuous with respect to (x, x', ξ, ξ') (uniformly with respect to t on any compact). Then, for every $t \geqslant 0$, $y(t,\cdot) \in \mathscr{C}^{0,\alpha}(\Omega, \mathbb{R}^d)$ with*

$$\mathrm{Hol}_\alpha(y(t,\cdot)) \leqslant e^{tL^y(t)} \left(1 + \mathrm{Hol}_\alpha(y(0,\cdot))\right) \qquad \forall t \geqslant 0 \tag{51}$$

  and, for every $N \in \mathbb{N}^*$,

$$\max_{i\in\{1,\dots,N\}} \|y(t,x_i) - \xi_i(t)\| \leqslant \frac{C_\Omega^\alpha}{N^\alpha} \left(1 + \mathrm{Hol}_\alpha(y^0)\right) e^{2tL(t)} \qquad \forall t \geqslant 0 \tag{52}$$

and actually,

$$\|y(t,\cdot) - y_{\Xi(t)}(\cdot)\|_{L^\infty(\Omega,\mathbb{R}^d)} \leqslant 2\frac{C_\Omega^\alpha}{N^\alpha}\left(1+\mathrm{Hol}_\alpha(y^0)\right)e^{2tL(t)} \qquad \forall t \geqslant 0 \tag{53}$$

where C_Ω is given by (12). The constant $L(t)$ in (52) and (53) is defined by

$$L(t) = \max_{0\leqslant s\leqslant t} \mathrm{Hol}_\alpha(G(s,\cdot,\cdot,\cdot,\cdot)_{|\Omega\times\Omega\times S(s)\times S(s)}) + \max_{\substack{x,x'\in\Omega\\ 0\leqslant s\leqslant t}} \mathrm{Lip}(G(s,x,x',\cdot,\cdot)_{|S(s)\times S(s)}) \tag{54}$$

where $S(s)\subset\mathbb{R}^d$ is the (compact) convex closure of all $\xi_i(s)$ for $i\in\{1,\dots,N\}$ and all $y(s,x)$ for $x\in\Omega$. The constant $L^y(t)$ in (51) is defined as $L(t)$ but with $S(s)$ replaced by $S^y(s)$ that is the convex closure of all $y(s,x)$ for $x\in\Omega$, i.e., like $S(s)$ but without the $\xi_i(s)$. We have $L^y(t)\leqslant L(t)$.

Theorem 4 is proved in Appendix B.5.

Remark 8. Theorem 4 can be extended to the case where Ω is not compact, under the following additional assumptions:

- the family of tagged partitions is such that the points x_i remain in a compact subset of Ω;
- the initial condition y^0 is of compact essential support;
- the set $S(s)\subset\Omega\times\mathbb{R}^d$ is defined as the compact closure of all $(x_i,\xi_i(s))$ for $i\in\{1,\dots,N\}$ and all $(x,y(s,x))$ for $x\in\mathrm{supp}(y(s,\cdot))$ (compact essential support).

The above assumptions imply that $y(t,\cdot)$ is of compact essential support, for every $t\geqslant 0$, and that $L(t)$ is well defined.

Theorem 5. *Given any $N\in\mathbb{N}^*$, let $\Xi_0\in\mathbb{R}^{dN}$. We consider on the one part the unique solution $t\mapsto\Xi(t)=(\xi_1(t),\dots,\xi_N(t))\in\mathbb{R}^{dN}$ (well defined for every $t\in\mathbb{R}$) of the particle system (16) such that $\Xi(0)=\Xi_0$, and we define $y_{\Xi(t)}(x)$ by (48). We consider on the other part the unique solution $t\mapsto y^N(t,\cdot)\in L^\infty(\Omega,\mathbb{R}^d)$ (well defined for every $t\in\mathbb{R}$) of the Euler equation (46) such that $y^N(0,\cdot)=y_{\Xi_0}(\cdot)$ (i.e., $y^N(0,x)=\xi_i(0)$ if $x\in\Omega_i$). Then*

$$\|y^N(t,\cdot) - y_{\Xi(t)}(\cdot)\|_{L^\infty(\Omega,\mathbb{R}^d)} = \mathrm{o}(1) \tag{55}$$

as $N\to+\infty$, where the remainder term $\mathrm{o}(1)$ is uniform with respect to t on any compact interval. If moreover G is locally α-Hölder continuous with respect to (x,x',ξ,ξ') (uniformly with respect to t on any compact), then, for every $N\in\mathbb{N}^*$,

$$\|y^N(t,\cdot) - y_{\Xi(t)}(\cdot)\|_{L^\infty(\Omega,\mathbb{R}^d)} \leqslant 2\frac{C_\Omega^\alpha}{N^\alpha}e^{2tL(t)} \qquad \forall t\geqslant 0, \tag{56}$$

where $L(t)$ is defined by (54).

Theorem 5 is proved in Appendix B.6.

Note that, in particular, taking $x=x_i$ in (56), we have

$$\max_{i\in\{1,\dots,N\}} \|y^N(t,x_i)-\xi_i(t)\| \leqslant 2\frac{C_\Omega^\alpha}{N^\alpha}e^{2tL(t)},$$

which improves the estimates obtained in [4].

Remark 9. In Appendix A.5, we provide estimates on the discrepancy between empirical measures and ν-monokinetic measures. Lemma 24 of that appendix, combined with Theorem 4 and with the proof of that proposition (given below), yields estimates on the discrepancy of the empirical measure $\mu^E_{(X,\Xi(t))}$ with respect to the ν-monokinetic measures $\mu^\nu_{y(t,\cdot)}$ or $\mu^\nu_{y_{\Xi(t)}(\cdot)}$.

Remark 10. The proofs of Propositions 4 and 5 that we provide in Appendices B.5 and B.6 are direct, but actually one can also prove these propositions by applying Theorem 2 with $\mu(t)=\mu^\nu_{y(t,\cdot)}=\nu\otimes\delta_{y(t,\cdot)}$ (the ν-monokinetic measure) and use Lemma 24 of Appendix A.5.

4.2 Additional remarks: from Liouville to Euler

In Section 4.1 we considered the direct passage from the particle system (16) to the Euler (graph limit) equation (46) through the system of ODEs defining the particle dynamics and in the previous Sections 2 and 3 we also reached the same Euler equation via the Vlasov equation in the pure mean field paradigm.

The Liouville equation (29) being the transport equation lifting the particle system (16), a natural question is to wonder whether there exists a direct way to pass from Liouville to Euler. Our objective in this section is to provide a quantity cooked up out of the solution $\rho(\cdot)$ of the Liouville equation (29), converging to the solution of the Euler equation as $N \to +\infty$. The question may fill a gap in the general micro-meso-macroscopic landscapes.

Let us explain how this can be done. Considering a system of N particles, each of them living in a phase space $\Omega \times \mathbb{R}^d$, the meaning of the solution $\rho(t) \in \mathcal{P}(\Omega \times \mathbb{R}^d)$ of the Liouville equation (29) is the following, when it has a density with respect to the Lebesgue measure: for any $X = (x_1, \dots, x_N) \in \Omega^N$ and any $\Xi = (\xi_1, \dots, \xi_N) \in \mathbb{R}^{dN}$, $\frac{d\rho}{dX\, d\Xi}(X, \Xi)$ is the joint probability that the ith particle has position and momentum (x_i, ξ_i), for every $i \in \{1, \dots, N\}$. In Section 2.4 we have shown that, for appropriate initial conditions $\rho(0)$, we recover the mean field limit by taking the limit $N \to +\infty$ of the average over all particles but one and then by taking marginals.

The Liouville paradigm enlarges the moment setting to a probabilistic one: every agent has a moment, but it hesitates randomly between several values that can be assigned to it. Of course the monokinetic case through the Vlasov equation exhausts this random feature by assigning a single moment. But it is quite remarkable, and one has to say still mysterious for us, that, for the opinion propagation model outside monokineticity, the marginal of the full density, namely a probability "average over all particles but one" leads through, and *after* the large N limit, its first moment to the same limit as the fundamentally different "discrete to continuous" passage emblematic to the graph limit.

It is therefore interesting to remove this "after the large N limit" and pass directly from Liouville to Euler and answer the aforementioned question: what is the solution to the graph limit equation the large N limit of? Hereafter, we describe two ways to pass directly from the Liouville equation (29) to the Euler equation (46).

First way. The first way, described below, is in our sense the most interesting because it does not require to consider the Vlasov equation. Given any $\Xi = (\xi_1, \dots, \xi_N) \in \mathbb{R}^{dN}$, we denote $\Xi_{\neq 1} = (\xi_2, \dots, \xi_N) \in \mathbb{R}^{d(N-1)}$. Given any $\rho \in \mathcal{P}(\Omega^N \times \mathbb{R}^{dN})$, we define the measure $\mathcal{M}_1[\rho] \in \mathcal{P}(\Omega^N \times \mathbb{R}^{d(N-1)}$ by

$$\int_{\Omega^N \times \mathbb{R}^{d(N-1)}} f \, d(\mathcal{M}_1[\rho]) = \int_{\Omega^N \times \mathbb{R}^{dN}} f(X, \Xi_{\neq 1}) \xi_1 \, d\rho^s(X, \Xi)$$

for every $f \in \mathscr{C}_c^\infty(\Omega^N \times \mathbb{R}^{d(N-1)})$. The measure $\mathcal{M}_1[\rho]$ is a kind of moment-measure of ρ, standing for the moment expectation of the first agent as a probability on the momenta of all other agents. Then, similarly to the definition of marginals, we define the measure $(\mathcal{M}_1[\rho])_{N:1} \in \mathcal{P}(\Omega)$ by

$$\int_\Omega f(x_1) \, d(\mathcal{M}_1[\rho])_{N:1}(x_1) = \int_{\Omega^N \times \mathbb{R}^{d(N-1)}} f(x_1) \, d(\mathcal{M}_1[\rho])(X, \Xi_{\neq 1})$$

for every $f \in \mathscr{C}_c^\infty(\Omega)$.

Lemma 1. *In the contexts of Theorems 2 or 3, $\mathcal{M}_1[\rho(t)^s]_{N:1}$ converges weakly to $y(t, \cdot)$ as $N \to +\infty$, uniformly with respect to t on compact intervals, with convergence estimates in Wasserstein distance W_1 under additional regularity assumptions on G as in those theorems.*

This lemma shows how to pass directly from Liouville to Euler at the limit $N \to +\infty$.

Second way. We have seen in Section 2.4 how to pass from Liouville to Vlasov, i.e., given any solution $\mu(\cdot)$ of the Vlasov equation (21), choosing adequate initial solutions $\rho(0) = \rho_0$ for the Liouville equation (29), we have established that

$$W_1(\rho(t)^s_{N:1}, \mu(t)) = \sup_{\mathrm{Lip}(f)\leqslant 1} \int_{\Omega\times\mathbb{R}^d} f(x,\xi)\, d\left(\rho(t)^s_{N:1} - \mu(t)\right)(x,\xi)$$

converges to 0 as $N \to +\infty$, with convergence estimates under additional regularity assumptions on G. Taking the particular test function $f(x,\xi) = \xi$, and defining the moment of order 1 of $\rho(t)^s_{N:1} \in \mathcal{P}(\Omega\times\mathbb{R}^d)$ by

$$\rho^N_1(t,x) = \int_{\mathbb{R}^d} \xi\, d\left(\rho(t)^s_{N:1}\right)_x(\xi)$$

for ν-almost every $x \in \Omega$ (where the measures $(\rho(t)^s_{N:1})_x \in \mathcal{P}(\mathbb{R}^d)$ are the disintegration of $\rho(t)^s_{N:1}$ with respect to its marginal on Ω), we thus infer that, in the conditions of Theorems 2 or 3, $x \mapsto \rho^N_1(t,x)$ converges weakly as $N \to +\infty$ to $x \mapsto y(t,x) = \int_{\mathbb{R}^d} \xi\, d\mu_{t,x}(\xi)$, the moment of order 1 of $\mu(t)$, with convergence estimates in Wasserstein distance under additional regularity assumptions on G.

Lemma 2. *In the contexts of Theorems 2 or 3, we have*

$$\rho^N_1(t,x) = \begin{cases} \displaystyle\int_{\Omega^N\times\mathbb{R}^{dN}} \xi_i\, d\rho_t & \text{if } x = x_i \text{ for } i \in \{1,\dots,N\}, \\ 0 & \text{otherwise}, \end{cases} \tag{57}$$

and $x \mapsto \rho^N_1(t,x)$ converges weakly to $x \mapsto y(t,x) = \int_{\mathbb{R}^d} \xi\, d\mu_{t,x}(\xi)$ as $N \to +\infty$ (with convergence estimates in Wasserstein distance W_1).

Proof. In Theorems 2 and 3, we have defined $\rho_0 = \delta_X \otimes \rho_{0,X}$ with $\rho_{0,X} = \delta_{\Xi_0}$ in the first case and $\rho_{0,X} = \mu_{0,x_1} \otimes \cdots \otimes \mu_{0,x_N}$ in the second case. In both cases, we have $\rho(t) = \delta_X \otimes \rho_{t,X}$ and thus, applying Lemma 21 in Appendix A.4, $\rho(t)^s_{N:1} = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \tilde{p}^i_* \rho_{t,X}$ where $\tilde{p}^i(\Xi) = \xi_i$. Therefore, the marginal of $\rho(t)^s_{N:1}$ on Ω is the empirical measure $\nu^E_X = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, and its disintegration with respect to ν^E_X is given by $(\rho(t)^s_{N:1})_x = \tilde{p}^i_* \rho_{t,X}$ if $x = x_i$ for $i \in \{1,\dots,N\}$ and 0 otherwise. The lemma follows. □

The formula (57) shows that, for the choices of $\rho(0)$ done in Theorems 2 and 3, the quantity $\rho^N_1(t,\cdot)$, defined by (57) directly from the solution $\rho(t)$ of the Liouville equation (29), converges weakly to the solution $y(t,\cdot)$ of the Euler equation (46).

Note anyway that, in the context of Theorem 2, the inferred result is quite obvious, because, there, we have $\rho(t) = \delta_X \otimes \delta_{\Xi(t)}$ and $\rho(t)^s_{N:1} = \mu^E_{X,\Xi(t)}$ and thus $(\rho(t)^s_{N:1})_x = \delta_{\xi_i(t)}$ if $x = x_i$ for $i \in \{1,\dots,N\}$ and 0 otherwise. Therefore $\rho^N_1(t,x) = \xi_i(t)$ if $x = x_i$ for $i \in \{1,\dots,N\}$ and 0 otherwise. The convergence results that we can recover here are then particular cases of those obtained in Propositions 4 and 5.

5 Summary: relationships between various scales

In the previous sections, we have investigated the following three scales:

- the microscopic model, which is the particle system

$$\dot{\xi}_i(t) = \frac{1}{N}\sum_{j=1}^N G(t, x_i, x_j, \xi_i(t), \xi_j(t)), \qquad i = 1,\dots,N; \tag{58}$$

when extending this system by setting $\dot{x}_i(t) = 0$, in some sense we perform an extension of the particle system to the phase space;

- the mesoscopic model, which is the (kinetic) Vlasov equation

$$\partial_t \mu + \mathrm{div}_\xi(\mathcal{X}[\mu]\mu) = 0 \tag{59}$$

where $\mathcal{X}[\mu](x,\xi) = \int_{\Omega\times\mathbb{R}^d} G(t,x,x',\xi,\xi')\, d\mu(x',\xi')$ for every $(x,\xi) \in \Omega \times \mathbb{R}^d$, obtained by mean field limit;

- the macroscopic model, which is the Euler equation

$$\partial_t y(t,x) = (A(t,y(t)))(x) = \int_\Omega G(t,x,x',y(t,x),y(t,x'))\, d\nu(x') \tag{60}$$

where $\nu \in \mathcal{P}(\Omega)$, obtained by graph limit.

Additionally, we have also considered the Liouville equation, having a probabilistic interpretation,

$$\partial_t \rho + \mathrm{div}_\Xi(Y\rho) = 0 \tag{61}$$

where Y is a vector field in $\mathbb{R}^{dN}$ representing the system of all particles.

Figure 1 illustrates the various relationships that we have investigated in the paper, and that we comment hereafter.

Particle to Liouville. Any solution $\Xi(\cdot)$ of (58) can also be embedded as a Dirac measure $\rho(\cdot) = \delta_X \otimes \delta_{\Xi(\cdot)}$ that is a solution of the Liouville equation (61).

Particle to Vlasov. By Proposition 1, any solution $\Xi(\cdot)$ of the particle system (58) can be embedded to an empirical measure $\mu(\cdot) = \mu^E_{(X,\Xi(\cdot))} = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \delta_{\xi_i(\cdot)}$ that is a solution of the Vlasov equation (59). Conversely if an empirical measure $\mu(\cdot) = \mu^E_{(X,\Xi(\cdot))}$ is a solution of the Vlasov equation (59) then $\Xi(\cdot)$ must be a solution of (58).

In this context, the mean field limit consists of taking the limit $N \to +\infty$.

Particle to Euler. Any solution $\Xi(\cdot)$ of the particle system (58) can be embedded to a solution of the (nonlinear) Euler equation (60) by setting $\nu = \nu^E_X = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, $y(t,x) = \xi_i(t)$ if $x = x_i$ and 0 otherwise. Conversely, if $\nu = \nu^E_X = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and if $y(\cdot,\cdot)$ is a locally Lipschitz solution of the Euler equation (60), then $\Xi(\cdot) = (\xi_1(\cdot),\dots,\xi_N(\cdot)$, with $\xi_i(\cdot) = y(\cdot,x_i)$ for every $i \in \{1,\dots,N\}$, is a locally Lipschitz solution of the particle system (58). Note that, however, $y(t,x)$ may not be zero for $x \notin \{x_1,\dots,x_N\}$.

Alternatively, to pass from the microscopic to the macroscopic scale, by Propositions 4 and 5, one can also take the graph limit of the particle system and thus obtain the Euler equation, with estimates of convergence as $N \to +\infty$.

Liouville to Vlasov. By Theorems 2 or 3, one can recover the solutions of the Vlasov equation (59) from those of the Liouville equation (61), for some appropriate initial conditions $\rho(0)$, by taking marginals and taking the limit $N \to +\infty$.

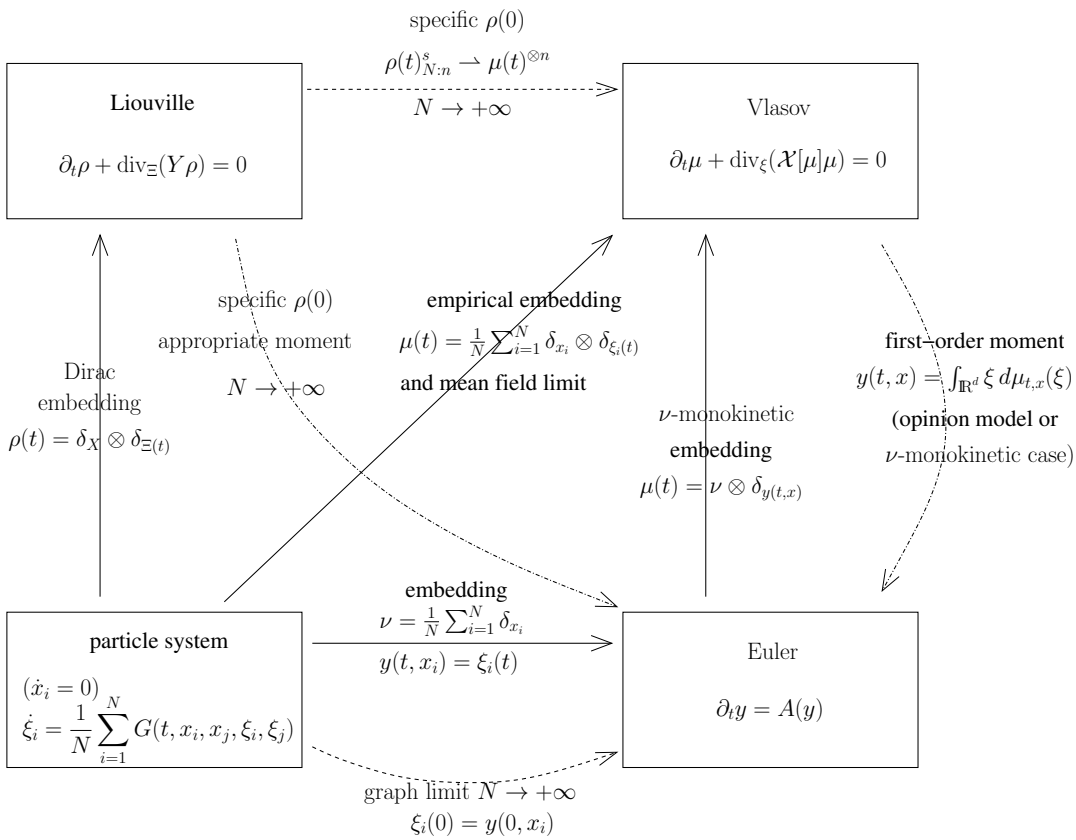

Figure 1: Relationships between particle (microscopic) system, Liouville (probabilistic) equation, Vlasov (mesoscopic, mean field) equation, Euler (macroscopic, graph limit) equation.

Euler to Vlasov. By Section 3.1.3, given any $\nu \in \mathcal{P}(\Omega)$ and any solution $y(\cdot,\cdot)$ of (60), the ν-monokinetic measure $\mu(\cdot) = \mu^\nu_{y(\cdot)} = \nu \otimes \delta_{y(\cdot,\cdot)}$ defined by (45) is a solution of the Vlasov equation (59). This embedding from the macroscopic to the mesoscopic scale is completely general and is valid for the general mean field $\mathcal{X}[\mu]$ defined by (20) and for the general nonlinear operator A defined by (47).

Vlasov to Euler. Here, and only here, we assume, first, that we are in the opinion propagation model of Example 1, i.e., $G(t,x,x',\xi,\xi') = \sigma(x,x')(\xi' - \xi)$. Proposition 3 says that, given any solution $\mu(\cdot)$ of the Vlasov equation (59), defining $\nu = \pi_*\mu(\cdot)$ (marginal of $\mu(\cdot)$, which does not depend on t), the moment y of order 1, defined by $y(t,x) = \int_{\mathbb{R}^d} \xi \, d\mu_{t,x}(\xi)$, is a solution of the Euler equation (60) (which is linear in this case).

As discussed in Section 3.1.3, there is a second way, still not general, of passing from Vlasov to Euler, by assuming that the solution $\mu(\cdot)$ of the Vlasov equation is ν-monokinetic. In this case, its first moment y is solution of the nonlinear Euler equation (46).

This projection from the mesoscopic to the macroscopic scale is not general because, in general, y does not satisfy a closed equation.

Liouville to Euler. Lemmas 2 or 1 in Section 4.2 show how to pass from Liouville to Euler, for specific initial conditions $\rho(0)$, by taking an adequate moment of $\rho(t)$ and then passing to the limit $N \to +\infty$.

All in all, above, all relationships are general (i.e., valid for a general interaction mapping G) except the transition from the mesoscopic (kinetic, mean field) model to the macroscopic (Euler) model, which is valid for the opinion propagation model but fails in general. The graph limit procedure is of a different nature and rather relies on the usual limit in Riemann integration theory, as explained in Section 4.

Anyway, what is interesting in the above arguments is that it may not be relevant to place the mesoscopic level in-between the microscopic level and the macroscopic one.

6 Finite particle approximations of evolution equations

6.1 Setting

Considering the metric space (Ω, d_Ω) of Assumption $(\mathbf{G_1})$, throughout this section, we assume either that:

$(\mathbf{O_1})$ Ω is the compact closure of a bounded open subset of $\mathbb{R}^n$ with a Lipschitz boundary, d_Ω is the induced Euclidean distance, and ν is the restriction to Ω of the Lebesgue measure of $\mathbb{R}^n$;

or that:

$(\mathbf{O_2})$ Ω is a smooth compact Riemannian manifold of dimension n, d_Ω is its Riemannian distance, and ν is the canonical Riemannian measure.

In the case $(\mathbf{O_1})$, Ω is usually called a Lipschitz compact domain of $\mathbb{R}^n$ (of interior denoted by $\mathring{\Omega}$). In the case $(\mathbf{O_2})$, for example Ω may be the sphere or the torus of dimension n.

Without loss of generality, we assume that the volume $|\Omega|$ of Ω is equal to 1, so that ν, hereafter, is the probability Lebesgue measure on Ω (with $\frac{d\nu}{dx} = 1$ in local coordinates).

Note that, under $(\mathbf{O_1})$ or $(\mathbf{O_2})$, tagged partitions associated with ν (satisfying (12), see Section 1.2) always exist for any $N \in \mathbb{N}^*$.

Let X be a Banach functional space containing $\mathscr{C}_c^\infty(\Omega, \mathbb{R}^d)$ as a dense subspace (for instance, $X = H^k(\Omega, \mathbb{R}^d)$ for some $k \in \mathbb{Z}$; but $L^\infty(\Omega, \mathbb{R}^d)$ is not allowed). We consider the abstract quasilinear evolution equation

$$\boxed{\partial_t y(t,x) = A(t, y(t,x))y(t,x)} \tag{62}$$

written for short as $\dot{y}(t) = A(t, y(t))y(t)$, where, for all $t \geqslant 0$ and $\xi \in \mathbb{R}^d$, $A(t,\xi)$ is a linear operator on X, continuous from $\mathscr{C}_c^\infty(\Omega, \mathbb{R}^d)$ to the distributional space $\mathscr{C}_c^\infty(\Omega, \mathbb{R}^d)'$. The latter (very general) assumption ensures that $A(t,\xi)$ has a Schwartz kernel $[A(t,\xi)] \in \mathscr{C}_c^\infty(\Omega \times \Omega, \mathbb{R}^d)'$, i.e., $A(t,\xi)y(x) = \int_\Omega [A(t,\xi)](x,x')y(x')$ for every $y \in \mathscr{C}_c^\infty(\Omega, \mathbb{R}^d)$ where the integral is understood as $\langle [A(t,z)](x,\cdot), y\rangle$ with the distribution bracket.

Assuming that $t \mapsto y(t) \in X$ is a solution of (62), since $y(t)$, as an element of the functional space X, is a function on Ω, in the sequel we denote indifferently $y(t)(x) = y(t,x)$ for all $t \geqslant 0$ and $x \in \Omega$.

Our objective is to prove that, under appropriate assumptions, solutions $t \mapsto y(t)$ of (62) can be approximated by the solutions of a family of finite particle systems.

Particle approximations are well known for some classes of PDEs, like fluid equations: for fluid Euler or Navier-Stokes equations, one often speaks of "fluid particles", in accordance with the

classical Eulerian or Lagrangian viewpoints. In this section, we show that particle approximations can be achieved for much more general PDEs and even for abstract evolution systems like (62).

The idea relies on Theorem 4 in Section 4.1, which shows that the solutions of Euler equations (46), with A defined by (47) with a continuous interaction mapping G, can be approximated with the solutions of the family of particle systems (16) (indexed by N) corresponding to G. A linear PDE, with an unbounded operator, cannot be written as the Euler equation (46) because, with a continuous mapping G, one cannot generate by (47) an unbounded operator A. This is why we are going to introduce, in addition to the parameter N, another (small) parameter ε, in order to approximate unbounded operators A by a family of bounded operators A_ε to which we can then apply the particle approximation result of Theorem 4.

Preliminaries and strategy. We have seen in Proposition 3 in Section 3.1.1 that, taking $G(t,x,x',\xi,\xi') = \sigma(x,x')(\xi'-\xi)$ (opinion propagation model), we obtain the linear Euler equation (43) with A defined by (44).

Actually, if $G(t,x,x',\xi,\xi') = \sigma(x,x')\xi'$, setting $y(x') = \int_{\mathbb{R}^d} \xi'\, d\mu_{x'}(\xi')$ as in Section 3 (moment of order 1 of μ), the mean field $\mathcal{X}[\mu](t,x,\xi)$, defined by (20), does not depend on (t,ξ) and is given by

$$\mathcal{X}[\mu](x) = \int_\Omega \sigma(x,x')y(x')\, d\nu(x') = (Ay)(x)$$

which thus defines the Hilbert-Schmidt operator A of kernel σ with respect to ν. Following Section 3, if $t \mapsto \mu(t)$ is solution of the Vlasov equation (21) then its first-order moment $t \mapsto y(t,\cdot)$ is solution of the linear Euler equation $\partial_t y = Ay$.

The above operator A is bounded, but replacing σ with a general distributional Schwartz kernel $[A]$ and having in mind the Schwartz kernel theorem, one is led to consider a general linear operator $Ay(x) = \int_\Omega [A](x,x')y(x')$. For instance if $[A](x,x') = \delta'_x$, the distributional derivative of the Dirac measure δ_x at x, then $A = -\partial_x$. The differential equation $\partial_t y = Ay$ is then the transport equation $\partial_t y + \partial_x y = 0$.

Let us use the above example as a paradigm to approximate arbitrary unbounded operators, by designing a sufficiently smooth approximation σ_ε of an arbitrary Schwartz kernel $[A]$. Following this idea, we set

$$G_\varepsilon(t,x,x',\xi,\xi') = \sigma_\varepsilon(t,x,x',\xi)\,\xi' \tag{63}$$

where $\varepsilon > 0$ is a small parameter. Recalling that ν is here the probability Lebesgue measure on Ω, given any $\varepsilon \in (0,1]$ we consider the Euler equation corresponding to (63), given by

$$\dot{y}_\varepsilon(t) = A_\varepsilon(t, y_\varepsilon(t))y_\varepsilon(t) \tag{64}$$

where

$$A_\varepsilon(t,\xi)y(x) = \int_\Omega \sigma_\varepsilon(t,x,x',\xi)f(x')\, dx' \qquad \forall f \in \mathscr{C}^\infty(\Omega) \qquad \forall (t,x,\xi) \in \mathbb{R}\times\Omega\times\mathbb{R}^d. \tag{65}$$

In what follows we are going to design an adequate interaction function σ_ε such that A_ε defined by (65) converges to A defined by (75) as $\varepsilon \to 0$, in an appropriate sense. For instance, if $\sigma_\varepsilon(x,\cdot)$ is a smooth function (not depending on (t,ξ)) approximating the distributional derivative δ'_x of the Dirac measure at x, then $A_\varepsilon \to A = -\partial_x$ and hence at the limit $\varepsilon \to 0$ we recover the transport equation $\partial_t y + \partial_x y = 0$.

Given any $\varepsilon \in (0,1]$, let us now introduce the particle approximation of (64). Let $(\mathcal{A}^N, X^N)_{N\in\mathbb{N}^*}$ be a family of tagged partitions associated with ν (satisfying (12), see Section 1.2), with $\mathcal{A}^N =$

$(\Omega_1^N, \dots, \Omega_N^N)$ and $X^N = (x_1^N, \dots, x_N^N)$. We consider the particle system corresponding to (63), given for every $N \in \mathbb{N}^*$ by

$$\boxed{\dot{\xi}_{\varepsilon,i}^N(t) = \frac{1}{N} \sum_{i=1}^N \sigma_\varepsilon(t, x_i^N, x_j^N, \xi_{\varepsilon,i}^N(t)) \xi_{\varepsilon,j}^N(t)} \tag{66}$$

Denoting by $t \mapsto \Xi_\varepsilon^N(t) = (\xi_{\varepsilon,1}^N(t), \dots, \xi_{\varepsilon,N}^N(t))$ an arbitrary solution (well defined and smooth on $\mathbb{R}$) of (80), we set

$$y_\varepsilon^N(t,x) = y_{\Xi_\varepsilon^N(t)}(x) = \sum_{i=1}^N \xi_{\varepsilon,i}^N(t) \mathbb{1}_{\Omega_i^N}(x) \qquad \forall (t,x) \in \mathbb{R} \times \Omega. \tag{67}$$

Note that y_ε^N is $\mathscr{C}^\infty$ in t and piecewise constant in x, and that $y_\varepsilon^N(t, x_i^N) = \xi_{\varepsilon,i}^N(t)$ for every $i \in \{1, \dots, N\}$.

The particle system (66) is expected to provide a particle approximation of the evolution equation (62), in the sense that it is expected that solutions y of (62) are limits of y_ε^N as $N \to +\infty$ and $\varepsilon \to 0$.

However, since the particle system (66) does not have any (classical) limit as $\varepsilon \to 0$, in order to derive convergence estimates we will have to let N tend to $+\infty$ and ε to 0 at some appropriate scale. Our strategy will be in two steps:

1. For any ε fixed, by Theorem 4 in Section 4.1, we obtain a convergence estimate as $N \to +\infty$ for the particle approximation (66) of the solutions y_ε of the "ε-Euler" equation (64), with constants keeping track of the dependence with respect to ε, of the form, for any $T > 0$ fixed,

$$\|y_\varepsilon(t) - y_\varepsilon^N(t)\|_{L^\infty(\Omega, \mathbb{R}^d)} \leqslant \frac{C}{N} e^{2tL_\varepsilon(t)} \qquad \forall t \in [0,T]$$

   for some constant C not depending on (ε, N), where $L_\varepsilon(t)$ is related to the Lipschitz constant of G_ε (defined by (63)) and thus to that of σ_ε. The dependence with respect to ε is thus encoded in the definition of the function σ_ε that approximates the Schwartz kernel of A.

2. Derive a convergence estimate of solutions y_ε of (64) to solutions y of (76): in what follows we do this in two possible ways. A first way is to use semigroup theory and its generalizations to the quasilinear case, in order to obtain estimates from the Duhamel formula; in this way, we will obtain convergence estimates in L^2 norm. A second way, without any semigroup assumption, is to perform a monokinetic lift of (64) to a Vlasov equation (see Proposition 4 in Section 3.1.3) and then to use stability properties for Vlasov equations; in this way, we will obtain convergence estimates in Wasserstein distance W_1.

The convergence estimates of particle solutions of (66) to solutions of (76) are then obtained by the triangular inequality. They depend on N and of ε, but as already alluded the estimates blow up when $\varepsilon \to 0$ with N being fixed and thus the limits must be done with some appropriate scaling.

Structure of the section. In Section 6.2 we give a general particle approximation result for the abstract evolution equation (62). To clarify the exposition we first treat the case where the operator A does not depend on (t, y), under the assumption that A generates a C_0 semigroup. This is already a very wide and interesting class. The main result is Theorem 6. We then show how to extend it to the non-autonomous case and to the quasilinear case under appropriate assumptions that generalize the concept of C_0 semigroup.

Although very general, these results are however abstract and not constructive. In Section 6.3, we consider quasilinear PDEs, with classical differential operators, and thanks to convolution we design explicit finite particle approximations. In this section, the main result is Theorem 7.

Finally, in Section 6.3.2, we extend our results to the case where A does not necessarily generate a semigroup. The price to pay is that, instead of obtaining error estimates in L^2 norm, we derive error estimates in Wasserstein distance W_1.

6.2 A general abstract result within semigroup theory

6.2.1 The autonomous case

Semigroup assumption. We assume that the operator A on the Banach space X is defined on a dense domain $D(A) \subset X$ and generates a C_0 semigroup $(e^{tA})_{t\geqslant 0}$ in X. In particular, there exist $M \geqslant 1$ and $\beta \geqslant 0$ such that

$$\|e^{tA}\|_{L(X)} \leqslant M e^{\beta t} \qquad \forall t \geqslant 0. \tag{68}$$

Given $y^0 \in D(A)$, there exists a unique solution $y \in C^0([0,+\infty), D(A)) \cap C^1((0,+\infty), X)$ of (78) such that $y(0) = y^0$, which is $y(t) = e^{tA} y^0$ (see [20]). In what follows we are going to approximate this solution y by finite particles.

Bounded operator approximation. We assume that there exists a family of bounded operators A_ε on X, indexed by $\varepsilon \in (0,1]$, satisfying the following properties:

- Denoting by e^{tA_ε} the usual exponential of a bounded operator, we have

$$\|e^{tA_\varepsilon}\|_{L(X)} \leqslant M e^{\beta t} \qquad \forall t \geqslant 0 \qquad \forall \varepsilon \in (0,1]. \tag{69}$$

- For every $T > 0$, there exist a Banach subspace $Z \subset X$, dense in X, $C_A > 0$ and a continuous function $\chi : [0,1] \to [0,+\infty)$, satisfying $\chi(0) = 0$, such that

$$\|(A_\varepsilon - A)y(t)\|_X \leqslant C_A\, \chi(\varepsilon) \|y(t)\|_Z \qquad \forall t \in [0,T] \qquad \forall \varepsilon \in (0,1]. \tag{70}$$

- For any $\varepsilon \in (0,1]$, the (density of the) Schwartz kernel σ_ε of A_ε (defined by (65)), is Lipschitz continuous on $\Omega \times \Omega$.

An example of bounded approximation operator A_ε satisfying (69), not explicit but fully general, is given by the *Yosida approximant*

$$A_\varepsilon = J_\varepsilon A \qquad \text{where} \qquad J_\varepsilon = \frac{1}{\varepsilon}\left(\frac{1}{\varepsilon}\mathrm{id} - A\right)^{-1},$$

which satisfies (69) (see [20, 36]. The assumption (70) entails two things: first, it requires that $y(t,\cdot)$ be sufficiently regular; second, it refers to convergence estimates, which are often proved by explicit approximation constructions (see also [30] for finite-dimensional approximations with error estimates), as we will do hereafter. For instance when $X = L^2(\Omega, \mathbb{R}^d)$, the Banach space Z may be a subspace of functions of X having a certain number of bounded derivatives. The third assumption is related to the regularity of A_ε; for instance if A is a differential operator then it is satisfied by iterating the Yosida approximation, taking $A_\varepsilon = J_\varepsilon^j A$ for j large enough. Of course, if A is unbounded then $\|\sigma_\varepsilon\|_{L^\infty(\Omega^2)} + \mathrm{Lip}(\sigma_\varepsilon) \to +\infty$ as $\varepsilon \to 0$. All in all, the first assumption (69) (which is classical in the Trotter-Kato theorem) is the most stringent; it is usually established in practice by means of dissipativity properties, and this is also what we will do in the explicit construction hereafter.

Theorem 6. *Let $T > 0$ be arbitrary. In addition to the above semigroup and bounded approximation operator assumptions, we assume that $y^0 = y(0)$ is Lipschitz continuous on Ω and that $y \in L^1([0,T], Z)$. For any $\varepsilon \in (0,1]$ and any $N \in \mathbb{N}^*$, let $t \mapsto \Xi_\varepsilon^N(t) = (\xi_{\varepsilon,1}^N(t), \dots, \xi_{\varepsilon,N}^N(t))$ be the unique solution of (80) such that $\xi_{\varepsilon,i}^N(0) = y^0(x_i^N)$ for every $i \in \{1, \dots, N\}$, and let y_ε^N be defined by (67). If there exists a continuous and dense embedding $L^\infty(\Omega, \mathbb{R}^d) \hookrightarrow X$, i.e., if there exists $C_\infty > 0$ such that $\|z\|_X \leqslant C_\infty \|z\|_{L^\infty}$ for any $z \in L^\infty(\Omega, \mathbb{R}^d)$, then*

$$\|y_\varepsilon^N - y\|_{\mathscr{C}^0([0,T],X)} \leqslant C_A M e^{\beta T} \chi(\varepsilon) \|y\|_{L^1([0,T],Z)} + \frac{2}{N} C_\Omega C_\infty \exp\left(2T\|\sigma_\varepsilon\|_{L^\infty} + 2T\|y^0\|_{L^\infty} \operatorname{Lip}(\sigma_\varepsilon) e^{2T\|\sigma_\varepsilon\|_{L^\infty}}\right) \quad (71)$$

for every $N \in \mathbb{N}^*$ and every $\varepsilon \in (0,1]$.

Proof. Given any $\varepsilon \in (0,1]$, since A_ε is bounded, there exists a unique locally Lipschitz (in t) solution of (64), i.e., of $\partial_t y_\varepsilon = A_\varepsilon y_\varepsilon$, such that $y_\varepsilon(0,\cdot) = y(0,\cdot) = y^0(\cdot)$, which is given by $y_\varepsilon(t) = e^{tA_\varepsilon} y^0$.

The proof, which is easy, is done in three steps: in the first, we establish that y_ε converges to y; in the second, we apply Theorem 4 (see Section 4.1) to prove that y_ε is approximated by the solutions of the particle system (66); we conclude by the triangular inequality.

First step: convergence of y_ε towards y. Given any $\varepsilon \in (0,1]$, writing that $\partial_t(y_\varepsilon - y) = A_\varepsilon y_\varepsilon - Ay = A_\varepsilon(y_\varepsilon - y) + (A_\varepsilon - A)y$ and integrating, we obtain (Duhamel formula)

$$y_\varepsilon(t) - y(t) = \int_0^t e^{(t-s)A_\varepsilon}(A_\varepsilon - A)y(s)\, ds,$$

and using the Jensen inequality ($\|\cdot\|_X$ is convex) and the estimates (69) and (70), we infer that

$$\|y_\varepsilon(t) - y(t)\|_X \leqslant \int_0^t \left\| e^{(t-s)A_\varepsilon}(A_\varepsilon - A)y(s) \right\|_X ds \leqslant \int_0^t M e^{\beta(t-s)} \|(A_\varepsilon - A)y(s)\|_X\, ds \leqslant C_A M e^{\beta t} \chi(\varepsilon) \|y\|_{L^1([0,t],Z)} \quad (72)$$

for every $t \in [0,T]$.

Second step: particle approximation. Given any $\varepsilon \in (0,1]$, since $y_\varepsilon(0) = y^0 \in L^\infty(\Omega, \mathbb{R}^d)$ and since $\partial_t y_\varepsilon(t,\cdot) = \int_\Omega \sigma_\varepsilon(\cdot, x') y_\varepsilon(t,x')\, dx'$ and $\dot{\xi}_{\varepsilon,i}^N(t) = \frac{1}{N}\sum_{j=1}^N \sigma_\varepsilon(x_i^N, x_j^N)\xi_{\varepsilon,j}^N(t)$, estimating roughly and integrating we get that

$$\max\left(\|\Xi_\varepsilon^N(t)\|_\infty, \|y_\varepsilon(t)\|_{L^\infty(\Omega,\mathbb{R}^d)}\right) \leqslant e^{t\|\sigma_\varepsilon\|_{L^\infty(\Omega^2)}} \|y^0\|_{L^\infty(\Omega,\mathbb{R}^d)} \qquad \forall t \geqslant 0. \quad (73)$$

Since y^0 is Lipschitz continuous, it follows from the second item of Theorem 4 in Section 4.1 that, for every $t \geqslant 0$, $y_\varepsilon(t)$ is Lipschitz continuous on Ω, with $\operatorname{Lip}(y_\varepsilon(t)) \leqslant e^{tL_\varepsilon(t)}(1 + \operatorname{Lip}(y^0))$ and

$$\|y_\varepsilon^N(t) - y_\varepsilon(t)\|_{L^\infty(\Omega,\mathbb{R}^d)} \leqslant \frac{2C_\Omega}{N}(1 + \operatorname{Lip}(y^0)) e^{2tL_\varepsilon(t)} \qquad \forall t \geqslant 0, \quad (74)$$

where, using (73) and the particular form of the mapping G_ε defined by (63), $L_\varepsilon(t)$ is defined by

$$L_\varepsilon(t) = \|\sigma_\varepsilon\|_{L^\infty} + \operatorname{Lip}(\sigma_\varepsilon) e^{t\|\sigma_\varepsilon\|_{L^\infty}} \|y^0\|_{L^\infty}.$$

Conclusion. We have $\|y_\varepsilon^N(t) - y_\varepsilon(t,\cdot)\|_X \leqslant C_\infty \|y_\varepsilon^N(t) - y_\varepsilon(t)\|_{L^\infty}$. Using the triangular inequality, we infer from (72) and (74) that, for every $t \in [0,T]$ and every $\varepsilon \in (0,1]$,

$$\|y(t) - y_\varepsilon^N(t)\|_X \leqslant C_A M e^{\beta T} \chi(\varepsilon) \|y\|_{L^1([0,T],Z)} + \frac{2}{N} C_\Omega C_\infty (1 + \mathrm{Lip}(y^0)) e^{2T L_\varepsilon(T)}$$

and (71) follows, using (73). □

Comments on Theorem 6. To illustrate and understand the convergence estimate (71), let us assume that $\chi(\varepsilon) \sim \varepsilon$ and that $\|\sigma_\varepsilon\|_{L^\infty} + \mathrm{Lip}(\sigma_\varepsilon) \sim \frac{1}{\varepsilon^k}$ for some $k \in \mathbb{N}^*$ (this will be the case in the explicit construction hereafter). Then, ignoring constants, the right-hand side of (71) is of the order of

$$\varepsilon + \frac{1}{N} \exp\left(\frac{1}{\varepsilon^k} \exp\left(\frac{1}{\varepsilon^k}\right)\right).$$

In order to pass to the limit as $N \to +\infty$ and $\varepsilon \to 0$, it is appropriate to choose parameters such that this term tends to 0 An optimization argument shows that the best choice for ε in function of N is

$$\varepsilon_N \sim \left(\frac{1}{\ln \ln N}\right)^{\frac{1}{k}}$$

as $N \to +\infty$, and in this case the estimate (71) gives

$$\|y_{\varepsilon_N}^N - y\|_{\mathscr{C}^0([0,T],L^2(\Omega,\mathbb{R}^d))} \leqslant \mathrm{Cst} \left(\frac{\mathrm{Cst}}{\ln \ln N}\right)^{\frac{1}{k}} (1 + \mathrm{o}(1)).$$

Such estimates are reminiscent of those found in [9] concerning the linear Boltzmann and the (hydrodynamic limit) heat equation.

The above double exponential (or double logarithm) is a general estimate that can be improved under additional assumptions. Indeed, the estimate (73) is very rough and can be improved for example if the norm of e^{tA_ε} as a bounded linear operator on $L^\infty(\Omega)$ is uniformly bounded with respect to ε on compact intervals of time.

6.2.2 Extensions

Non-autonomous case. In the previous section, we have considered a linear *autonomous* operator A. Theorem 6 can be straightforwardly extended to the case where the operator $A(t)$ depends on t, under the assumption that the family of operators $(A(t))_{0 \leqslant t \leqslant T}$ satisfies the requirements of [36, Chapter 5, Section 5.3, Theorem 3.1] (see also [20, Chapter VI, Section 9.a]), which are standard assumptions in order to extend classical results for semigroups to the instationary case. In particular, it is required that there exists $\beta \in \mathbb{R}$ such that $A(t) - \beta\,\mathrm{id}$ is m-dissipative for every $t \in [0,T]$.

In the proof, the semigroups e^{tA} and e^{tA_ε} are replaced with the respective (so-called) evolution systems $U(t,s)$ and $U_\varepsilon(t,s)$, defined by $\partial_t U(t,s) = A(t)U(t,s)$ for $0 \leqslant s \leqslant t \leqslant T$, $U(s,s) = \mathrm{id}$ and $U(t,r)U(r,s) = U(t,s)$ (same for U_ε). Moreover, we have $\|U(t,s)\|_{L(L^2(\Omega,\mathbb{R}^d))} \leqslant M e^{\beta(t-s)}$. The assumption that $\|U_\varepsilon(t,s)\|_{L(L^2(\Omega,\mathbb{R}^d))} \leqslant M e^{\beta(t-s)}$, for any $\varepsilon \in (0,1]$, is the instrumental fact in the proof of the estimate (72) inferred from the Duhamel formula.

Quasilinear case. Finally, Theorem 6 can be extended to the general case of the quasilinear evolution equation (62), by assuming that the family of operators $A(t,y)$, for $0 \leqslant t \leqslant T$ and $y \in X$, satisfies the requirements of [36, Chapter 6, Section 6.4, Theorems 4.3 or 4.5]. As previously, then, we deal with evolution systems $U(t,s,y)$ and $U_\varepsilon(t,s,y_\varepsilon)$, we still have the Duhamel formula and the instrumental assumption $\|U_\varepsilon(t,s,y_\varepsilon)\|_{L(L^2(\Omega,\mathbb{R}^d))} \leqslant M e^{\beta(t-s)}$.

6.3 Application to PDEs and explicit construction

In local coordinates x on Ω, we denote $D^\alpha = \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}$ where ∂_i is the partial derivative with respect to the ith variable of x (which we do not denote by x_i because the notation is already used for the tagged partitions), where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ and we set $|\alpha| = \sum_{i=1}^n \alpha_i$.

Let $p \in \mathbb{N}^*$ be arbitrary. For every $\alpha \in \mathbb{N}^n$ such that $|\alpha| \leqslant p$, let $a_\alpha \in L^\infty(\mathbb{R} \times \Omega \times \mathbb{R}^d)$ be a function of (t, x, ξ). Throughout the section, we assume that $X = L^2(\Omega, \mathbb{R}^d)$ and that

$$A(t, \xi) = \sum_{|\alpha| \leqslant p} a_\alpha(t, \cdot, \xi)\, D^\alpha \qquad \forall (t, \xi) \in \mathbb{R} \times \mathbb{R}^d, \tag{75}$$

i.e., we consider the quasilinear partial differential equation

$$\boxed{\partial_t y(t, x) = \sum_{|\alpha| \leqslant p} a_\alpha(t, x, y(t, x)) D^\alpha y(t, x) = A(t, y(t, x)) y(t, x)} \tag{76}$$

with some prescribed conditions at the boundary of Ω when there is a boundary. As an application of Theorem 6, our objective is to prove that, under appropriate assumptions, the solutions of (76) can be approximated by the solutions of a family of finite particle systems, of which we design an explicit construction.

6.3.1 The autonomous case

In this section, we assume that the coefficients a_α in (75) do not depend on (t, ξ), i.e., $a_\alpha \in L^\infty(\Omega)$ for every $\alpha \in \mathbb{N}^n$ such that $|\alpha| \leqslant p$. The operator A defined by (75) is then a classical differential operator, given by

$$A = \sum_{|\alpha| \leqslant p} a_\alpha\, D^\alpha \tag{77}$$

and (76) is then the linear autonomous PDE

$$\partial_t y = Ay = \sum_{|\alpha| \leqslant p} a_\alpha\, D^\alpha y. \tag{78}$$

For every $k \in [1, +\infty]$, we denote by $W^{p,k}(\Omega)$ the Sobolev space of functions f on Ω whose partial (distributional) derivatives up to order p are identified with functions of $L^k(\Omega)$, endowed with the norm

$$\|f\|_{W^{p,k}(\Omega)} = \max_{|\alpha| \leqslant p} \|D^\alpha f\|_{L^k(\Omega)}.$$

For $k = 2$, we denote $H^p(\Omega) = W^{p,2}(\Omega)$.

Semigroup assumption. We assume that the operator A on $L^2(\Omega, \mathbb{R}^d)$ is defined on a domain $D(A) \subset H^p(\Omega, \mathbb{R}^d)$, dense in $L^2(\Omega, \mathbb{R}^d)$, which may encode some Dirichlet or Neumann like boundary conditions, maybe of higher order, and that there exists $\beta \geqslant 0$ such that $A - \beta\,\mathrm{id}$ generates a C_0 semigroup of contractions in $L^2(\Omega, \mathbb{R}^d)$. Hence we have $M = 1$ in the inequality (68) and, by the Lumer-Phillips theorem, $A - \beta\,\mathrm{id}$ is m-dissipative (see [20]), which means that it is dissipative (i.e., $\langle (A - \beta\,\mathrm{id}) f, f \rangle_{L^2(\Omega)} \leqslant 0$ for every $f \in D(A)$) and that $\mathrm{Ran}((\beta + 1)\,\mathrm{id} - A) = ((\beta + 1)\,\mathrm{id} - A) D(A) = L^2(\Omega)$.

Particle approximation. Let $\eta \in \mathscr{C}_c^\infty(\mathbb{R}^n)$ be a nonnegative symmetric smooth function on $\mathbb{R}^n$, of compact support contained in the closed unit ball $\overline{B}(0,1)$, such that $\int_{\mathbb{R}^n} \eta(x)\,dx = 1$. Here, symmetric means that $\eta(x) = \eta(-x)$ for every $x \in \mathbb{R}^n$. We set $C_\eta = \int_{\mathbb{R}^n} \|x\|\eta(x)\,dx$. For example, we can take

$$\eta(x) = \begin{cases} c\,e^{1/(\|x\|^2-1)} & \text{if } \|x\| < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a normalization constant. Given any $\varepsilon \in (0,1]$, we denote by $\eta_\varepsilon \in \mathscr{C}_c^\infty(\mathbb{R}^n)$ the (mollifier) function given by

$$\eta_\varepsilon(x) = \frac{1}{\varepsilon^n}\eta\left(\frac{x}{\varepsilon}\right) \qquad \forall x \in \mathbb{R}^n$$

and we define the function σ_ε (not depending on (t,ξ)) on $\Omega \times \Omega$ by

$$\sigma_\varepsilon(x,x') = \sum_{|\alpha|\leqslant p} \int_\Omega \eta_\varepsilon(x-z)a_\alpha(z)(D^\alpha\eta_\varepsilon)(z-x')\,dz \qquad \forall x,x' \in \Omega\times\Omega. \tag{79}$$

We have $\sigma_\varepsilon \in \mathscr{C}^\infty(\Omega\times\Omega)$ (it is smooth up to the boundary). The mapping G_ε is defined by (63); it does not depend on (t,ξ).

Given any $\varepsilon \in (0,1]$, we consider the family of particle systems (66), indexed by N, associated with the family $(\mathcal{A}^N, X^N)_{N\in\mathbb{N}^*}$ of tagged partitions. Here, for any $N \in \mathbb{N}^*$, the particle system (66) is autonomous and is written as

$$\dot{\xi}^N_{\varepsilon,i}(t) = \frac{1}{N}\sum_{i=1}^N \sigma_\varepsilon(x_i^N, x_j^N)\xi^N_{\varepsilon,j}(t). \tag{80}$$

Theorem 7. *We assume that the coefficients a_α of A, in (77), are Lipschitz continuous on Ω, i.e., are in $W^{1,\infty}(\Omega)$.*

Let $T > 0$. We assume that $y \in L^1([0,T], W^{p+1,\infty}(\Omega, \mathbb{R}^d))$ is a solution of $\partial_t y = Ay$ such that $y(0,\cdot) \in \mathrm{Lip}(\Omega, \mathbb{R}^d)$.

For any $\varepsilon \in (0,1]$ and any $N \in \mathbb{N}^*$, let $t \mapsto \Xi^N_\varepsilon(t) = (\xi^N_{\varepsilon,1}(t), \dots, \xi^N_{\varepsilon,N}(t))$ be the unique solution of (80) such that $\xi^N_{\varepsilon,i}(0) = y(0,x_i^N)$ for every $i \in \{1,\dots,N\}$, and let y^N_ε be defined by (67).

Then, there exists $C > 0$ such that

$$\|y^N_\varepsilon - y\|_{\mathscr{C}^0([0,T],L^2(\Omega,\mathbb{R}^d))} \leqslant C\left(\varepsilon + \frac{1}{N}\exp\left(\frac{C}{\varepsilon^{n+p+1}}\exp\left(\frac{C}{\varepsilon^{n+p}}\right)\right)\right) \qquad \forall N \in \mathbb{N}^* \quad \forall \varepsilon \in (0,1], \tag{81}$$

except in case ($\mathbf{O_1}$) when $n = 1$, in which case the first term ε in the parenthesis at the right-hand side of (81) must be replaced with $\sqrt{\varepsilon}$.

As discussed after Theorem 6, in order to ensure convergence as $N \to +\infty$ and $\varepsilon \to 0$, the best choice for ε in function of N is $\varepsilon_N \sim \left(\frac{C}{\ln\ln N}\right)^{\frac{1}{n+p}}$ as $N \to +\infty$, and in this case the estimate (81) gives

$$\|y^N_{\varepsilon_N} - y\|_{\mathscr{C}^0([0,T],L^2(\Omega,\mathbb{R}^d))} \leqslant \left(\frac{C}{\ln\ln N}\right)^{\frac{1}{n+p}}(1+\mathrm{o}(1)).$$

The rest of the section is devoted to proving Theorem 7. We are going to apply the item **??** of Theorem 6 with $X = L^2(\Omega,\mathbb{R}^d)$, $Z = W^{p+1,\infty}(\Omega,\mathbb{R}^d)$, $C_\infty = 1$ (because $|\Omega| = 1$), $M = 1$.

Recall that the operator A_ε is defined by (65), i.e., $A_\varepsilon y(x) = \int_\Omega \sigma_\varepsilon(x,x')f(x')\,dx'$ for every $f \in \mathscr{C}^\infty(\Omega)$ and every $x \in \Omega$, where, in the present case, σ_ε is defined by (79), and we have

$$\|\sigma_\varepsilon\|_{L^\infty} \leqslant \frac{C_L}{\varepsilon^{n+p}} \qquad \text{and} \qquad \mathrm{Lip}(\sigma_\varepsilon) \leqslant \frac{C_L}{\varepsilon^{n+p+1}}$$

for some constant $C_L > 0$ depending on η and on the Lipschitz constants of the coefficients a_α but not depending on ε.

We have to prove that the assumptions (69) and (70) are satisfied, and we are going to compute C_A and $\chi(\varepsilon)$. To do that, we first express A_ε using an unusual convolution that we introduce next.

Definition and properties of a convolution operator. Given any $k \in [1, +\infty)$ and any $f \in L^k(\Omega)$, let us define and give some properties of the smooth approximation $\eta_\varepsilon \star_\Omega f \in \mathscr{C}^\infty(\Omega)$ of f for any $\varepsilon \in (0, 1]$.

In the case $(\mathbf{O_1})$, i.e., when Ω is a smooth compact Riemannian manifold of dimension n, using a smooth partition of unity over an atlas of Ω we can always write $f = \sum_{i=1}^m f_i$ for some $m \in \mathbb{N}^*$ and for some functions $f_i \in L^k(\Omega)$ whose essential support is sufficiently small, contained in a chart of the atlas. In each chart, we are locally in $\mathbb{R}^n$ and we can then define $\eta_\varepsilon \star f_i$ with the standard convolution for $\varepsilon > 0$ sufficiently small. At the global level, this defines the function $\eta_\varepsilon \star_\Omega f \in \mathscr{C}^\infty(\Omega)$.

In the case $(\mathbf{O_2})$, i.e., when Ω is the compact closure of a bounded open subset of $\mathbb{R}^n$ with a Lipschitz boundary, we have to be careful with the boundary. Given any $f \in L^2(\Omega)$, for any $\varepsilon \in (0, 1]$, we define the function $\eta_\varepsilon \star_\Omega f$ on Ω by

$$\eta_\varepsilon \star_\Omega f(x) = \int_\Omega \eta_\varepsilon(x - x') f(x')\, dx' \qquad \forall x \in \Omega$$

but we stress that this is not a usual convolution and thus the usual properties of the convolution cannot be used directly. This is why, hereafter, we relate this unusual convolution with the usual one, by extending functions on Ω to $\mathbb{R}^n$ by 0 outside of Ω. Given any $f \in L^2(\Omega)$, we denote by $\tilde{f}$ the extension of f to $\mathbb{R}^n$ by 0. For any $\varepsilon \in (0, 1]$, the consider the function $\eta_\varepsilon \star \tilde{f} \in \mathscr{C}_c^\infty(\mathbb{R}^n)$ defined by the usual convolution

$$(\eta_\varepsilon \star \tilde{f})(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x - x') \tilde{f}(x')\, dx' = \int_\Omega \eta_\varepsilon(x - x') f(x')\, dx' \qquad \forall x \in \mathbb{R}^n,$$

whose support satisfies $\mathrm{supp}(\eta_\varepsilon \star \tilde{f}) \subset \Omega + \overline{B}(0, \varepsilon)$. We have

$$\eta_\varepsilon \star_\Omega f = (\eta_\varepsilon \star \tilde{f})_{|\Omega},$$

i.e., $\eta_\varepsilon \star_\Omega f$ is the restriction of $\eta_\varepsilon \star \tilde{f}$ to Ω. Hence $\eta_\varepsilon \star_\Omega f \in \mathscr{C}^\infty(\Omega)$: it is smooth up to the boundary of the compact domain Ω. We also have $\eta_\varepsilon \star_\Omega f = (\tilde{f} \star \eta_\varepsilon)_{|\Omega}$. Finally, as a consequence of the properties of the usual convolution, we have $\eta_\varepsilon \star_\Omega f \to f$ in $L^2(\Omega)$ as $\varepsilon \to 0$.

More generally, for every $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, the respective functions $D^\alpha(\eta_\varepsilon \star_\Omega f)$, $D^\alpha(\eta_\varepsilon) \star_\Omega f$ and $\eta_\varepsilon \star_\Omega D^\alpha f$ (provided that $D^\alpha f \in L^2(\Omega)$ in the latter case) are smooth on Ω and are the restrictions to Ω of the respective smooth functions $D^\alpha(\eta_\varepsilon \star \tilde{f})$, $D^\alpha(\eta_\varepsilon) \star \tilde{f}$ and $\eta_\varepsilon \star D^\alpha \tilde{f}$ on $\mathbb{R}^n$.[3] In particular, the function $A(\eta_\varepsilon \star_\Omega f)$ is the restriction to Ω of $A(\eta_\varepsilon \star \tilde{f})$.

With these definitions, in both cases $(\mathbf{O_1})$ and $(\mathbf{O_2})$, for every $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ we have $D^\alpha(\eta_\varepsilon \star_\Omega f) = (D^\alpha \eta_\varepsilon) \star_\Omega f = \eta_\varepsilon \star_\Omega D^\alpha f$ for every $f \in L^2(\Omega)$ (such that $D^\alpha f \in L^2(\Omega)$ for the last equality) and $D^\alpha(\eta_\varepsilon \star_\Omega f) \to D^\alpha f$ in $L^2(\Omega)$ as $\varepsilon \to 0$ if $D^\alpha f \in L^2(\Omega)$. We also have

$$\sigma_\varepsilon(x, x') = \left(\eta_\varepsilon \star_\Omega \sum_{|\alpha| \leqslant p} a_\alpha(\cdot)(D^\alpha \eta_\varepsilon)(\cdot - x') \right)(x) \qquad \forall x, x' \in \Omega.$$

[3] There is a small subtlety for defining the latter one, removed by noting that $D^\alpha \tilde{f} = \widetilde{D^\alpha f}$ a.e. if $D^\alpha f \in L^2(\Omega)$.

Expressing A_ε with the convolution operator $\eta_\varepsilon \star_\Omega$.

Lemma 3. *Given any $\varepsilon \in (0,1]$ and any $f \in \mathscr{C}^\infty(\Omega)$, we have*[4]

$$A_\varepsilon f = \eta_\varepsilon \star_\Omega A(\eta_\varepsilon \star_\Omega f) = \Big(\eta_\varepsilon \star \big(A(\eta_\varepsilon \star \tilde{f})\mathbb{1}_\Omega\big)\Big)_{|\Omega}.$$

Proof. For $x \in \Omega$ fixed, we have, using that $D^\alpha \eta_\varepsilon \star_\Omega f = D^\alpha(\eta_\varepsilon \star_\Omega f)$,

$$(A_\varepsilon f)(x) = \int_\Omega \sigma_\varepsilon(x,x') f(x')\,dx' = \Big(\eta_\varepsilon \star_\Omega \sum_{|\alpha| \leqslant p} a_\alpha D^\alpha \eta_\varepsilon \star_\Omega f\Big)(x) = \Big(\eta_\varepsilon \star_\Omega A(\eta_\varepsilon \star_\Omega f)\Big)(x),$$

thus giving the lemma. □

Uniform stability property (69). Thanks to Lemma 3, we can now establish (69).

Lemma 4. *For all $f, g \in L^2(\Omega)$, we have $\langle \eta_\varepsilon \star_\Omega g, f\rangle_{L^2(\Omega)} = \langle g, \eta_\varepsilon \star_\Omega f\rangle_{L^2(\Omega)}$.*

Proof. By definition, we have $\eta_\varepsilon \star_\Omega g = (\eta_\varepsilon \star \tilde{g})_{|\Omega}$, but since $\tilde{f} = 0$ on $\mathbb{R}^n \setminus \Omega$, we have $\langle \eta_\varepsilon \star_\Omega g, f\rangle_{L^2(\Omega)} = \langle \eta_\varepsilon \star \tilde{g}, \tilde{f}\rangle_{L^2(\mathbb{R}^n)}$. Now, using the fact that η_ε is symmetric, i.e., that $\eta_\varepsilon(z) = \eta_\varepsilon(-z)$ for any $z \in \mathbb{R}^n$, and that this property ensures that the usual convolution by η_ε is symmetric in $L^2(\mathbb{R}^n)$, we infer that $\langle \eta_\varepsilon \star_\Omega g, f\rangle_{L^2(\Omega)} = \langle \tilde{g}, \eta_\varepsilon \star \tilde{f}\rangle_{L^2(\mathbb{R}^n)}$. But the latter term is equal to $\langle g, \eta_\varepsilon \star_\Omega f\rangle_{L^2(\Omega)}$. The lemma is proved. □

Lemma 5. *Like the operator $A - \beta\,\mathrm{id}$, the operator $A_\varepsilon - \beta\,\mathrm{id}$ is m-dissipative on $L^2(\Omega, \mathbb{R}^d)$, for any $\varepsilon \in (0,1]$. As a consequence, we have*

$$\|e^{tA_\varepsilon}\|_{L(L^2(\Omega,\mathbb{R}^d))} \leqslant e^{\beta t} \quad \text{and} \quad \|e^{tA}\|_{L(L^2(\Omega,\mathbb{R}^d))} \leqslant e^{\beta t} \qquad \forall t \geqslant 0 \qquad \forall \varepsilon \in (0,1].$$

Therefore, (69) is satisfied (with $M = 1$).

Proof. Given any $f \in \mathscr{C}^\infty(\Omega)$, applying Lemma 4 to $g = (A - \beta\,\mathrm{id})(\eta_\varepsilon \star_\Omega f)$, we have

$$\langle (A_\varepsilon - \beta\,\mathrm{id})f, f\rangle_{L^2(\Omega)} = \langle \eta_\varepsilon \star_\Omega (A - \beta\,\mathrm{id})(\eta_\varepsilon \star_\Omega f), f\rangle_{L^2(\Omega)} = \langle (A - \beta\,\mathrm{id})(\eta_\varepsilon \star_\Omega f), \eta_\varepsilon \star_\Omega f\rangle_{L^2(\Omega)} \leqslant 0$$

because $A - \beta\,\mathrm{id}$ is dissipative. Since A_ε is bounded, we have $D(A_\varepsilon) = L^2(\Omega)$, and thus its adjoint A_ε^* is bounded and $D(A_\varepsilon^*) = L^2(\Omega)$. Therefore, obviously, $A_\varepsilon^* - \beta\,\mathrm{id}$ is also dissipative. The conclusion now follows from the Lumer-Phillips theorem (see [20, Chapter II, Corollary 3.17] or [36, Chapter 1, Theorem 4.3]). □

Remark 11. Lemma 5 is the key step where we use the particular form $A_\varepsilon f = \eta_\varepsilon \star_\Omega A(\eta_\varepsilon \star_\Omega f)$, in order to ensure dissipativity. It would not work if we had chosen $A_\varepsilon f = A(\eta_\varepsilon \star_\Omega f)$. Note that, as mentioned in Section 6.2, (69) is always satisfied when choosing the Yosida approximant $A_\varepsilon = \frac{1}{\varepsilon}\left(\frac{1}{\varepsilon}\mathrm{id} - A\right)^{-1} A$. The interest of the above construction is that it is completely explicit.

[4] The function $A(\eta_\varepsilon \star_\Omega f)$ is the restriction to Ω of the function $A(\eta_\varepsilon \star \tilde{f})$ and, denoting by $g_\varepsilon = A(\eta_\varepsilon \star \tilde{f})\mathbb{1}_\Omega$ the extension to $\mathbb{R}^n$ by 0 of the function $A(\eta_\varepsilon \star_\Omega f)$, the function $\eta_\varepsilon \star_\Omega A(\eta_\varepsilon \star_\Omega f)$ is the restriction to Ω of the function $\eta_\varepsilon \star g_\varepsilon$.

A first convergence property of A_ε.

Lemma 6. *Given any $f \in \mathscr{C}^\infty(\Omega)$, we have $A_\varepsilon f \to Af$ in $L^2(\Omega)$ as $\varepsilon \to 0$.*

Proof. By the triangular inequality, using the expression of A_ε given by Lemma 3,

$$\|A_\varepsilon f - Af\|_{L^2(\Omega)} \leqslant \|\eta_\varepsilon \star_\Omega (A(\eta_\varepsilon \star_\Omega f) - Af)\|_{L^2(\Omega)} + \|\eta_\varepsilon \star_\Omega Af - Af\|_{L^2(\Omega)}.$$

The second term at the right-hand side of that inequality converges to 0 as $\varepsilon \to 0$, because $Af \in L^2(\Omega)$. To handle the first term, we use the Young inequality $\|g \star h\|_{L^r(\mathbb{R}^n)} \leqslant B_{p,q}\|g\|_{L^p(\mathbb{R}^n)}\|h\|_{L^q(\mathbb{R}^n)}$ with $g = \eta_\varepsilon$, $h = \tilde{r}_\varepsilon$ where $r_\varepsilon = A(\eta_\varepsilon \star_\Omega f) - Af$, and with $r = 2$, $p = 1$ and $q = 2$, obtaining

$$\|\eta_\varepsilon \star_\Omega r_\varepsilon\|_{L^2(\Omega)} = \|(\eta_\varepsilon \star \tilde{r}_\varepsilon)_{|\Omega}\|_{L^2(\mathbb{R}^n)} \leqslant \|\eta_\varepsilon \star \tilde{r}_\varepsilon\|_{L^2(\mathbb{R}^n)} \leqslant B_{1,2}\|\eta_\varepsilon\|_{L^1(\Omega)}\|\tilde{r}_\varepsilon\|_{L^2(\mathbb{R}^n)} = B_{1,2}\|r_\varepsilon\|_{L^2(\Omega)}$$

because $\|\eta_\varepsilon\|_{L^1(\Omega)} = 1$, and we conclude that $A_\varepsilon f \to Af$ in $L^2(\Omega)$ by noticing that $r_\varepsilon \to 0$ in $L^2(\Omega)$ because

$$A(\eta_\varepsilon \star_\Omega f) = \sum_{|\alpha| \leqslant p} a_\alpha D^\alpha(\eta_\varepsilon \star_\Omega f) = \sum_{|\alpha| \leqslant p} a_\alpha \eta_\varepsilon \star_\Omega D^\alpha f \longrightarrow \sum_{|\alpha| \leqslant p} a_\alpha D^\alpha f = Af$$

in $L^2(\Omega)$ as $\varepsilon \to 0$. □

In terms of Schwartz kernels, the kernel of A_ε is obtained by convoluting to the left and "to the right" (in some sense) the Schwartz kernel of A with η_ε, and that, for every $x \in \Omega$ fixed, the function $x' \mapsto \sigma_\varepsilon(x, x')$ converges in the distributional sense to the distribution $\sum_{|\alpha| \leqslant p}(-1)^{|\alpha|} a_\alpha \delta_x^{(\alpha)}$ as $\varepsilon \to 0$ (where $\delta_x^{(\alpha)}$ is a distributional derivative of the Dirac δ_x at x), which is the Schwartz kernel $[A](x, \cdot)$ of A.

The convergence property stated in Lemma 6 is not enough to get (70). We need to refine the analysis and establish some error estimates. We start by refining our analysis of the unusual convolution operator introduced previously.

Convergence estimates for the convolution operator $\eta_\varepsilon \star_\Omega$. Set $C_a = \max_{|\alpha| \leqslant p} \|a_\alpha\|_{W^{1,\infty}(\Omega)}$.

We introduce the following notation: in the case **($\mathbf{O_2}$)**, for any $\varepsilon \in (0, 1]$ we define the compact subset Ω_ε of $\mathring{\Omega}$ by

$$\Omega_\varepsilon = \{x \in \Omega \ \mid \ d_\Omega(x, \partial\Omega) \geqslant \varepsilon\} = \mathbb{R}^n \setminus (\mathbb{R}^n \setminus \Omega + B(0, \varepsilon)).$$

There exists a constant $C_{\partial\Omega} > 0$ such that

$$|\Omega \setminus \Omega_\varepsilon| \leqslant C_{\partial\Omega}\, \varepsilon^n \qquad \forall \varepsilon \in (0, 1].$$

In the case **($\mathbf{O_1}$)** we simply set $\Omega_\varepsilon = \Omega$ and $C_{\partial\Omega} = 0$.

We will also need to use extension operators in the case **($\mathbf{O_2}$)**: according to [42, Chap. VI, Sec. 3, Theorem 5] (see also [43, Chap. 12]), there exist $C_E > 0$ and a linear continuous operator E mapping functions on Ω to functions on $\mathbb{R}^n$, such that the restriction of Ef to Ω coincides with f and $\|Ef\|_{W^{j,k}(\mathbb{R}^n)} \leqslant C_E\|f\|_{W^{j,k}(\Omega)}$ for every $f \in W^{j,k}(\Omega)$ and for every $j \in \mathbb{N}$ and every $k \in [1, +\infty]$ (Stein extension). In the case **($\mathbf{O_1}$)**, accordingly, we set $C_E = 1$.

Lemma 7. *Given any $\varepsilon \in (0, 1]$, we have*

$$\|\eta_\varepsilon \star_\Omega f - f\|_{L^\infty(\Omega)} \leqslant 2\|f\|_{L^\infty(\Omega)} \qquad \forall f \in L^\infty(\Omega),$$

$$|\eta_\varepsilon \star_\Omega f(x) - f(x)| \leqslant C_E C_\eta \varepsilon \|f\|_{W^{1,\infty}(\Omega)} \qquad \forall x \in \Omega_\varepsilon \qquad \forall f \in W^{1,\infty}(\Omega).$$

Proof. The first inequality is obviously obtained by using that $\|\eta_\varepsilon \star \tilde{f}\|_{L^\infty(\mathbb{R}^n)} \leqslant \|\tilde{f}\|_{L^\infty(\mathbb{R}^n)}$ because $\|\eta_\varepsilon\|_{L^1(\mathbb{R}^n)} = 1$.

Let $x \in \Omega_\varepsilon$ be arbitrary. Since $\mathrm{supp}(\eta_\varepsilon) \subset \overline{B}(0,\varepsilon)$ and thus $\mathrm{supp}(\eta_\varepsilon(x-\cdot)) \subset \overline{B}(x,\varepsilon) \subset \Omega$, we have $\int_\Omega \eta_\varepsilon(x-x')\,dx' = \int_{\mathbb{R}^n} \eta_\varepsilon(x-x')\,dx' = 1$, hence $f(x) = \int_\Omega \eta_\varepsilon(x-x')f(x)\,dx'$ and

$$|\eta_\varepsilon \star f(x) - f(x)| \leqslant \int_\Omega \eta_\varepsilon(x-x')|f(x') - f(x)|\,dx' \leqslant \int_{\mathbb{R}^n} \eta_\varepsilon(x-x')|Ef(x') - Ef(x)|\,dx'$$

where Ef is the Stein extension of f, defined above (actually the latter inequality is even an equality because $\eta_\varepsilon(x-x') = 0$ for any $x' \in \mathbb{R}^n \setminus \Omega$, since $x \in \Omega_\varepsilon$). It follows from the mean value theorem that

$$|Ef(x') - Ef(x)| \leqslant \|Ef\|_{W^{1,\infty}(\mathbb{R}^n)}\|x-x'\| \leqslant C_E\|f\|_{W^{1,\infty}(\Omega)}\|x-x'\|.$$

Hence

$$|\eta_\varepsilon \star f(x) - f(x)| \leqslant C_E\|f\|_{W^{1,\infty}(\Omega)} \int_{\mathbb{R}^n} \frac{1}{\varepsilon^n}\eta\left(\frac{x-x'}{\varepsilon}\right)\|x-x'\|\,dx' = C_E C_\eta \varepsilon \|f\|_{W^{1,\infty}(\Omega)}$$

by using the change of variable $x' = x - \varepsilon s$.

Note that, in the above argument, we have used a $W^{1,\infty}$ extension of f (and not the extension by 0, which is not in $W^{1,\infty}(\mathbb{R}^n)$) in order to use the mean value theorem, because, in the case $(\mathbf{O_2})$, Ω may not be convex. □

Convergence properties (70) of A_ε. We are now in a position to establish (70).

Lemma 8. *Given any $\varepsilon \in (0,1]$, we have*

$$\|(A_\varepsilon - A)f\|_{L^\infty(\Omega)} \leqslant 4pC_a\|f\|_{W^{p,\infty}(\Omega)} \qquad \forall f \in W^{p,\infty}(\Omega), \tag{82}$$

$$|(A_\varepsilon - A)f(x)| \leqslant 2pC_E C_\eta C_a \varepsilon \|f\|_{W^{p+1,\infty}(\Omega)} \qquad \forall x \in \Omega_\varepsilon \qquad \forall f \in W^{p+1,\infty}(\Omega). \tag{83}$$

As a consequence,

$$\|(A_\varepsilon - A)f\|_{L^2(\Omega)} \leqslant 2pC_a\sqrt{C_E^2 C_\eta^2 \varepsilon^2 + 4C_{\partial\Omega}\varepsilon^n}\|f\|_{W^{p+1,\infty}(\Omega)} \qquad \forall f \in W^{p+1,\infty}(\Omega), \tag{84}$$

and therefore (70) is satisfied with $C_A = 2pC_a$ and $\chi(\varepsilon) = \sqrt{C_E^2 C_\eta^2 \varepsilon^2 + 4C_{\partial\Omega}\varepsilon^n}$.

Note that, in the case $(\mathbf{O_2})$, we have $C_{\partial\Omega} = 0$ and then the above estimate is in ε. Actually, in both cases $(\mathbf{O_1})$ and $(\mathbf{O_2})$ the estimate is in ε as $\varepsilon \to 0$ except in the case $(\mathbf{O_1})$ when moreover $n = 1$, in which case the estimate is in $\sqrt{\varepsilon}$.

Proof. For any $\alpha \in \mathbb{N}^n$ such that $|\alpha| \leqslant p$, noting that $D^\alpha(\eta_\varepsilon \star_\Omega f - f) = \eta_\varepsilon \star_\Omega D^\alpha f - D^\alpha f$, we infer from Lemma 7 applied to $D^\alpha f$ that

$$\|D^\alpha(\eta_\varepsilon \star_\Omega f - f)\|_{L^\infty(\Omega)} \leqslant 2\|f\|_{W^{p,\infty}(\Omega)}$$

$$|D^\alpha(\eta_\varepsilon \star_\Omega f - f)(x)| = |\eta_\varepsilon \star_\Omega (D^\alpha f)(x) - D^\alpha f(x)| \leqslant C_E C_\eta \varepsilon \|f\|_{W^{1+|\alpha|,\infty}(\Omega)} \qquad \forall x \in \Omega_\varepsilon$$

and thus, using that $A = \sum_{|\alpha|\leqslant p} a_\alpha D^\alpha$ and that $\|a_\alpha\|_{L^\infty(\Omega)} \leqslant C_a$, we obtain

$$\|A(\eta_\varepsilon \star_\Omega f - f)\|_{L^\infty(\Omega)} \leqslant 2pC_a\|f\|_{W^{p,\infty}(\Omega)}$$

$$|A(\eta_\varepsilon \star_\Omega f - f)(x)| \leqslant pC_E C_\eta C_a \varepsilon \|f\|_{W^{p+1,\infty}(\Omega)} \qquad \forall x \in \Omega_\varepsilon.$$

Besides, we infer from Lemma 7 applied to $A(\eta_\varepsilon \star_\Omega f)$, using that $\|a_\alpha\|_{W^{1,\infty}(\Omega)} \leqslant C_a$, that

$$\begin{aligned}\|\eta_\varepsilon \star_\Omega A(\eta_\varepsilon \star_\Omega f) - A(\eta_\varepsilon \star_\Omega f)\|_{L^\infty(\Omega)} &\leqslant 2\|A(\eta_\varepsilon \star_\Omega f)\|_{L^\infty(\Omega)} \\ &\leqslant 2pC_a\|\eta_\varepsilon \star_\Omega f\|_{W^{p,\infty}(\Omega)} \leqslant 2pC_a\|f\|_{W^{p,\infty}(\Omega)}\end{aligned}$$

and

$$\begin{aligned}|\eta_\varepsilon \star_\Omega A(\eta_\varepsilon \star_\Omega f)(x) - A(\eta_\varepsilon \star_\Omega f)(x)| &\leqslant C_E C_\eta \varepsilon \|A(\eta_\varepsilon \star_\Omega f)\|_{W^{1,\infty}(\Omega)} \\ &\leqslant pC_E C_\eta C_a \varepsilon \|\eta_\varepsilon \star_\Omega f\|_{W^{p+1,\infty}(\Omega)} \leqslant pC_E C_\eta C_a \varepsilon \|f\|_{W^{p+1,\infty}(\Omega)} \qquad \forall x \in \Omega_\varepsilon\end{aligned}$$

where we have used that $\|\eta_\varepsilon \star g\|_{L^\infty(\mathbb{R}^n)} \leqslant \|g\|_{L^\infty(\mathbb{R}^n)}$ for any $g \in L^\infty(\mathbb{R}^n)$ (note that $\|\eta_\varepsilon\|_{L^1(\mathbb{R}^n)} = 1$). Finally, by the triangular inequality, we have

$$|(A_\varepsilon - A)f(x)| \leqslant |\eta_\varepsilon \star_\Omega A(\eta_\varepsilon \star_\Omega f)(x) - A(\eta_\varepsilon \star_\Omega f)(x)| + |A(\eta_\varepsilon \star_\Omega f - f)(x)|$$

and the estimates (82) and (83) follow.

To establish (84), we write

$$\|(A_\varepsilon - A)f\|^2_{L^2(\Omega)} = \int_{\Omega_\varepsilon} |(A_\varepsilon - A)f(x)|^2\, dx + \int_{\Omega\setminus\Omega_\varepsilon} |(A_\varepsilon - A)f(x)|^2\, dx.$$

Using (83), the first term is estimated by

$$\int_{\Omega_\varepsilon} |(A_\varepsilon - A)f(x)|^2\, dx \leqslant 4p^2 C_E^2 C_\eta^2 C_a^2 \varepsilon^2 \|f\|^2_{W^{p+1,\infty}(\Omega)},$$

and using (82), the second term is estimated by

$$\int_{\Omega\setminus\Omega_\varepsilon} |(A_\varepsilon - A)f(x)|^2\, dx \leqslant \|(A_\varepsilon - A)f\|^2_{L^\infty(\Omega)} |\Omega \setminus \Omega_\varepsilon| \leqslant 16p^2 C_a^2 \|f\|^2_{W^{p,\infty}(\Omega)} C_{\partial\Omega}\, \varepsilon^n,$$

and the conclusion follows. □

6.3.2 Extensions

As explained in Section 6.2.2, Theorem 7 can be straightforwardly extended to non-autonomous linear and quasilinear PDEs, under appropriate assumptions.

Non-autonomous linear PDEs. Theorem 7 is first extended to the case where the coefficients a_α depend on t, i.e.,

$$(A(t)y)(x) = \sum_{|\alpha|\leqslant p} a_\alpha(t,x)\,(D^\alpha y)(x) \qquad \forall y \in \mathscr{C}^\infty(\Omega, \mathbb{R}^d) \qquad \forall (t,x) \in \mathbb{R} \times \Omega,$$

by assuming that the norms of the coefficients $a_\alpha(t,\cdot)$ in $W^{1,\infty}(\Omega)$ are uniformly bounded with respect to t on any compact interval, and by assuming that the family of operators $(A(t))_{0\leqslant t\leqslant T}$ satisfies the requirements of [36, Chapter 5, Section 5.3, Theorem 3.1].

Quasilinear PDEs. For the general quasilinear PDE (76), i.e., the case where the coefficients a_α depend on $(t,x,\xi) \in \mathbb{R} \times \Omega \times \mathbb{R}^d$, G_ε is defined by $G_\varepsilon(t,x,x',\xi,\xi') = \sigma_\varepsilon(t,x,x',\xi)\xi'$ with

$$\sigma_\varepsilon(t,x,x',\xi) = \int_\Omega \eta_\varepsilon(x-z) \sum_{|\alpha| \leqslant p} a_\alpha(t,z,\xi)(D^\alpha \eta_\varepsilon)(z-x').$$

Theorem 7 is also be extended to that case, by assuming that the norms of the coefficients $a_\alpha(t,\cdot,\xi)$ in $W^{1,\infty}(\Omega)$ are bounded, uniformly with respect to (t,ξ) on any compact, and by assuming that the family of operators $A(t,\xi)$, for $0 \leqslant t \leqslant T$ and $\xi \in \mathbb{R}^d$, satisfies the requirements of [36, Chapter 6, Section 6.4, Theorem 4.3].

6.4 Particle approximation of PDEs without semigroup property

In this section, we provide another type of particle approximation of a general class of non linear PDEs non necessitating the semigroup property of the preceding section. The strategy and the proofs are different from those of Section 6. We restrict the study to the 1D case and we take $\Omega = [0,1]$. Let us consider the quasilinear PDE

$$\partial_t y(t,x) = \sum_{j=0}^{p} a_j(y(x),x) D_x^l y(t,x), \;\; y(0,x) = y^0(x), \; x \in [0,1]. \tag{85}$$

with a_j Lipschitz continuous, $l = 0,\dots,L$.

In order to avoid boundary conditions issues, we suppose $a_j(\cdot,0) = a_j(\cdot,1) = 0, \;\; l = 1,\dots,L$. Therefore (85) can be alternatively seen as a PDE on $\mathbb{R}$ by taking $a_j(\cdot,x) = 0, \;\; x \notin (0,1), l = 1,\dots,L$.

We will show in this section that (85) is the graph limit, as $N \to \infty, \varepsilon \to 0$, of the system of particles

$$\dot{\xi} = \frac{1}{N}\sum_{j=1}^{N} G_\varepsilon(i,j,\xi_i,\xi_j), \quad G_\varepsilon(x,x',\xi,\xi') = \sum_{j=0}^{p} a_j(\xi,x)\xi'(i\partial_{x'})^l \frac{e^{-\frac{(x-x')^2}{2\varepsilon}}}{(\pi\varepsilon)^{\frac{1}{2}}}. \tag{86}$$

More precisely, let us associate as previously to (86) the system

$$\begin{cases} \dot{x}_i(t) = 0 \\ \dot{\xi}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} G_\varepsilon(t,x_i,x_j,\xi_i(t),\xi_j(t)), \qquad i = 1,\dots,N \end{cases} \tag{87}$$

we have the following result (we take for simplicity $x_i = \frac{i}{N}, \; i = 1,\dots,N$).

Theorem 8. *Let $T > 0$ and y^0 be such that there exists a unique solution $y(t,x) =: y^t(x), \; 0 \leqslant t \leqslant T, \; x \in [0,1]$ to (85), L differentiable with $D_x^L y^t$ Lipschitz for all $0 \leqslant t \leqslant T$.*

Let Φ_t^ε be the flow generated by the system (87) and $\rho(t) = (\Phi_t^\varepsilon)_* \rho_0$ with $\rho_0 = \prod_{i=1}^N \delta_{x_i} \otimes \delta_{y^0(x_i)}$. Finally, set

$$y_{\varepsilon,N}^t = \int \xi(\rho(t))_{N;1}(\cdot, d\xi).$$

Then, for each $0 \leqslant t \leqslant T$, ε small enough and $N \in \mathbb{N}$,

$$W_1(y^t_{\varepsilon,N}, y^t) \leqslant e^{\varepsilon^{-(L+1)d}\bar{L}_{\mu_0}(t)^2 t/2}\left(\varepsilon^2 \bar{B}_{\mu_0}(t) + \frac{\varepsilon^{-(L+1)d}}{N}\bar{K}_{\mu_0}(t)\right)^{\frac{1}{2}} + \frac{e^{\left(\sup_{x,\xi,s\leqslant t}\mathrm{Lip}\left(\sum_{j=0}^p a_j(y^s(x),x)D^l_x y^s(x)\right)\right)t}}{N} \tag{88}$$

where the constants $\bar{L}_{\mu_0}(t)$, $\bar{L}_{\mu_0}(t)$, $\bar{L}_{\mu_0}(t)$ are defined in (112).

Proof. Out of the particle system (86), we define the Vlasov equation

$$\dot{\mu}(x,\xi) = \nabla_\xi \cdot (\int G_\varepsilon(x,x',\xi,\xi')\mu(dx',d\xi')\mu(x,\xi)) \tag{89}$$

and the associated flow ϕ^t_ε, defined for any solution μ^t to (89) by

$$\mu^t = \phi^t_{\varepsilon *}\mu_0.$$

The limit as $\varepsilon \to 0$ of (89) is the equation,

$$\dot{\mu}(x,\xi) = \nabla_\xi \cdot [G(x,\xi,D_x)\int \xi'\mu(x,d\xi')]\mu(x,\xi),\ G(x,\xi,D_x) = \sum_{j=0}^p a_j(\xi,x)D^l_x, \tag{90}$$

posed on monkinetic measures and associated to the flow ϕ^t the same way as for (89).

Thanks to Lemma 23, Theorem 8 is an immediate corollary of the following Vlasov type equivalent result.

Theorem 9. *Let Φ^ε_t be the flow generated by (87), let μ^t the solution to (90) with the monokinetic initial condition $\mu_0(x,\xi) := \delta(\xi - y^0(x))$ where y^0 satisfies the same hypothesis as in Theorem 8 and let again $\rho_0(X,\Xi) = \prod_{i=1}^N \delta(x_i - \frac{i}{N})\delta(\xi_i - y^0(\frac{i}{N}))$.*

Note that under the hypothesis on μ_0, μ^t is unique for all $0 \leqslant t \leqslant T$, with a first moment $\int \xi'\mu^t(x,d\xi')]$ L-differentiable and Lipschitz continuous.

Then, for all $0 \leqslant t \leqslant T$,

$$W_1(((\Phi^\varepsilon_t)_*\rho_0)^s_{N;1}, \mu^t) \leqslant e^{\varepsilon^{-(L+1)d}\bar{L}_{\mu_0}(t)^2 t/2}\left(\varepsilon^2 \bar{B}_{\mu_0}(t) + \frac{\varepsilon^{-(L+1)d}}{N}\bar{K}_{\mu_0}(t)\right)^{\frac{1}{2}} + \frac{e^{t\sup_{x,\xi,s\leqslant t}\mathrm{Lip}\left(\int[G(x,\xi,D_x)\xi'\mu^s(x,d\xi')]\right)}}{N} \tag{91}$$

where the constants $\bar{L}_{\mu_0}(t)$, $\bar{L}_{\mu_0}(t)$, $\bar{L}_{\mu_0}(t)$ are defined in (112).

Indeed, taking $\mu_0(x,\xi) = \delta(\xi - y^0(x))$ we know that $\mu^t(x,\xi) = \delta(\xi - y^t(x))$ solves (90) if and only if y^t solves (85). Therefore (88) is a consequence of (91), and Theorem 8 follows by $\int \xi'\mu^t(\cdot,d\xi')] = y^t$. Theorem 8 is proved. □

Proof of Theorem 9. Let us consider the flow $\phi^t_{\mu_0}$ associated to the linear equation

$$\dot{\nu}(x,\xi) = \nabla_\xi \cdot \left([G(x,\xi,D_x)\int \xi'\mu^t(x,d\xi')]\nu^t(x,\xi)\right).$$

Let as before $\mu_E^{in} = (\rho_0)^s_{N;1}$. By the Lipschitz hypothesis on $\int \xi' \mu^t(x, d\xi')$, i.e. on $\int G(x,\xi,D_x)\xi'\mu^t(x,d\xi']$ we know that

$$W_1(\mu^t, (\phi^t_{\mu_0})_*\mu_E^{in}) = W_1((\phi^t_{\mu_0})_*\mu_0, (\phi^t_{\mu_0})_*\mu_E^{in}) \leqslant e^{t \sup\limits_{x,\xi,s\leqslant t} \mathrm{Lip}\,(\int[G(x,\xi,D_x)\xi'\mu^s(x,d\xi')])t} W_1(\mu_0^E, \mu_0)$$
$$\leqslant \mathrm{Lip}\,(\mu_0)\frac{e^{t \sup\limits_{x,\xi,s\leqslant t} \mathrm{Lip}\,(\int[G(x,\xi,D_x)\xi'\mu^s(x,d\xi')])}}{N}. \quad (92)$$

Therefore, by the triangle inequality and the fact that (see (94) below) $(\phi^t_{\mu_0})_*\mu_E^{in} = (\phi^t_{\mu_0})_*(\rho_0)_{N;1} = \left(((\phi^t_{\mu_0})^{\otimes N})_*\rho_0\right)_{N;1}$, we get

$$W_1(((\Phi^\varepsilon_t)_*\rho_0)_{N;1}, \mu^t) \leqslant W_1\left(((\Phi^\varepsilon_t)_*\rho_0)_{N;1}, (((\phi^t_{\mu_0})^{\otimes N})_*\rho_0)_{N;1}\right) + \frac{e^{t \sup\limits_{x,\xi,s\leqslant t} \mathrm{Lip}\,(\int[G(x,\xi,D_x)\xi'\mu^s(x,d\xi')])}}{N}. \quad (93)$$

It remains to estimate $W_1\left(((\Phi^\varepsilon_t)_*\rho^{in})_{N;1}, (((\phi^t_{\mu_0})^{\otimes N})_*\rho_0)_{N;1}\right)$.

Let us consider the following optimal coupling of ρ_0 with itself

$$\pi^0(X,\Xi;Y,V) = \rho_0(X,\Xi)\delta_{(X,\Xi)}(Y,V).$$

Let us define

$$\pi^t(X,\Xi;Y,V) = ([\Phi^\varepsilon_t]_{(X,\Xi)} \otimes [(\phi^t_{\mu_0})^{\otimes N}]_{(Y,V)})_*\pi^0(X,\Xi;Y,V)$$

where the subscripts $_{(X,\Xi),(Y,V)}$ denote the variables on which the flows act.

Obviously, π^t is a coupling between $(\Phi^\varepsilon_t)_*\rho_0$ and $((\phi^t_{\mu_0})^{\otimes N})_*\rho_0$.

We define

$$D_N(t) = \frac{1}{N}\int((X-Y)^2 + (\Xi - V)^2)\pi^t(dX, d\Xi, dY, dV).$$

Set $Z = (X,\Xi), Z' = (Y,V)$ and, for any $\rho \in \mathcal{P}(\mathbb{R}^{4dN})$,

$$\rho_s(Z,Z') = \frac{1}{N!}\sum_{\sigma\in\Sigma_N}\rho(\sigma(Z),\sigma(Z'))$$

where the action of σ on Z, Z' is defined by $(\sigma(Z))_i = Z_{\sigma(i)}$, $(\sigma(Z'))_i = Z'_{\sigma(i)}$. Obviously, by invariance of the Lebesgue measure by permutations,

$$\int_{\mathbb{R}^{2dN}}\pi^t_s(Z,Z')d^NZ' = ((\Phi^\varepsilon_t)_*\rho_0)_s(Z) = \frac{1}{N!}\sum_{\sigma\in\Sigma_N}(\Phi^\varepsilon_t)_*\rho_0(\sigma(Z)) = ((\Phi^\varepsilon_t)_*\rho_0)_s(Z),$$

and, the same way,

$$\int_{\mathbb{R}^{2dN}}\pi^t_s(Z,Z')d^NZ = (((\phi^t_{\mu_0})^{\otimes N})_*\rho_0)_s$$

Hence, π^t_s couples $((\Phi^\varepsilon_t)_*\rho_0)_s$ and $(((\phi^t_{\mu_0})^{\otimes N})_*\rho_0)_s$. Therefore, defining, for each $n = 1,\dots,N$,

$$(\pi^t_s)_n(z_1,\dots,z_n,z'_1,\dots,z'_n) = \int_{\mathbb{R}^{4d(N-n)}}\pi^t_s(z_1,\dots,z_N,z'_1,\dots,z'_N)dz_{n+1}\dots dz_N dz'_{n+1}\dots dz'_N,$$

we check that $(\pi^t_s)_n$ couples $(((\Phi^\varepsilon_t)_*\rho_0)_s)_{N:n} = ((\Phi^\varepsilon_t)_*\rho_0)^s_{N:n}$ and $((((\phi^t_{\mu_0})^{\otimes N})_*\rho_0)_s)_{N;n}$. Moreover, one easily check that

$$((((\phi^t_{\mu^0})^{\otimes N})_*\rho_0)_s)_{N;n} = (((\phi^t_{\mu^0})^{\otimes N})_*(\rho_0)_s)_{N;n} = ((\phi^t_{\mu^0})^{\otimes n})_*((\rho_0)_s)_{N;n} = ((\phi^t_{\mu^0})^{\otimes n})_*(\rho_0)^s_{N;n}, \quad (94)$$

so that, coming back to D_N we have, by invariance by permutations of the cost function,

$$\begin{aligned} D_N(t) &= \frac{1}{N}\int((X-Y)^2+(\Xi-V)^2)\pi^t(dX,d\Xi,dY,dV). \\ &= \frac{1}{N}\int((X-Y)^2+(\Xi-V)^2)\pi^t_s(dX,d\Xi,dY,dV). \\ &= \int\frac{1}{n}\sum_{i=1}^n((x_i-y_i)^2+(\xi_i-y_i)^2)(\pi^t_s)_n(dx_1\dots dx_n,d\xi_1\dots d\xi_n,dy_1\dots dy_n,dv_1\dots dv_n). \end{aligned}$$

We conclude, since $(\pi^t_s)_n$ couples $((\Phi^\varepsilon_t)_*\rho^{in})^s_{N:n}$ and $((\phi^t_{\mu^0})^{\otimes n})_*(\rho^{in})^s_{N;n}$, that

$$\begin{aligned} W_1\left(((\Phi^\varepsilon_t)_*\rho^{in})_{N;n},((\phi^t_{\mu^0})^{\otimes n})_*(\rho^{in})^s_{N;n}\right)^2 &\leqslant W_2\left(((\Phi^\varepsilon_t)_*\rho^{in})_{N;n},((\phi^t_{\mu^0})^{\otimes n})_*(\rho^{in})^s_{N;n}\right)^2 \\ &\leqslant \int\sum_{i=1}^n((x_i-y_i)^2+(\xi_i-v_i)^2)d(\pi^t_s)_n = nD_N(t) \end{aligned} \tag{95}$$

We have now to estimate $D_N(t)$. Let us first recall that, by definition,

$$\dot\pi^t = \sum_{i=1}^N\nabla_{\xi_i}\left(\frac{1}{N}\sum_{j=1}^N G_\varepsilon(x_j,x_i,\xi_j,\xi_i)\pi^t(X,\Xi,Y,V)\right)+\sum_{i=1}^N\nabla_{v_i}\left(G_{\mu^t}(y_i,v_i)\pi^t(X,\Xi,Y,V)\right)$$

where, see (90), $G_{\mu^t}(y_i,v_i)=\int[G(y_i,v_i,D_{y_i})\xi'\mu(y_i,d\xi')]$.

Therefore, since $\nabla_\Xi(\Xi-V)^2=2(\Xi-V)=-\nabla_V(\Xi-V)^2$, we get, after integration by part,

$$\begin{aligned} &\dot D_N(t) \\ &= \frac{2}{N}\int\sum_{i=1}^N\left((\xi_i-v_i)\cdot(\frac{1}{N}\sum_{j=1}^N G_\varepsilon(x_j,x_i,\xi_j,\xi_i)-G_{\mu^t}(y_i,v_i)\right)\pi^t(dX,d\Xi,dY,dV) \\ &= \frac{2}{N}\int\sum_{i=1}^N\left((\xi_i-v_i)\cdot(\frac{1}{N}\sum_{j=1}^N(G_\varepsilon(x_j,x_i,\xi_j,\xi_i)-G_\varepsilon(y_j,y_i,v_j,v_i))\right. \\ &\qquad\qquad\left.+(G_\varepsilon(y_j,y_i,v_j,v_i)-G_{\mu^t}(y_i,v_i)))\right)\pi^t(dX,d\Xi,dY,dV) \\ &= \frac{2}{N}\int\sum_{i=1}^N\left((\xi_i-v_i)\cdot(\frac{1}{N}\sum_{j=1}^N(G_\varepsilon(x_j,x_i,\xi_j,\xi_i)-G_\varepsilon(y_j,y_i,v_j,v_i))\right)\pi^t(dX,d\Xi,dY,dV) \quad (96) \\ &+ \frac{2}{N}\int\sum_{i=1}^N\left((\xi_i-v_i)\cdot(\frac{1}{N}\sum_{j=1}^N G_\varepsilon(y_j,y_i,v_j,v_i)-G_{\mu^t}(y_i,v_i))\right)\pi^t(dX,d\Xi,dY,dV)r \quad (97) \end{aligned}$$

The (absolute value of the) r.h.s. of (96) can be easily estimated by

$$2(1+\sup_{(X,\Xi,Y,V)\in supp(\pi^t)}\sup_{1\leqslant i,j\leqslant N}\mathrm{Lip}(G)_{(y_i,y_j,v_i,v_j)})^2D_N(t) := 2(1+\varepsilon^{-(L+1)/2}L_{\mu_0}(t))^2D_N(t) \tag{98}$$

Indeed, using $2u\cdot v\leqslant u^2+v^2$ and the Lipschitz property og G, we have, for $(X,\Xi,Y,V)\in supp(\pi^t)$,

$$\begin{aligned} &\sum_{i=1}^N\left|(\xi_i-v_i)\cdot(\frac{1}{N}\sum_{j=1}^N(G_\varepsilon(x_j,x_i,\xi_j,\xi_i)-G_\varepsilon(y_j,y_i,v_j,v_i))\right| \\ \leqslant\ &(1+\sup_{(X,\Xi,Y,V)\in supp(\pi^t)}\sup_{1\leqslant i,j\leqslant N}\mathrm{Lip}(G)_{(x_i,y_j,\xi_i,v_j)})^2(|X-Y)^2+(\Xi-V)^2). \end{aligned}$$

It remains to estimate the second term (97). By the same trick, its absolute value is bounded by

$$\begin{aligned}
& 2D_N(t) + \frac{2}{N}\int \sum_{i=1}^N \left(\frac{1}{N}\sum_{j=1}^N G_\varepsilon(y_j,y_i,v_j,v_i) - G_{\mu^t}(y_i,v_i)\right)^2 \pi_N^t(dX,dV,dY,d\Xi) \\
= \; & 2D_N(t) + \frac{2}{N}\int \sum_{i=1}^N \left(\frac{1}{N}\sum_{j=1}^N G_\varepsilon(y_j,y_i,v_j,v_i) - G_{\mu^t}(y_i,v_i)\right)^2 ((\phi^t)^{\otimes N})_*\rho^{in}(dY,dV)
\end{aligned}$$

What is left now is to estimate, uniformly on $i = 0,\dots,N$,

$$\begin{aligned}
& \int \left|\frac{1}{N}\sum_j G_\varepsilon(y_i,y_j,v_i,v_j) - G(y_i,v_i,D_{y_i})\int v'\mu^t(y_i,dv')\right|^2 ((\phi^t)^{\otimes N})_*\rho^{in}(dY,dV) \\
\leqslant \; & \int \left|\frac{1}{N}\sum_j G_\varepsilon(y_i,y_j,v_i,v_j) - \int G_\varepsilon(y_i,y',v_i,v')\mu^t(dy',dv')\right|^2 ((\phi^t)^{\otimes N})_*\rho^{in}(dY,dV) \qquad (99) \\
+ \; & \int \left|\int G_\varepsilon(y_i,y',v_i,v')\mu^t(dy',dv') - G(y_i,v_i,D_{y_i})\int v'\mu^t(y_i,dv')\right|^2 ((\phi^t)^{\otimes N})_*\rho^{in}(dY,dV) \qquad (100)
\end{aligned}$$

We get first that, by (86)and (90), and the fact that the a_j identically vanish at $x = 0, 1$,

$$\left|\int G_\varepsilon(y_i,y',v_i,v')\mu^t(dy',dv') - G(y_i,v_i,D_{y_i})\int v'\mu^t(y_i,dv')\right| \leqslant \varepsilon \sup_{l=0,\dots,L,x,\xi} |a_j(\xi,x)|\,\mathrm{Lip}\,(D_x^l\mu^t).$$

so that the second term (100) satisfies

$$\begin{aligned}
& \int \left|\int G_\varepsilon(y_i,y',v_i,v')\mu^t(dy',dv') - G_{\mu^t}(y_i,v_i)\right|^2 (\phi^t)_*^{\otimes N}\rho^{in}(dY,dV) \\
& \qquad \leqslant \varepsilon^2(\sup_{l=0,\dots,L,x,\xi} |a_j(\xi,x)|\,\mathrm{Lip}\,(D_x^l\mu^t))^2 = \varepsilon^2 B_{\mu_0}(t). \qquad (101)
\end{aligned}$$

The first term (99) gives rise to

$$\begin{aligned}
& \int \left|\frac{1}{N}\sum_j G_\varepsilon(y_i,y_j,v_i,v_j) - \int G_\varepsilon(y_i,y',v_i,v')\mu^t(dy',dv')\right|^2 ((\phi^t)^{\otimes N})_*\rho^{in}(dY,dV) \\
\leqslant \; & \frac{\varepsilon^{-Ld}}{N}C_{\mu_0}(t) + \frac{\varepsilon^{-(L+1)d}}{N}D_{\mu_0}(t) + \frac{\varepsilon^{-(L+1)/2}}{N}F^2_{\mu_0}(t) + \frac{\varepsilon^{-(L+1)d}}{N^2}(F^1_{\mu_0}(t))^2 \\
\leqslant \; & \frac{\varepsilon^{-(L+1)d}}{N}(C_{\mu_0}(t) + D_{\mu_0}(t) + F^2_{\mu_0}(t) + (F^1_{\mu_0}(t))^2) = \frac{\varepsilon^{-(L+1)d}}{N}K_{\mu_0}(t), \qquad (102)
\end{aligned}$$

where the constants $C_{\mu_0}(t), D_{\mu_0}(t), F^1_{\mu_0}(t), F^2_{\mu_0}(t)$ are given below by (108), (109), (110) and (111) respectively.

Indeed, for each $j = 1,\dots,N$, let us denote

$$\nu(y,v) = \int G_\varepsilon(y_j,y^*,v_j,v^*)\mu^t(y^*,v^*)dy^*dv^* - G_\varepsilon(y_j,y,v_j,v).$$

Note that, since $\int \mu^t(y,v)dydv = 1$,

$$\int \nu(y,v)\mu^t(y,v)dxdv = 0. \tag{103}$$

One has

$$\begin{aligned}
&\int \left| \int G(y_i,y^*,v_j,v^*)\mu^t(y^*,v^*)dy^*dv^* - \frac{1}{N}\sum_{y=1}^{N} G(y_i,y_y,v_j,v_y)\right|^2 ((\phi^t)^{\otimes N})_*\rho^{in}(dY,dV) \\
= &\int \left|\frac{1}{N}\sum_{k=1}^{N}\nu(y_k,v_k)\right|^2 ((\phi^t)^{\otimes N})_*\rho^{in}(dY,dV) \\
= &\frac{1}{N^2}\sum_{k,l=1,\dots N}\int \nu(y_k,v_k)\nu(y_l,v_l)((\phi^t)^{\otimes N})_*\rho^{in}(dY,dV) \\
= &\frac{1}{N^2}\sum_{k\neq l=1,\dots N}\int \nu(y_k,v_k)\nu(y_l,v_l)\left(((\phi^t)^{\otimes N})_*\rho^{in}\right)_{N;2}(dy_l,dy_k,dv_l,dv_k) \\
+ &\frac{1}{N^2}\sum_{k=1,\dots,N}\int \nu(y_k,v_k)^2\left(((\phi^t)^{\otimes N})_*\rho^{in}\right)_{N;1}(dy_k,dv_k) \\
= &\frac{1}{N^2}\sum_{k\neq l=1,\dots N}\int \nu(y_k,v_k)\nu(y_l,v_l)((\phi^t_{\mu_0})^{\otimes 2})_*(\rho^{in})_{N;2}(dy_l,dy_k,dv_l,dv_k) \qquad (104) \\
+ &\frac{1}{N^2}\sum_{k=1,\dots,N}\int \nu(y_k,v_k)^2(\phi^t_{\mu_0})_*(\rho^{in})_{N;1}(dy_k,dv_k). \qquad (105)
\end{aligned}$$

The integral in (105) is

$$\begin{aligned}
\int \nu(y_k,v_k)^2(\phi^t_{\mu_0})_*(\rho^{in})_{N;1}(dy_k,dv_k) &= \int \nu(y_k,v_k)^2(\phi^t_{\mu_0})_*\mu_0^E(dy_k,dv_k) \\
&= \int \nu(y_k,v_k)^2(\phi^t_{\mu_0})_*\mu_0(dy_k,dv_k) \qquad (106) \\
&+ \int \nu(y_k,v_k)^2(\phi^t_{\mu_0})_*(\mu_0^E-\mu_0)(dy_k,dv_k) \qquad (107)
\end{aligned}$$

The first term (106) can be estimated by

$$\begin{aligned}
&\left|\int \nu(y_k,v_k)^2(\phi^t_{\mu_0})_*\mu_0(dy_k,dv_k)\right| = |\int \nu(y_k,v_k)^2\mu^t(dy_k,dv_k)| \leqslant (\sup_{(y,v)\in supp(\mu^t)}|\nu(y,v)|)^2 \\
&\leqslant 4\|G_\varepsilon\|^2_{L^\infty(supp(\mu^t)\times 2)} \leqslant 4(\pi\varepsilon)^{-L}\|\sum_{j=0}^{p} a_j(x,\xi)\xi'(x-x')^l\|^2_{L^\infty(supp(\mu^t)\times 2)} = \varepsilon^{-L}C_{\mu_0}(t) \qquad (108)
\end{aligned}$$

The second term (107) can be estimated thanks to (92):

$$\begin{aligned}
&|\int \nu(y_k,v_k)^2(\phi^t_{\mu_0})_*(\mu_0^E-\mu_0)(dy_k,dv_k)| \\
\leqslant\ &(\pi\varepsilon)^{-(L+1)}\|\sum_{j=0}^{p} a_j(x,\xi)\xi'(x-x')^l\|^2_{L^\infty(supp(\mu^t)\times 2)}\,\mathrm{Lip}\,(\mu_0)\frac{e^{t\sup_{x,\xi,s\leqslant t}\mathrm{Lip}\,(\int[G(x,\xi,D_x)\xi'\mu^s(x,d\xi')])}}{N} \\
=\ &\frac{\varepsilon^{-(L+1)}}{N}D_{\mu_0}(t) \qquad (109)
\end{aligned}$$

Let us now turn back to the integral in (104). We note first that, by Lemma 22, $W_1((\rho^{in})^s_{N;2}, ((\rho^{in})^s_{N;1})^{\otimes 2}) \leqslant \frac{1}{N} E_{\rho^{in}}$ where $E_{\rho^{in}}$ is given by (126). Therefore

$$|\int \nu(y_k,v_k)\nu(y_l,v_l)((\phi^t_{\mu_0})^{\otimes 2})_*(\rho^{in})_{N;2}(dy_l,dy_k,dv_l,dv_k)| \leqslant |\int \nu(y,v)(\phi^t_{\mu_0})_*\mu^E_0(dy,dv)|^2$$
$$+ |\int \nu(y_k,v_k)\nu(y_l,v_l)((\phi^t_{\mu_0})^{\otimes 2})_*((\rho^{in})_{N;2} - (\mu^E_0)^{\otimes 2})(dy_l,dy_k,dv_l,dv_k)|$$

We have, by (103), the fact that $(\phi^t_{\mu_0})_*\mu_0 = \mu^t$ and (92),

$$|\int \nu(y,v)(\phi^t_{\mu_0})_*\mu^E_0(dy,dv)| = |\int \nu(y,v)(\phi^t_{\mu_0})_*(\mu^E_0 - \mu_0)(dy,dv)|$$
$$\leqslant \mathrm{Lip}\,(\nu|_{supp(\mu^t)})\,\mathrm{Lip}\,(\mu_0)\frac{e^{\left(\sup_{x,\xi,s\leqslant t}\mathrm{Lip}\,(\int[G(x,\xi,D_x)\xi'\mu^s(x,d\xi')])\right)t}}{N} = \frac{\varepsilon^{-(L+1)/2}}{N}F^1_{\mu_0}(t) \quad (110)$$

and the same way

$$|\int \nu(y_k,v_k)\nu(y_l,v_l)((\phi^t_{\mu_0})^{\otimes 2})_*((\rho^{in})_{N;2} - (\mu^E_0)^{\otimes 2})(dy_l,dy_k,dv_l,dv_k)|$$
$$\leqslant \mathrm{Lip}\,(\nu^{\otimes 2}|_{supp(\mu^t)^{\times 2}})\frac{e^{\left(\sup_{x,\xi,s\leqslant t}\mathrm{Lip}\,(\int[G(x,\xi,D_x)\xi'\mu^s(x,d\xi')])\right)t}E_{\mu_0}}{N} = \frac{\varepsilon^{-(L+1)/2}}{N}F^2_{\mu_0}(t). \quad (111)$$

Collecting the estimates (108), (109), (110) and (111) gives (102).

Therefore, by (98) and (101) and (102), we get that $D_N(t)$ satisfies

$$\dot{D}_N(t) \leqslant 2(1+\varepsilon^{-(L+1)/2}L_{\mu_0}(t))^2 D_N(t) + (\varepsilon^2 B_{\mu_0}(t) + \frac{\varepsilon^{-(L+1)d}}{N}K_{\mu_0}(t)).$$

By the Gronwall Lemma we get, for ε small enough,

$$D_N(t) \leqslant e^{\varepsilon^{-(L+1)d}\bar{L}_{\mu_0}(t)^2 t}\left(\varepsilon^2\bar{B}_{\mu_0}(t) + \frac{\varepsilon^{-(L+1)d}}{N}\bar{K}_{\mu_0}(t)\right)$$

where (we recall that $L_{\mu_0}(s)$ is defined in (98))

$$\bar{L}_{\mu_0}(t) = \sup_{0\leqslant s\leqslant t} L_{\mu_0}(s),\ \bar{B}_{\mu_0}(t) = \sup_{0\leqslant s\leqslant t} B_{\mu_0}(s) \text{ and } \bar{K}_{\mu_0}(t) = \sup_{0\leqslant s\leqslant t} K_{\mu_0}(s). \quad (112)$$

Specializing (95) to $n=1$ and using (93), we get finally

$$W_1(((\Phi^\varepsilon_t)_*\rho_0)_{N;1},\mu^t) \leqslant e^{\varepsilon^{-(L+1)d}\bar{L}_{\mu_0}(t)^2 t/2}\left(\varepsilon^2\bar{B}_{\mu_0}(t) + \frac{\varepsilon^{-(L+1)d}}{N}\bar{K}_{\mu_0}(t)\right)^{\frac{1}{2}}$$
$$+ \frac{e^{\left(\sup_{x,\xi,s\leqslant t}\mathrm{Lip}\,(\int[G(x,\xi,D_x)\xi'\mu^s(x,d\xi')])\right)t}}{N}.$$

Theorem 9 is proved. □

Acknowledgment. We are indebted to Nicolas Fournier for a useful discussion.

A Appendix

A.1 Some general facts on the Wasserstein distance

Let $p \in \mathbb{N}^*$. Given any $\mu \in \mathcal{P}(\mathbb{R}^p)$, the measure $\mu^s \in \mathcal{P}(\mathbb{R}^p)$, called the symmetrization under permutations of ρ, is defined by

$$\int_{\mathbb{R}^p} f(x)\, d\mu^s(x) = \frac{1}{p!} \sum_{\sigma \in \mathfrak{S}_p} \int_{\mathbb{R}^p} f(\sigma \cdot x)\, d\mu(x) \qquad \forall f \in \mathscr{C}_c^\infty(\mathbb{R}^p),$$

where $\sigma \cdot x = (x_{\sigma(1)}, \dots, x_{\sigma(N)})$ for every $x \in \mathbb{R}^p$ and $\mathfrak{S}_p$ is the group of permutations of p elements.

Lemma 9. *Given any $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^p)$, we have*

$$W_1(\mu_1^s, \mu_2^s) \leqslant W_1(\mu_1, \mu_2).$$

Proof. We use the definition (9) of W_1. Given any $f \in \mathrm{Lip}(\mathbb{R}^p)$ such that $\mathrm{Lip}(f) \leqslant 1$, the mapping $x \mapsto f(\sigma \cdot x)$ is Lipzchitz on $\mathbb{R}^p$, with the same Lipschitz constant $\mathrm{Lip}(f)$, and thus

$$\int_{\mathbb{R}^p} f\, d(\mu_1^s - \mu_2^s) = \frac{1}{p!} \sum_{\sigma \in \mathfrak{S}_p} \int_{\mathbb{R}^p} f(\sigma \cdot x)\, d(\mu_1 - \mu_2)(x) \leqslant \frac{1}{p!} \sum_{\sigma \in \mathfrak{S}_p} W_1(\mu_1, \mu_2) = W_1(\mu_1, \mu_2)$$

and taking the supremum over all f, the result follows. □

Given any $\mu \in \mathcal{P}(\mathbb{R}^p)$ and any $n \in \{1, \dots, p\}$, the nth-order marginal $\mu_{p:n} \in \mathcal{P}(\mathbb{R}^n)$ of μ is the image of μ under the canonical projection $\pi_n : \mathbb{R}^p = \mathbb{R}^n \times \mathbb{R}^{p-n} \to \mathbb{R}^n$.

Lemma 10. *Given any $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^p)$ and any $n \in \{1, \dots, p\}$, we have*

$$W_1((\mu_1)_{p:n}, (\mu_2)_{p:n}) \leqslant W_1(\mu_1, \mu_2).$$

Proof. Given any $g \in \mathrm{Lip}(\mathbb{R}^n)$ such that $\mathrm{Lip}(g) \leqslant 1$, we have $\mathrm{Lip}(g \circ \pi_n) \leqslant 1$ and thus

$$\int_{\mathbb{R}^n} g\, d((\mu_1)_{p:n} - (\mu_2)_{p:n}) = \int_{\mathbb{R}^p} g \circ \pi_n\, d(\mu_1 - \mu_2) \leqslant W_1(\mu_1, \mu_2)$$

and taking the supremum over all g, the first result follows. □

Lemma 11. *Given any $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^p)$, given any $q \in \mathbb{N}^*$ and any $\mu' \in \mathcal{P}(\mathbb{R}^q)$, we have*

$$W_1(\mu_1, \mu_2) = W_1(\mu_1 \otimes \mu', \mu_2 \otimes \mu') = W_1(\mu' \otimes \mu_1, \mu' \otimes \mu_2).$$

Proof. Let us prove the first equality, the second being obviously similar. By Lemma 10, we already have $W_1(\mu_1, \mu_2) \leqslant W_1(\mu_1 \otimes \mu', \mu_2 \otimes \mu')$. Let us prove the converse inequality. The distance $W_1(\mu_1 \otimes \mu', \mu_2 \otimes \mu')$ is the supremum of $\int_{\mathbb{R}^q} \int_{\mathbb{R}^p} f(x, x')\, d(\mu_1 - \mu_2)(x)\, d\mu'(x')$ over all $f \in \mathrm{Lip}(\mathbb{R}^p \times \mathbb{R}^q)$ such that $\mathrm{Lip}(f) \leqslant 1$. But, for every $x' \in \mathbb{R}^q$, we have $\mathrm{Lip}(f(\cdot, x')) \leqslant 1$ and hence $\int_{\mathbb{R}^p} f(x, x')\, d(\mu_1 - \mu_2)(x) \leqslant W_1(\mu_1, \mu_2)$, and the inequality follows by integrating with respect to x', since μ' is a probability measure. □

Lemma 12. *Given any $\mu_1, \mu_2 \in \mathcal{P}_c(\mathbb{R}^p)$, we have*

$$W_1(\mu_1, \mu_2) \leqslant \|\mathrm{supp}(\mu_1)\|_\infty + \|\mathrm{supp}(\mu_2)\|_\infty.$$

Proof. By (7), since $W_1(\mu_1, \mu_2)$ is the infimum of $\int_{\mathbb{R}^{2p}} \|x - y\|\, d\Pi(x, y)$ over all probability measures Π on $\mathbb{R}^{2p}$ coupling μ_1 and μ_2, we have $W_1(\mu_1, \mu_2) \leqslant \int_{\mathbb{R}^p} \|x\|\, d\mu_1(x) + \int_{\mathbb{R}^p} \|y\|\, d\mu_2(y)$, and the result follows. □

Lemma 13. *Given any $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^p)$ and given any $n \in \mathbb{N}^*$, we have*

$$W_1(\mu_1, \mu_2)^n \leqslant W_1(\mu_1^{\otimes n}, \mu_2^{\otimes n}) \leqslant n\, W_1(\mu_1, \mu_2).$$

Proof. We use the definition (9) of W_1. In particular, using the Fubini theorem, we have

$$\prod_{i=1}^n \int_{\mathbb{R}^p} f_i \, d(\mu_1 - \mu_2) = \int_{\mathbb{R}^{np}} f_1 \otimes \cdots \otimes f_n \, d(\mu_1^{\otimes n} - \mu_2^{\otimes n}) \leqslant W_1(\mu_1^{\otimes n}, \mu_2^{\otimes n})$$

for all $f_i \in \mathrm{Lip}(\mathbb{R}^p)$ such that $\mathrm{Lip}(f_i) \leqslant 1$, for $i = 1, \dots, n$, and taking suprema we get that $W_1(\mu_1, \mu_2)^n \leqslant W_1(\mu_1^{\otimes n}, \mu_2^{\otimes n})$, which is the left-hand side inequality of the lemma.

To establish the right-hand side inequality, we now use the fact that, by (7), $W_1(\mu_1^{\otimes n}, \mu_2^{\otimes n})$ is the infimum of $\int_{\mathbb{R}^{2pn}} \|X - Y\| \, d\Pi(X, Y)$ over all probability measures Π on $\mathbb{R}^{2pn}$ coupling $\mu_1^{\otimes n}$ and $\mu_2^{\otimes n}$ (i.e., of marginals $\mu_1^{\otimes n}$ and $\mu_2^{\otimes n}$ on the two respective copies of $\mathbb{R}^{pn}$). Setting $X = (x^1, \dots, x^n)$ and $Y = (y^1, \dots, y^n)$ with $x^i, y^i \in \mathbb{R}^p$ for every $i \in \{1, \dots, n\}$, we have $\|X - Y\| \leqslant \sum_{i=1}^n \|x^i - y^i\|$ (here, we recall that $\| \ \|$ is the Euclidean norm, either in $\mathbb{R}^{pn}$ or in $\mathbb{R}^p$). For every $i \in \{1, \dots, n\}$, we define the projection $\pi^i : \mathbb{R}^{2pn} \to \mathbb{R}^{2p}$ by $\pi^i(X, Y) = (x^i, y^i)$ and we set $\Pi^i = (\pi^i)_* \Pi$. We also define the projections $\pi_1^i : \mathbb{R}^{2pn} \to \mathbb{R}^p$ and $\pi_2^i : \mathbb{R}^{2pn} \to \mathbb{R}^p$ by $\pi_1^i(X, Y) = x^i$ and $\pi_2^i(X, Y) = y^i$. By definition of Π, we have $(\pi_1^1 \otimes \cdots \pi_1^n)_* \Pi = (\mu_1)^{\otimes n}$ and $(\pi_2^1 \otimes \cdots \pi_2^n)_* \Pi = (\mu_2)^{\otimes n}$, and thus, for every $i \in \{1, \dots, n\}$, we have $(\pi_1^i)_* \Pi = \mu_1$ and $(\pi_2^i)_* \Pi = \mu_2$, which implies that the probability measure $(\pi^i)_* \Pi$ on $\mathbb{R}^{2p}$ is coupling μ_1 with μ_2. Now, since the integral

$$\int_{\mathbb{R}^{2pn}} \|x^i - y^i\| \, d\Pi(X, Y) = \int_{\mathbb{R}^{2n}} \|x^i - y^i\| \, d\Pi^i(x^i, y^i)$$

does not depend on i, taking the sum over $i = 1, \dots, n$ and then taking the infimum gives the result. □

Lemma 14. *Let $\mu_1, \mu_2, \beta \in \mathcal{P}(\mathbb{R}^p)$ and let $\varepsilon > 0$ be such that $\mu_1 = (1 + \varepsilon)\mu_2 - \varepsilon\beta$. Then*

$$W_1(\mu_1, \mu_2) = \varepsilon W_1(\mu_2, \beta).$$

Proof. Given any $f \in \mathscr{C}_c^\infty(\mathbb{R}^p)$, we have $\int_{\mathbb{R}^p} f \, d(\mu_1 - \mu_2) = \varepsilon \int_{\mathbb{R}^p} f \, d(\mu_2 - \beta)$, and taking (in two steps) the supremum over all f such that $\mathrm{Lip}(f) \leqslant 1$, the result follows. □

Lemma 15. *For $i = 1, 2$, let $Y^i(t, \cdot)$ be a continuous time-varying locally Lipschitz vector field on $\mathbb{R}^p$, generating a flow $(\Phi^i(t, t_0, \cdot))_{t \in \mathbb{R}}$ (assumed to be well defined for every $t \in \mathbb{R}$) for any $t_0 \in \mathbb{R}$, that is, $\partial_t \Phi^i(t, t_0, x) = Y^i(t, \Phi^i(t, t_0, x))$ and $\Phi^i(t_0, t_0, x) = x$ for all $t, t_0 \in \mathbb{R}$ and $x \in \mathbb{R}^p$.*

Given any $t_0 \in \mathbb{R}$ and any $\mu^1(t_0), \mu^2(t_0) \in \mathcal{P}_c(\mathbb{R}^p)$, setting $\mu^i(t) = \Phi^i(t, t_0, \cdot)_* \mu^i(t_0)$ for $i = 1, 2$, we have

$$W_1(\mu^1(t), \mu^2(t)) \leqslant e^{(t - t_0) L([t_0, t])} W_1(\mu^1(t_0), \mu^2(t_0)) + M([t_0, t]) \frac{e^{(t - t_0) L([t_0, t])} - 1}{L([t_0, t])} \qquad \forall t \geqslant t_0$$

where, setting $S(t) = \mathrm{Conv}(\mathrm{supp}(\mu^1(t)) \cup \mathrm{supp}(\mu^2(t)))$,

$$L([t_0, t]) = \operatorname{ess\,sup}\{\|\partial_x Y^i(s, x)\| \ \mid \ t_0 \leqslant s \leqslant t, \ x \in S(s), \ i = 1, 2\}$$

is the maximal Lipschitz constant of the vector fields $Y^i(s, \cdot)$ restricted to $S(s)$, for $i = 1, 2$ and $0 \leqslant s \leqslant t$, and

$$M([t_0, t]) = \max\{\|Y^1(s, x) - Y^2(s, x)\| \ \mid \ t_0 \leqslant s \leqslant t, \ x \in S(s)\}.$$

Note that if $Y^1 = Y^2$ then $M(\cdot) = 0$. Note also that, when $t_0 = 0$, in this paper we denote $\Phi^i(t,x) = \Phi^i(t,0,x)$, $L(t) = L([0,t])$ and $M(t) = M([0,t])$.

Proof. The proof is a variant of that of [39, Prop. 4]. By definition, we have $\partial_t \partial_x \Phi^i(t,t_0,x) = \partial_x Y^i(t,\Phi^i(t,t_0,x)).\partial_x \Phi^i(t,t_0,x)$ for $i = 1,2$, for almost every $x \in \mathbb{R}^p$. Therefore, on $S(t)$, we have $\mathrm{Lip}(\Phi^i(t,t_0,\cdot)) \leqslant e^{(t-t_0)L([t_0,t])}$.

Taking an arbitrary $\Pi_{t_0} \in \mathcal{P}(\mathbb{R}^{2p})$ coupling $\mu^1(t_0)$ with $\mu^2(t_0)$, the probability measure $\Pi_t = (\Phi^1(t,t_0,\cdot) \otimes \Phi^2(t,t_0,\cdot))_* \Pi_{t_0}$ couples $\mu^1(t)$ with $\mu^2(t)$. Therefore, using the definition (7) of W_1,

$$\begin{aligned} W_1(\mu^1(t),\mu^2(t)) &\leqslant \int_{\mathbb{R}^{2p}} \|x-y\| \, d\Pi_t(x,y) = \int_{\mathbb{R}^{2p}} \|\Phi^1(t,t_0,x) - \Phi^2(t,t_0,y)\| \, d\Pi_{t_0}(x,y) \\ &\leqslant \int_{\mathbb{R}^{2p}} \|\Phi^1(t,t_0,x) - \Phi^1(t,t_0,y)\| \, d\Pi_{t_0}(x,y) + \int_{\mathbb{R}^{2p}} \|\Phi^1(t,t_0,y) - \Phi^2(t,t_0,y)\| \, d\Pi_0(x,y) \end{aligned}$$

The first term at the right-hand side of the inequality is less than $e^{(t-t_0)L([t_0,t])} \int_{\mathbb{R}^{2p}} \|x-y\| \, d\Pi_{t_0}(x,y)$ and thus than $e^{(t-t_0)L([t_0,t])} W_1(\mu^1(t_0),\mu^2(t_0))$ by taking the infimum over Π_{t_0}. For the second term, we observe that, for every $y \in S(t)$,

$$\begin{aligned} \partial_t \|\Phi^1(t,t_0,y) - \Phi^2(t,t_0,y)\| &\leqslant \|Y^1(t,\Phi^1(t,t_0,y)) - Y^1(t,\Phi^2(t,t_0,y))\| \\ &\qquad + \|Y^1(t,\Phi^2(t,t_0,y)) - Y^2(t,\Phi^2(t,t_0,y))\| \\ &\leqslant L([t_0,t]) \|\Phi^1(t,t_0,y) - \Phi^2(t,t_0,y)\| + M([t_0,t]) \end{aligned}$$

and thus, by the Gronwall lemma,

$$\|\Phi^1(t,t_0,y) - \Phi^2(t,t_0,y)\| \leqslant \int_{t_0}^t M([t_0,s]) e^{\int_s^t L([t_0,\tau])\, d\tau} \, ds \leqslant M([t_0,t]) \frac{e^{(t-t_0)L([t_0,t])} - 1}{L([t_0,t])}$$

Therefore, the second term is less than $M([t_0,t]) \frac{e^{(t-t_0)L([t_0,t])}-1}{L([t_0,t])}$, and the lemma follows. □

A.2 Density of empirical measures in the set of probability measures

Let $p \in \mathbb{N}^*$. Given any $N \in \mathbb{N}^*$ and any $Y = (y_1,\ldots,y_N) \in (\mathbb{R}^p)^N$, we define the *empirical measure* $\mu_Y^E \in \mathcal{P}(\mathbb{R}^p)$ by

$$\mu_Y^E = \frac{1}{N} \sum_{i=1}^N \delta_{y_i}.$$

The points y_i are not required to be distinct, so that the empirical measure μ_Y^E can equivalently be defined as a *convex combination with rational coefficients of Dirac masses*. Note that

$$\int_{\mathbb{R}^p} f \, d\mu_Y^E = \frac{1}{N} \sum_{i=1}^N f(y_i) \qquad \forall f \in \mathscr{C}^0(\mathbb{R}^p).$$

By the Riesz-Markov-Kakutani theorem, $\mathcal{P}(\mathbb{R}^p)$ is identified with a subspace of $\mathscr{C}_0(\mathbb{R}^p)'$, the topological dual $\mathscr{C}_0(\mathbb{R}^p)'$ of the Banach space $\mathscr{C}_0(\mathbb{R}^p)$ of continuous functions on $\mathbb{R}^p$ vanishing at infinity endowed with the sup norm. Then, $\mathcal{P}(\mathbb{R}^p)$ inherits of the weak star topology of $\mathscr{C}_0(\mathbb{R}^p)'$, and we say that a sequence $(\mu_k)_{k\in\mathbb{N}^*}$ of $\mathcal{P}(\mathbb{R}^p)$ converges weakly to $\mu \in \mathcal{P}(\mathbb{R}^p)$ if $\int_{\mathbb{R}^p} f \, d\mu_k \to \int_{\mathbb{R}^p} f \, d\mu$ for any $f \in \mathscr{C}_0(\mathbb{R}^p)$. Since we deal with probability measures, this is equivalent to $\int_{\mathbb{R}^p} f \, d\mu_k \to \int_{\mathbb{R}^p} f \, d\mu$ for any $f \in \mathscr{C}_b(\mathbb{R}^p)$ (narrow convergence), where $\mathscr{C}_b(\mathbb{R}^p)$ is the Banach space of bounded functions on $\mathbb{R}^p$.

Lemma 16. *The set* $\{\mu_Y^E \mid N \in \mathbb{N}^*,\ Y \in (\mathbb{R}^p)^N\}$ *is weakly dense in* $\mathcal{P}(\mathbb{R}^p)$*. In other words, any probability measure on* $\mathbb{R}^p$ *is the weak limit of a sequence of empirical measures.*

Proof. This is a well known consequence of the Krein-Milman theorem (see, e.g., [32, Lemma 7]). Let us anyway recall a proof. The set $\mathcal{P}(\mathbb{R}^p)$ is convex and weak star compact, and its extreme points are Dirac masses. The Krein-Milman theorem implies that any $\mu \in \mathcal{P}(\mathbb{R}^p)$ is the limit of a finite convex combination $\sum_i \lambda_i \delta_{y_i}$ of Dirac masses. By density of rationals, without loss of generality we can moreover assume that $\lambda_i \in \mathbb{Q}$. The statement follows. □

Recall that the Wasserstein distance W_1 metrizes the weak convergence in $\mathcal{P}_1(\mathbb{R}^p)$ (which also entails the convergence of first moments). We have then the following variant of the above lemma (see [45, Theorem 6.18]).

Lemma 17. *The set* $\{\mu_Y^E \mid N \in \mathbb{N}^*,\ Y \in (\mathbb{R}^p)^N\}$ *is dense in* $\mathcal{P}_1(\mathbb{R}^p)$ *for the Wasserstein distance* W_1*. In other words, any probability measure on* $\mathbb{R}^p$ *having a finite first moment is the limit of a sequence of empirical measures for the Wasserstein distance* W_1*.*

Proof. It suffices to consider $R > 0$ sufficiently large such that $\int_{\mathbb{R}^p \setminus B(0,R)} \|x\|\, d\mu(x) < \varepsilon$, for $\varepsilon > 0$ small enough, so that the argument can be performed in the compact set $\overline{B}(0,R)$, and the statement readily follows (see also [41, Chap. 5]). □

There exist a number of results in the literature quantifying the convergence of empirical measures μ_Y^E towards $\mu \in \mathcal{P}(\mathbb{R}^p)$ and providing rates of convergence, most in a probabilistic context, like [23] where Y consists of N random variables having the same distribution as μ. In the result hereafter, Y is deterministic and the rate of convergence is the one obtained by Riemann integration.

Lemma 18. *Let* $\mu \in \mathcal{P}_c(\mathbb{R}^p)$ *and let* $N \in \mathbb{N}^*$*. We assume that there exists a tagged partition of* $\mathrm{supp}(\mu)$ *associated with* μ *(see (12)), i.e., a partition of* $\mathrm{supp}(\mu) = \cup_{i=1}^N F_i$ *such that all subsets* F_i *are* μ*-measurable, pairwise disjoint,* $\mu(F_i) = \frac{1}{N}$*, and there exists* $C > 0$ *such that* $\mathrm{diam}(F_i) \leqslant \frac{C}{N}$*. Then, given any* $Y = (y_1, \dots, y_N) \in (\mathbb{R}^p)^N$ *such that* $y_i \in F_i$ *for every* $i \in \{1, \dots, N\}$*, we have*

$$W_1(\mu_Y^E, \mu) \leqslant \frac{C}{N}.$$

The assumption made on μ implies that the mass of μ is quite well uniformly distributed; for instance it is satisfied if μ is absolutely continuous with respect to the Lebesgue measure, with a density that is bounded. This result is quite obvious and has nothing to see with much deeper and general results like those of [23].

Proof. For every $i \in \{1, \dots, N\}$, we have $\int_{F_i} f(y_i)\, d\mu(y) = f(y_i)\mu(F_i) = \frac{1}{N} f(y_i)$ and thus, for every $f \in \mathrm{Lip}(\mathbb{R}^p)$ such that $\mathrm{Lip}(f) \leqslant 1$,

$$\begin{aligned}\left|\int_{\mathbb{R}^p} f\, d(\mu - \mu_Y^E)\right| &= \left|\sum_{i=1}^N \int_{F_i} f(y)\, d\mu(y) - \frac{1}{N}\sum_{i=1}^N f(y_i)\right| = \left|\sum_{i=1}^N \int_{F_i} (f(y) - f(y_i))\, d\mu(y)\right| \\ &\leqslant \sum_{i=1}^N \int_{F_i} |f(y) - f(y_i)|\, d\mu(y) \leqslant \sum_{i=1}^N \int_{F_i} \|y - y_i\|\, d\mu(y) \leqslant \sum_{i=1}^N \mu(F_i)\mathrm{diam}(F_i) \leqslant \frac{C}{N}\end{aligned}$$

and the conclusion follows by taking the supremum over all f. □

A.3 Convergence of empirical and semi-empirical measures

Let (Ω, d_Ω) be a complete metric space and let $\nu \in \mathcal{P}(\Omega)$. We assume that, for any $N \in \mathbb{N}^*$, there exists a tagged partition $(\mathcal{A}, X)$ of Ω associated with ν (satisfying (12), see Section 1.2), with $\mathcal{A} = (\Omega_1, \dots, \Omega_N)$ and $X = (x_1, \dots, x_N)$. We do not put any superscript N to $(\mathcal{A}, X)$ nor to Ω_i, x_i to keep a better readability. We define the *empirical measure* $\nu_X^E \in \mathcal{P}(\Omega)$ by

$$\nu_X^E = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}.$$

A.3.1 Convergence of empirical measures on Ω

Lemma 19.

- *Let f be a bounded and ν-almost everywhere continuous (i.e., ν-Riemann integrable) function on Ω, of compact support. Then*

$$\int_\Omega f \, d(\nu - \nu_X^E) = \int_\Omega f \, d\nu - \frac{1}{N} \sum_{i=1}^N f(x_i) = \mathrm{o}(1) \tag{113}$$

as $N \to +\infty$. As a consequence, ν_X^E converges weakly to ν as $N \to +\infty$; equivalently, $W_1\left(\nu_X^E, \nu\right) = \mathrm{o}(1)$ as $N \to +\infty$ if moreover Ω is compact.

- *Given any $\alpha \in (0, 1]$ and any $N \in \mathbb{N}^*$, we have*

$$\left| \int_\Omega f \, d\nu - \frac{1}{N} \sum_{i=1}^N f(x_i) \right| \leqslant \frac{C_\Omega^\alpha}{N^\alpha} \operatorname{Hol}_\alpha(f) \tag{114}$$

for every $f \in \mathscr{C}_c^{0,\alpha}(\Omega)$. As a consequence of (114) for $\alpha = 1$, we have

$$W_1\left(\nu_X^E, \nu\right) \leqslant \frac{C_\Omega}{N}. \tag{115}$$

Proof. In the first item, (113) follows from the theorem of convergence of Riemann sums, as already recalled in (14). Interpreted in terms of the empirical measure ν_X^E, this means that ν_X^E converges weakly to ν as $N \to +\infty$. In accordance with the Portmanteau theorem (see, e.g., [8, Chapter 1, Section 2, Theorem 2.1]), since W_1 metrizes the weak convergence, we have $W_1\left(\nu_X^E, \nu\right) = \mathrm{o}(1)$ as $N \to +\infty$ if moreover Ω is compact.

Writing $\int_\Omega f \, d\nu = \sum_{i=1}^N \int_{\Omega_i} f \, d\nu$ and using that $\nu(\Omega_i) = \frac{1}{N}$ and that $\operatorname{diam}(\Omega_i) \leqslant \frac{C_\Omega}{N}$ (see (12)), we have

$$\begin{aligned} \left| \int_\Omega f \, d\nu - \frac{1}{N} \sum_{i=1}^N f(x_i) \right| &\leqslant \sum_{i=1}^N \int_{\Omega_i} |f(x) - f(x_i)| \, d\nu(x) \leqslant \operatorname{Hol}_\alpha(f) \sum_{i=1}^N \int_{\Omega_i} d_\Omega(x, x_i)^\alpha \, d\nu(x) \\ &\leqslant \operatorname{Hol}_\alpha(f) \sum_{i=1}^N \nu(\Omega_i) \operatorname{diam}(\Omega_i)^\alpha \leqslant \frac{C_\Omega^\alpha}{N^\alpha} \operatorname{Hol}_\alpha(f) \end{aligned}$$

which gives (114). Taking $\alpha = 1$, (115) follows by the definition (9) of W_1. □

A.3.2 Convergence of semi-empirical measures

Let $\mu \in \mathcal{P}(\Omega \times \mathbb{R}^d)$, disintegrated as $\mu = \int_\Omega \mu_x \, d\nu(x)$ with respect to its marginal $\nu = \pi_* \mu$ on Ω and $(\mu_x)_{x \in \Omega}$ is a family of probability measures on $\mathbb{R}^d$. We define the *semi-empirical measure*

$\mu_X^{SE} \in \mathcal{P}(\Omega \times \mathbb{R}^d)$ by

$$\mu_X^{SE} = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \otimes \mu_{x_i} = \int_\Omega \mu_x \, d\nu_X^E(x).$$

Its marginal on Ω is the empirical measure ν_X^E. In other words, the disintegration of μ_X^{SE} with respect to ν_X^E is the family of probability measures given by μ_{x_i} when $x = x_i$ for some $i \in \{1, \dots, N\}$ and 0 otherwise.

Lemma 20.

- *We assume that $x \mapsto \mu_x$ is ν-almost everywhere continuous for the Wasserstein distance W_1. Let f be a bounded and μ-almost everywhere continuous (i.e., μ-Riemann integrable) function on $\Omega \times \mathbb{R}^d$, of compact support, Lipschitz with respect to $\xi \in \mathbb{R}^d$ with a Lipschitz constant that is uniform with respect to $x \in \Omega$. Then*

$$\int_{\Omega \times \mathbb{R}^d} f \, d(\mu - \mu_X^{SE}) = \mathrm{o}(1) \tag{116}$$

as $N \to +\infty$. As a consequence, μ_X^{SE} converges weakly to μ. If moreover $\mu \in \mathcal{P}_1(\Omega \times \mathbb{R}^d)$ then $W_1(\mu_X^{SE}, \mu) = \mathrm{o}(1)$ as $N \to +\infty$.

- *We assume that $x \mapsto \mu_x$ is Lipschitz for the Wasserstein distance W_1, i.e., that there exists $L > 0$ such that $W_1(\mu_x, \mu_y) \leqslant L \, d_\Omega(x, y)$ for ν-almost all $x, y \in \Omega$. Then, given any $N \in \mathbb{N}^*$,*

$$\left| \int_{\Omega \times \mathbb{R}^d} f \, d(\mu - \mu_X^{SE}) \right| \leqslant \frac{C_\Omega(L+1)}{N} \operatorname{Lip}(f) \tag{117}$$

for every $f \in \mathscr{C}_0^0(\Omega \times \mathbb{R}^d) \cap \operatorname{Lip}(\Omega \times \mathbb{R}^d)$. As a consequence,

$$W_1\left(\mu_X^{SE}, \mu\right) \leqslant \frac{C_\Omega(L+1)}{N}. \tag{118}$$

Proof. Let $f : \Omega \times \mathbb{R}^d \to \mathbb{R}$ be a bounded and μ-almost everywhere continuous function, of compact support, Lipschitz with respect to $\xi \in \mathbb{R}^d$. The function F defined by $F(x) = \int_{\mathbb{R}^d} f(x, \xi) \, d\mu_x(\xi)$ is bounded on Ω, and

$$\begin{aligned} |F(x) - F(y)| &\leqslant \int_{\mathbb{R}^d} |f(x,\xi) - f(y,\xi)| \, d\mu_x(\xi) + \left| \int_{\mathbb{R}^d} f(y,\xi) \, d(\mu_x - \mu_y)(\xi) \right| \\ &\leqslant \int_{\mathbb{R}^d} |f(x,\xi) - f(y,\xi)| \, d\mu_x(\xi) + W_1(\mu_x, \mu_y) \operatorname{Lip}(f(y, \cdot)) \end{aligned} \tag{119}$$

for all $x, y \in \Omega$. Now:

- First, if moreover $y \mapsto \operatorname{Lip}(f(y, \cdot))$ is bounded on Ω and if $x \mapsto \mu_x$ is ν-almost everywhere continuous for the Wasserstein distance W_1, then we infer from (119) that F is ν-almost everywhere continuous. Therefore

$$\int_{\Omega \times \mathbb{R}^d} f \, d(\mu - \mu_X^{SE}) = \int_\Omega F \, d(\nu - \nu_X^E) = \int_\Omega F \, d\nu - \frac{1}{N} \sum_{i=1}^{N} F(x_i) = \mathrm{o}(1)$$

as $N \to +\infty$ by convergence of Riemann sums (f and thus F being fixed), which gives (116). The statement on W_1 is because W_1 metrizes the weak convergence in $\mathcal{P}_1(\Omega \times \mathbb{R}^d)$.

- Second, if $f \in \mathrm{Lip}(\Omega \times \mathbb{R}^d)$ and if $x \mapsto \mu_x$ is L-Lipschitz for the Wasserstein distance W_1 then we infer from (119) that

$$|F(x) - F(y)| \leqslant \mathrm{Lip}(f)\, d_\Omega(x,y) + W_1(\mu_x, \mu_y)\, \mathrm{Lip}(f) \leqslant \mathrm{Lip}(f)(1+L)\, d_\Omega(x,y)$$

and thus, using Lemma 19, that $\int_\Omega F\, d(\nu - \nu_X^E) \leqslant \frac{C_\Omega}{N} \mathrm{Lip}(F)$, whence (117) and (118).

□

Remark 12. In the first item of Lemma 20, the boundedness assumption on f can be slightly weakened to: $x \mapsto f(x,0)$ bounded and $\mu \in \mathcal{P}_1(\Omega \times \mathbb{R}^d)$. Indeed, writing $|f(x,\xi)| \leqslant |f(x,0)| + \mathrm{Lip}(f(x,\cdot))|\xi|$, we infer that F is bounded. The rest of the proof is the same.

A.4 Symmetrization of measures and marginals

Let $N \in \mathbb{N}^*$ and let $\rho \in \mathcal{P}(\Omega^N \times \mathbb{R}^{dN})$. The symmetrization ρ^s of ρ is defined by (30), i.e., with more compact notations, by

$$\rho^s = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sigma_* \rho \tag{120}$$

where the measure $\sigma_*\rho$ is defined by $\langle \sigma_*\rho, f\rangle = \langle \rho, \sigma^* f\rangle$ and $(\sigma^* f)(X,\Xi) = f(X_\sigma, \Xi_\sigma)$, with $X_\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(N)})$ and $\Xi_\sigma = (\xi_{\sigma(1)}, \dots, \xi_{\sigma(N)})$. Here, $\langle\ ,\ \rangle$ is the duality bracket.

First marginal of the symmetrization. For every $i \in \{1, \dots, N\}$, we denote by p^i the projection of $\Omega^N \times \mathbb{R}^{dN}$ onto the product $\Omega \times \mathbb{R}^d$ of the ith copy of Ω with the ith copy of $\mathbb{R}^d$, i.e., in coordinates, $p^i(X,\Xi) = (x_i, \xi_i)$.

Let us compute the first marginal $\rho^s_{N:1} = p^1_* \rho^s$ of the symmetrization ρ^s of ρ.

Lemma 21. *We have*

$$\rho^s_{N:1} = \frac{1}{N} \sum_{i=1}^N p^i_* \rho$$

where $p^i_*\rho$ is the image of ρ under the projection p^i. In other words, $\rho^s_{N:1}$ is the average of the marginals of ρ on the copies of $\Omega \times \mathbb{R}^d$.

Proof. Given any $f \in \mathscr{C}^\infty_c(\Omega \times \mathbb{R}^d)$, we have

$$\begin{aligned}\langle \rho^s_{N:1}, f\rangle &= \langle p^1_*\rho^s, f\rangle = \langle \rho^s, (p^1)^* f\rangle = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \langle \sigma_*\rho, (p^1)^* f\rangle = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \langle \rho, \sigma^*(p^1)^* f\rangle \\ &= \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \int_{\Omega^N \times \mathbb{R}^{dN}} f \circ p^1(X_\sigma, \Xi_\sigma)\, d\rho(X,\Xi) = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \int_{\Omega^N \times \mathbb{R}^{dN}} f(x_{\sigma(1)}, \xi_{\sigma(1)})\, d\rho(X,\Xi)\end{aligned}$$

When designing a permutation $\sigma \in \mathfrak{S}_N$, we have N choices for $\sigma(1)$, among $\{1, \dots, N\}$, and the rest is a permutation of $N-1$ elements. Since $\mathrm{card}(\mathfrak{S}_{N-1}) = (N-1)!$, we get that

$$\begin{aligned}\langle \rho^s_{N:1}, f\rangle = \frac{1}{N} \sum_{i=1}^N \int_{\Omega^N \times \mathbb{R}^{dN}} f(x_i, \xi_i)\, d\rho(X,\Xi) &= \frac{1}{N} \sum_{i=1}^N \int_{\Omega^N \times \mathbb{R}^{dN}} f \circ p^i(X,\Xi)\, d\rho(X,\Xi) \\ &= \frac{1}{N} \sum_{i=1}^N \langle \rho, f \circ p^i\rangle = \frac{1}{N} \sum_{i=1}^N \langle p^i_*\rho, f\rangle\end{aligned}$$

whence the result. □

A technical lemma. Given any $n \in \{1, \dots, N\}$, the nth-order marginal of ρ is, by definition, the image of ρ under the projection of $\Omega^N \times \mathbb{R}^{dN}$ on the product $\Omega^n \times \mathbb{R}^{dn}$ of the n first copies of Ω with the n first copies of $\mathbb{R}^d$.

Lemma 22. *Let $X = (x_1, \dots, x_N) \in \Omega^N$ and let $\gamma_1, \dots, \gamma_N \in \mathcal{P}(\mathbb{R}^d)$. Setting $\Gamma = \gamma_1 \otimes \cdots \otimes \gamma_N$, we define $\rho \in \mathcal{P}(\Omega^N \times \mathbb{R}^{dN})$ by*

$$\rho = \delta_X \otimes \Gamma = \delta_{x_1} \otimes \cdots \otimes \delta_{x_N} \otimes \gamma_1 \otimes \cdots \otimes \gamma_N.$$

The symmetrization of ρ (defined by (30) or by (120)) is given by

$$\rho^s = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \delta_{x_{\sigma(1)}} \otimes \cdots \otimes \delta_{x_{\sigma(N)}} \otimes \gamma_{\sigma(1)} \otimes \cdots \otimes \gamma_{\sigma(N)}. \tag{121}$$

The first marginal $\rho^s_{N:1} \in \mathcal{P}(\Omega \times \mathbb{R}^d)$ of ρ^s is

$$\rho^s_{N:1} = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \otimes \gamma_i \tag{122}$$

and, for every $n \in \{2, \dots, N\}$, its nth-order marginal $\rho^s_{N:n} \in \mathcal{P}(\Omega^n \times \mathbb{R}^{dn})$ is

$$\rho^s_{N:n} = (1 + \varepsilon_n) \left(\rho^s_{N:1}\right)^{\otimes n} - \varepsilon_n \beta_n \tag{123}$$

where

$$\varepsilon_n = \frac{N^n (N-n)!}{N!} - 1 \in \left[0, e^{\frac{n^2}{2N}} - 1\right] \tag{124}$$

and

$$\beta_n = \frac{1}{\varepsilon_n} \frac{(N-n)!}{N!} \sum \delta_{x_{i_1}} \otimes \cdots \otimes \delta_{x_{i_n}} \otimes \gamma_{i_1} \otimes \cdots \otimes \gamma_{i_n} \in \mathcal{P}(\Omega^n \times \mathbb{R}^{dn}) \tag{125}$$

where the sum in (125) is taken over all n-tuples $(i_1, \dots, i_n) \in \{1, \dots, N\}^n$ for which at least two elements are equal. We have

$$W_1\left(\rho^s_{N:n}, (\rho^s_{N:1})^{\otimes n}\right) \leqslant \left(e^{\frac{n^2}{2N}} - 1\right) W_1\left((\rho^s_{N:1})^{\otimes n}, \beta_n\right) \qquad \forall n \in \{1, \dots, N\}. \tag{126}$$

Proof. The formula (121) straightforwardly follows from (120), and the formula (122) follows from Lemma 22 because $p^i_* \rho = \delta_{x_i} \otimes \gamma_i$.

Let us now compute the nth-order marginal $\rho^s_{N:n}$ of ρ^s, for every $n \in \{2, \dots, N\}$. Let I^N_n be the set of all n-tuples $(i_1, \dots, i_n)$ consisting of distinct integers chosen in $\{1, \dots, N\}$. We have $\text{card}(I^N_n) = \frac{N!}{(N-n)!}$. Denoting by $\mathfrak{S}_N^{i_1, \dots, i_n}$ the set of all $\sigma \in \mathfrak{S}_N$ such that $(\sigma(1), \dots, \sigma(n)) = (i_1, \dots, i_n)$, we have $\text{card}(\mathfrak{S}_N^{i_1, \dots, i_n}) = (N-n)!$. Now, since

$$\sum_{\sigma \in \mathfrak{S}_N} \delta_{x_{\sigma(1)}} \otimes \cdots \otimes \delta_{x_{\sigma(N)}} = \sum_{(i_1, \dots, i_N) \in I^N_n} \delta_{x_{i_1}} \otimes \cdots \otimes \delta_{x_{i_n}} \otimes \sum_{\sigma \in \mathfrak{S}_N^{i_1, \dots, i_n}} \delta_{x_{\sigma(n+1)}} \otimes \cdots \otimes \delta_{x_{\sigma(N)}}$$

we infer that

$$\rho^s_{N:n} = \frac{(N-n)!}{N!} \sum_{(i_1, \dots, i_n) \in I^N_n} \delta_{x_{i_1}} \otimes \cdots \otimes \delta_{x_{i_n}} \otimes \gamma_{i_1} \otimes \cdots \otimes \gamma_{i_n}. \tag{127}$$

Now, writing $I_n^N = \{1,\dots,N\}^n \setminus \left(\{1,\dots,N\}^n \setminus I_n^N\right)$, we write the sum in (127) as a sum over $\{1,\dots,N\}^n$ minus a sum over $\{1,\dots,N\}^n \setminus I_n^N$ (where at least two of the indices are equal). For the first sum, we have

$$\sum_{(i_1,\dots,i_n)\in\{1,\dots,N\}^n} \delta_{x_{i_1}} \otimes \cdots \otimes \delta_{x_{i_n}} \otimes \gamma_{i_1} \otimes \cdots \otimes \gamma_{i_n} = \left(\sum_{i=1}^N \delta_{x_i} \otimes \gamma_i\right)^{\otimes n} = N^n \left(\rho^s_{N:1}\right)^{\otimes n}. \quad (128)$$

We infer from (127) and (128) that

$$\rho^s_{N:n} = \frac{N^n (N-n)!}{N!} \left(\rho^s_{N:1}\right)^{\otimes n} - \frac{(N-n)!}{N!}\beta$$

where

$$\beta = \sum_{(i_1,\dots,i_n)\in\{1,\dots,N\}^n\setminus I_n^N} \delta_{x_{i_1}} \otimes \cdots \otimes \delta_{x_{i_n}} \otimes \gamma_{i_1} \otimes \cdots \otimes \gamma_{i_n}$$

is a nonnegative Radon measure of total mass $|\beta| = \mathrm{card}(\{1,\dots,N\}^n \setminus I_n^N) = N^n - \frac{(N-n)!}{N!}$. Besides, we have

$$1 \leqslant \frac{N^n (N-n)!}{N!} = \frac{N^n}{N(N-1)\cdots(N-n+1)} = \frac{1}{\prod_{i=1}^{n-1}\left(1-\frac{i}{N}\right)} \leqslant e^{\frac{n^2}{2N}}$$

where we have used the inequality

$$\ln \prod_{i=1}^{n-1}\left(1-\frac{i}{N}\right) = -\sum_{i=1}^{n-1} \ln\left(1-\frac{i}{N}\right) \geqslant -\frac{1}{N}\sum_{i=1}^{n-1} i = -\frac{(n-1)n}{N} \geqslant -\frac{n^2}{2N}.$$

Therefore, defining ε_n by (124) and

$$\beta_n = \frac{1}{\varepsilon_n}\frac{(N-n)!}{N!}\beta \in \mathcal{P}(\mathbb{R}^{dn}),$$

we obtain $\rho^s_{N:n} = (1+\varepsilon_n)\left(\rho^s_{N:1}\right)^{\otimes n} - \varepsilon_n \beta_n$, which is (123). It follows from Lemma 14 in Appendix A.1 that $W_1((\rho^s_{N:1})^{\otimes n}, \rho^s_{N:n}) = \varepsilon_n W_1((\rho^s_{N:1})^{\otimes n}, \beta_n)$. We thus obtain (126). □

Let us finish this section by the following useful lemma.

Lemma 23. *Let $\mu, \mu' \in \mathcal{P}([0,1]\times \mathbb{R}^d)$ and let y, y' their moment of order 1 as defined by ´(41), supposed finite. Then*

$$W_1(y, y') \leqslant W_1(\mu, \mu').$$

A.5 Discrepancy between the empirical measure and the ν-monokinetic measure

Recall that:

- given any $X = (x_1,\dots,x_N) \in \Omega^N$ and any $\Xi = (\xi_1,\dots,\xi_N) \in \mathbb{R}^{dN}$, the empirical measure $\mu^E_{(X,\Xi)}$ on $\Omega \times \mathbb{R}^d$ is defined by (28);
- given any $\nu \in \mathcal{P}(\Omega)$ and any measurable function $y : \Omega \to \mathbb{R}^d$, the ν-monokinetic measure μ^ν_y on $\Omega \times \mathbb{R}^d$ is defined by (45).

Lemma 24. *Let $\nu \in \mathcal{P}(\Omega)$ and let $(\mathcal{A}, X)$ be a tagged partition associated with ν (see (12)).*

(i) Let $y \in \mathscr{C}^{0,1}(\Omega, \mathbb{R}^d)$. Taking $\Xi = (\xi_1, \dots, \xi_N)$ with $\xi_i = y(x_i)$ for every $i \in \{1, \dots, N\}$, we have

$$\left|\left\langle \mu_y^\nu - \mu^E_{(X,\Xi)}, f \right\rangle\right| \leqslant \frac{C_\Omega}{N} \operatorname{Lip}(x \mapsto f(x, y(x)))$$

for every $f \in \mathscr{C}_c^\infty(\Omega \times \mathbb{R}^d)$.

(ii) Let $\Xi = (\xi_1, \dots, \xi_N) \in \mathbb{R}^d$. Defining the piecewise continuous function y_Ξ by

$$y_\Xi(x) = \sum_{i=1}^N \xi_i \mathbb{1}_{\Omega_i}(x) \qquad \forall x \in \Omega$$

so that $y_\Xi(x_i) = \xi_i$ for every $i \in \{1, \dots, N\}$, we have

$$\left|\left\langle \mu_{y_\Xi}^\nu - \mu^E_{(X,\Xi)}, f \right\rangle\right| \leqslant \frac{C_\Omega}{N} \max_{1 \leqslant i \leqslant N} \operatorname{Lip}(f(\cdot, \xi_i))$$

for every $f \in \mathscr{C}_c^\infty(\Omega \times \mathbb{R}^d)$.

Proof. Let us prove (i). We have

$$\langle \mu_y^\nu, f \rangle = \int_\Omega f(x, y(x))\, d\nu(x) = \sum_{i=1}^N \int_{\Omega_i} f(x, y(x))\, d\nu(x)$$

and

$$\left\langle \mu^E_{(X,\Xi)}, f \right\rangle = \frac{1}{N} \sum_{i=1}^N f(x_i, y(x_i)) = \sum_{i=1}^N \int_{\Omega_i} f(x_i, y(x_i))\, d\nu(x)$$

hence

$$\begin{aligned}\left|\left\langle \mu_y^\nu - \mu^E_{(X,\Xi)}, f \right\rangle\right| &\leqslant \sum_{i=1}^N \int_{\Omega_i} |f(x, y(x)) - f(x_i, y(x_i))|\, d\nu(x) \\ &\leqslant \operatorname{Lip}(x \mapsto f(x, y(x))) \sum_{i=1}^N \int_{\Omega_i} d_\Omega(x, x_i)\, d\nu(x)\end{aligned}$$

and the estimate of (i) follows because $\int_{\Omega_i} d_\Omega(x, x_i)\, d\nu(x) \leqslant \nu(\Omega_i)\operatorname{diam}(\Omega_i) \leqslant \frac{C_\Omega}{N^2}$ (using (12)).

The estimate of (ii) is proved similarly: we have

$$\langle \mu_{y_\Xi}^\nu, f \rangle = \sum_{i=1}^N \int_{\Omega_i} f(x, \xi_i)\, d\nu(x)$$

and thus

$$\left|\left\langle \mu_{y_\Xi}^\nu - \mu^E_{(X,\Xi)}, f \right\rangle\right| \leqslant \sum_{i=1}^N \int_{\Omega_i} |f(x, \xi_i) - f(x_i, \xi_i)|\, d\nu(x) \leqslant \sum_{i=1}^N \operatorname{Lip}(f(\cdot, \xi_i)) \int_{\Omega_i} d_\Omega(x, x_i)\, d\nu(x)$$

and the estimate of (ii) follows. □

Remark 13. Actually, we see from the proof that, in the estimates stated in the above lemma, it suffices that all functions of which we consider the Lipschitz constant, be Lipschitz on each subset Ω_i. In particular, they may be discontinuous at the boundary of Ω_i.

With that remark, we recover (ii) as a consequence of (i).

B Proofs

B.1 Proof of Theorem 1

We follow and extend [40, Appendix A.1: proof of Theorem 2.3]. The proof of the existence of the solution $\mu(t)$ is done first, by constructing a sequence of piecewise constant measures converging to the solution.

In the case (**B**) where G is locally Lipschitz with respect to (x, x', ξ, ξ'), there is no difference with [40] (so we do not repeat the complete argument), and the statement (**B**) and in particular the estimate (26) (as well as existence and uniqueness of solutions) is obtained by using that, for all $\mu^1, \mu^2 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$, for every $(t, x, \xi) \in \mathbb{R} \times \Omega \times \mathbb{R}^d$ fixed,

$$\left\|\mathcal{X}[\mu^1](t,x,\xi) - \mathcal{X}[\mu^2](t,x,\xi)\right\| = \left\|\int_{\Omega\times\mathbb{R}^d} G(t,x,x',\xi,\xi')\, d(\mu^1(x',\xi') - \mu^2(x',\xi'))\right\| \\ \leqslant \max_{(x,\xi)\in S} \mathrm{Lip}(G(t,x,\cdot,\xi,\cdot)_{|S})\, W_1(\mu^1,\mu^2)$$

where $S = \mathrm{supp}(\mu^1) \cup \mathrm{supp}(\mu^2)$ (compact set).

In the case (**A**), we assume that G is locally Lipschitz only with respect to (ξ, ξ') and thus the classical Wasserstein distance W_1 cannot be used as above. The main difference then comes from the following observation: given any two probability measures $\mu^1, \mu^2 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ *having the same marginal* $\nu \in \mathcal{P}_c(\Omega)$ on Ω, we have, by disintegration,

$$\mathcal{X}[\mu^1](t,x,\xi) - \mathcal{X}[\mu^2](t,x,\xi) = \int_\Omega \int_{\mathbb{R}^d} G(t,x,x',\xi,\xi')\, d(\mu^1_{x'}(\xi') - \mu^2_{x'}(\xi'))\, d\nu(x')$$

and thus

$$\left\|\mathcal{X}[\mu^1](t,x,\xi) - \mathcal{X}[\mu^2](t,x,\xi)\right\| \leqslant L^1_\nu W_1(\mu^1,\mu^2) \max_{\substack{(x,\xi)\in S \\ x'\in\mathrm{supp}(\nu)}} \mathrm{Lip}(G(t,x,x',\xi,\cdot)_{|S_{x'}}) \tag{129}$$

where $S = \mathrm{supp}(\mu^1) \cup \mathrm{supp}(\mu^2)$ (note that S and $\mathrm{supp}(\nu)$ are compact), $S_{x'} = \mathrm{supp}(\mu^1_{x'}) \cup \mathrm{supp}(\mu^2_{x'})$ (compact for any $x' \in \mathrm{supp}(\nu)$) and $L^1_\nu W_1(\mu^1,\mu^2) = \int_\Omega W_1(\mu^1_x, \mu^2_x)\, d\nu(x)$ is defined by (11). The proof of [40, (A.4) and (A.5)] is then similar, replacing W_1 with $L^1_\nu W_1$. Note that, along the construction done in [40] of the solution $\mu(t)$, for ν-almost every $x \in \Omega$ fixed, we need to consider the vector field $\mathcal{X}[\mu_t]$ (defined by (20)) only on $\mathrm{supp}(\mu(t)) = \varphi_{\mu_0}(t, \mathrm{supp}(\mu_0))$, and there, $\mathcal{X}[\mu_t](t,x,\cdot)$ is $L(t)$-Lipschitz and $\|\mathcal{X}[\mu](t,x,\xi)\| \leqslant L(t)(1 + \|\xi\|)$, where

$$L(t) = \max\left\{\|G(s,x,x',\xi,\xi')\| \;\middle|\; 0 \leqslant s \leqslant t,\ (x,\xi),(x',\xi') \in \mathrm{supp}(\mu(s))\right\} \\ + \max\left\{\mathrm{Lip}(G(s,x,x',\cdot,\cdot)_{|S_x\times S_{x'}}) \;\middle|\; 0 \leqslant s \leqslant t,\ x,x' \in \mathrm{supp}(\nu)\right\}.$$

At this step, we have obtained the existence of a solution.

Let us now establish (24) (which also entails uniqueness), that is, let us establish the item (**A₂**). Since we are going to apply Lemma 15 with $t_0 \neq 0$, for any $t_0 \in \mathbb{R}$ and any $\mu_{t_0} \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$, we denote by $t \mapsto \varphi_{\mu_{t_0}}(t, t_0, x, \cdot)$ the unique solution of $\partial_t \varphi_{\mu_{t_0}}(t,t_0,x,\cdot) = \mathcal{X}[\mu(t)](t,x) \circ \varphi_{\mu_{t_0}}(t,t_0,x,\cdot)$ such that $\varphi_{\mu_{t_0}}(t_0,t_0,x,\cdot) = \mathrm{id}_{\mathbb{R}^d}$ for ν-almost every $x \in \Omega$, where $\mu(t) = \varphi_{\mu_0}(t,t_0,\cdot,\cdot)_* \mu_{t_0}$.

With this more general notation, in view of establishing (24), let us consider two locally Lipschitz solutions $\mu^1(\cdot)$ and $\mu^2(\cdot)$ of (21) such that $\mu^1(t_0), \mu^2(t_0) \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ have the same marginal $\nu \in \mathcal{P}_c(\Omega)$ on Ω. Since $\mu^i_{t,x} = \varphi_{\mu^i_{t_0}}(t,t_0,x,\cdot)_* \mu^i_{t_0,x}$ for ν-almost every $x \in \Omega$, it follows from Lemma 15 (applied with the vector fields $\mathcal{X}[\mu^i_t](t,x,\cdot)$, for x fixed) that

$$W_1(\mu^1_{t,x}, \mu^2_{t,x}) \leqslant e^{(t-t_0)L([t_0,t])} W_1(\mu^1_{t_0,x}, \mu^2_{t_0,x}) + M([t_0,t]) \frac{e^{(t-t_0)L([t_0,t])} - 1}{L([t_0,t])} \qquad \forall t \geqslant t_0$$

where, setting $S(t) = \mathrm{supp}(\mu^1(t)) \cup \mathrm{supp}(\mu^2(t))$ (compact) and $S_x(t) = \mathrm{supp}(\mu^1_x(t)) \cup \mathrm{supp}(\mu^2_x(t))$ (compact) for any $x \in \mathrm{supp}(\nu)$,

$$L([t_0,t]) = \max\left\{\|G(s,x,x',\xi,\xi')\| \ \mid \ t_0 \leqslant s \leqslant t,\ (x,\xi),(x',\xi') \in S(s)\right\} \\ + \max\left\{\mathrm{Lip}(G(s,x,x',\cdot,\cdot)_{|S_x(s)\times S_{x'}(s)}) \ \mid \ t_0 \leqslant s \leqslant t,\ x,x' \in \mathrm{supp}(\nu)\right\}$$

and

$$M([t_0,t]) = \max\left\{\|\mathcal{X}[\mu^1_s](s,x,\xi) - \mathcal{X}[\mu^2_s](s,x,\xi)\| \ \mid \ t_0 \leqslant s \leqslant t,\ (x,\xi) \in S(s)\right\}.$$

Since μ^1_s and μ^2_s have the same marginal ν on Ω, it follows from (129) that

$$M([t_0,t]) \leqslant L([t_0,t]) \max_{t_0 \leqslant s \leqslant t} L^1_\nu W_1(\mu^1_s,\mu^2_s).$$

Therefore

$$W_1(\mu^1_{t,x},\mu^2_{t,x}) \leqslant e^{(t-t_0)L([t_0,t])} W_1(\mu^1_{t_0,x},\mu^2_{t_0,x}) + \left(e^{(t-t_0)L([t_0,t])} - 1\right) \max_{t_0 \leqslant s \leqslant t} L^1_\nu W_1(\mu^1_s,\mu^2_s).$$

Integrating with respect to $x \in \Omega$ for the measure ν, we obtain

$$L^1_\nu W_1(\mu^1(t),\mu^2(t)) \leqslant e^{(t-t_0)L([t_0,t])} L^1_\nu W_1(\mu^1(t_0),\mu^2(t_0)) + \left(e^{(t-t_0)L([t_0,t])} - 1\right) \max_{t_0 \leqslant s \leqslant t} L^1_\nu W_1(\mu^1_s,\mu^2_s).$$

Then, setting $h(t) = L^1_\nu W_1(\mu^1(t),\mu^2(t))$ and taking $t = t_0 + \varepsilon$ with $\varepsilon \geqslant 0$, we get

$$\frac{h(t_0+\varepsilon) - h(t_0)}{\varepsilon} \leqslant \frac{e^{\varepsilon L([t_0,t_0+\varepsilon])} - 1}{\varepsilon} \left(h(t_0) + \max_{t_0 \leqslant s \leqslant t_0+\varepsilon} h(s)\right)$$

and hence, taking the limit $\varepsilon \to 0$, we infer that $h'(t_0) \leqslant 2L(\{t_0\})h(t_0)$. Since t_0 is arbitrary, this implies that $h(t) \leqslant h(0)e^{2\int_0^t L(\{s\})\,ds} \leqslant h(0)e^{2L(t)}$ where $L(t) = L([0,t])$. This gives (24).

It remains to establish the item $(\mathbf{A_1})$. Let $T > 0$ be fixed, and let $\mu^k(\cdot) \in \mathscr{C}^0([0,T],\mathcal{P}_c(\Omega\times\mathbb{R}^d))$ be a sequence of solutions of the Vlasov equation, such that $\mu^k_0 = \mu^k(0)$ converges weakly to $\mu_0 = \mu(0)$ as $k \to +\infty$. Our objective is to prove that $\mu^k(t)$ converges weakly to $\mu(t)$, uniformly with respect to $t \in [0,T]$.

Let us denote by $\mathscr{C}_0(\Omega \times \mathbb{R}^d)$ the Banach space of continuous functions on $\Omega \times \mathbb{R}^d$ vanishing at infinity, and by $\mathcal{M}^1(\Omega \times \mathbb{R}^d) = (\mathscr{C}_0(\Omega \times \mathbb{R}^d))'$ the Banach space of Radon measures on $\Omega \times \mathbb{R}^d$, endowed with the total variation norm $\|\ \|_{TV}$ (which is the dual norm). Of course, we have $\mathcal{P}_c(\Omega \times \mathbb{R}^d) \subset \mathcal{M}^1(\Omega \times \mathbb{R}^d)$.

Since $\mu^k(t)$ is a probability measure for every $t \in [0,T]$, we have $\|\mu^k(t)\|_{TV} = 1 < +\infty$, and thus the sequence $(\mu^k(\cdot))_{k\in\mathbb{N}^*}$ is bounded in $L^\infty([0,T],\mathcal{M}^1(\Omega \times \mathbb{R}^d))$ (for the strong topology).

Second, recall the general fact, well known in Bochner integral theory, that $(L^1([0,T],E)' = L^\infty([0,T],E')$ (isometric isomorphism) for any Banach space E such that E' is separable. Applying this fact to the Banach space $E = \mathscr{C}_0(\Omega \times \mathbb{R}^d)$, observing that $E' = \mathcal{M}^1(\Omega \times \mathbb{R}^d)$ is separable (because the set of rational convex combinations of Dirac measures over points with rational coordinates is dense in it), we have $(L^1([0,T],\mathscr{C}_0(\Omega \times \mathbb{R}^d)))' = L^\infty([0,T],\mathcal{M}^1(\Omega \times \mathbb{R}^d))$.

Therefore, the sequence $(\mu^k(\cdot))_{k\in\mathbb{N}^*}$ is bounded in $(L^1([0,T],\mathscr{C}_0(\Omega\times\mathbb{R}^d)))'$ for the strong (dual norm) topology. By the Banach-Alaoglu theorem, let $\tilde{\mu}(\cdot) \in L^\infty([0,T],\mathcal{M}^1(\Omega \times \mathbb{R}^d))$ be any limit point of that sequence: there exists a subsequence of $(\mu^k(\cdot))_{k\in\mathbb{N}^*}$ converging to $\tilde{\mu}(\cdot)$ for the weak star topology.

Now, for every $k \in \mathbb{N}^*$, $\mu^k(\cdot)$ is solution of the Vlasov equation $\partial_t \mu^k + L_{\mathcal{X}[\mu^k]}\mu^k = 0$, which we write in the (time) integrated weaker form, by integrating against any $f \in \mathscr{C}_c^0([0,T] \times \Omega \times \mathbb{R}^d)$, as

$$\int_0^T \int_{\Omega\times\mathbb{R}^d} f(t,x,\xi)\, d\mu_t^k(x,\xi)\, dt = \int_0^T \int_{\Omega\times\mathbb{R}^d} f(t,x,\xi)\, d\mu_0^k(x,\xi)\, dt \\ + \int_0^T \int_0^t \int_{\Omega\times\mathbb{R}^d} \langle \nabla_\xi f(t,x,\xi), G(s,x,x',\xi,\xi')\rangle\, d\mu_s^k(x',\xi')\, d\mu_s^k(x,\xi)\, dt.$$

Passing to the limit, we obtain

$$\int_0^T \int_{\Omega\times\mathbb{R}^d} f(t,x,\xi)\, d\tilde{\mu}_t(x,\xi)\, dt = \int_0^T \int_{\Omega\times\mathbb{R}^d} f(t,x,\xi)\, d\mu_0(x,\xi)\, dt \\ + \int_0^T \int_0^t \int_{\Omega\times\mathbb{R}^d} \langle \nabla_\xi f(t,x,\xi), G(s,x,x',\xi,\xi')\rangle\, d\tilde{\mu}_s(x',\xi')\, d\tilde{\mu}_s(x,\xi)\, dt$$

for any $f \in \mathscr{C}_c^0([0,T] \times \Omega \times \mathbb{R}^d)$, and thus $\partial_t \tilde{\mu} + L_{\mathcal{X}[\tilde{\mu}]}\tilde{\mu} = 0$, with $\tilde{\mu}(0) = \mu(0)$. By uniqueness (already proved earlier), we infer that $\tilde{\mu} = \mu$. Since all limit points coincide, this shows that the sequence of $(\mu^k(\cdot))_{k\in\mathbb{N}^*}$ converges to $\mu(\cdot)$ for the weak star topology. This finishes the proof of the theorem.

B.2 Proof of Theorem 2

We have $\rho(t) = \delta_X \otimes \delta_{\Xi(t)}$. By (122) in Lemma 22 of Appendix A.4 (applied with $\gamma_i = \delta_{\xi_i(t)}$), we have

$$\rho(t)^s_{N:1} = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \delta_{\xi_i(t)} = \mu^E_{(X,\Xi(t))},$$

the empirical measure, which gives the preliminary remark to Theorem 2. The statement **(A)** for $n = 1$ then follows from the item **($\mathbf{A_1}$)** of Theorem 1, and the estimate (32) follows from the item **(B)** of Theorem 1. This is the case $n = 1$ of the argument below.

The nth-order marginal $\rho(t)^s_{N:n}$ is given by (123) in Appendix A.4 with $\gamma_i = \delta_{\xi_i(t)}$. By the triangular inequality, we have

$$W_1\left(\rho(t)^s_{N:n}, \mu(t)^{\otimes n}\right) \leqslant W_1\left(\rho(t)^s_{N:n}, \left(\mu^E_{(X,\Xi(t))}\right)^{\otimes n}\right) + W_1\left(\left(\mu^E_{(X,\Xi(t))}\right)^{\otimes n}, \mu(t)^{\otimes n}\right) \tag{130}$$

For the first term of the right-hand side of (130), since $\mu^E_{(X,\Xi(t))} = \rho(t)^s_{N:1}$, we infer from (126) in Lemma 22 of Appendix A.4 that

$$W_1\left(\rho(t)^s_{N:n}, \left(\mu^E_{(X,\Xi(t))}\right)^{\otimes n}\right) = W_1\left(\rho(t)^s_{N:n}, (\rho(t)^s_{N:1})^{\otimes n}\right) \leqslant \left(e^{\frac{n^2}{2N}} - 1\right) W_1\left((\rho(t)^s_{N:1})^{\otimes n}, \beta_n\right)$$

for every $n \in \{1,\dots,N\}$, and it follows from (125) (with $\gamma_i = \delta_{\xi_i(t)}$) and from Lemma 12 in Appendix A.1 that

$$W_1\left((\rho(t)^s_{N:1})^{\otimes n}, \beta_n\right) = W_1\left(\left(\mu^E_{(X,\Xi(t))}\right)^{\otimes n}, \beta_n\right) \leqslant 2\max(1, \|\Xi(t)\|_\infty),$$

hence

$$W_1\left(\rho(t)^s_{N:n}, \left(\mu^E_{(X,\Xi(t))}\right)^{\otimes n}\right) \leqslant 2\left(e^{\frac{n^2}{2N}} - 1\right)\max(1, \|\Xi(t)\|_\infty). \tag{131}$$

For the second term of the right-hand side of (130), as a consequence of Lemma 13 in Appendix A.1 and of Theorem 1, we have

$$W_1\left(\left(\mu^E_{(X,\Xi(t))}\right)^{\otimes n}, \mu(t)^{\otimes n}\right) \leqslant n\, W_1\left(\mu^E_{(X,\Xi(t))}, \mu(t)\right)$$
$$\leqslant n\, C\left(\mu^E_{(X,\Xi(t))}, \mu(t)\right) W_1\left(\mu^E_{(X,\Xi(0))}, \mu(0)\right). \quad (132)$$

The real number $C(\mu^E_{(X,\Xi(t))}, \mu(t))$ defined by (25) in Theorem 1, coincides with the real number $C(\mu(t))$ defined by (31), because $\Xi(t) \in \mathrm{supp}(\mu(t))$. Also, $\|\Xi(t)\|_\infty \leqslant \|\mathrm{supp}(\mu(t))\|_\infty$. Therefore, (33) follows from (130), (131) and (132). Note that, for $n = 1$, the first term of the right-hand side of (130) is equal to 0, which gives (32).

The statement **(A)** for any $n \in \mathbb{N}^*$ also follows, by replacing the right inequality in (132) with the application of the item **(A)** of Theorem 1.

B.3 Proof of Theorem 3

First of all, since $\rho_0 = \delta_X \otimes \rho_{0,X}$, with $\delta_X = \delta_{x_1} \otimes \cdots \otimes \delta_{x_N}$ and $\rho_{0,X} = \mu_{0,x_1} \otimes \cdots \otimes \mu_{0,x_N}$, it follows from (122) in Lemma 22 of Appendix A.4 (applied with $\gamma_i = \mu_{0,x_i}$) that $(\rho_0)^s_{N:1} = (\mu_0)^{SE}_X$ (semi-empirical measure), which gives (36) (and the weak convergence to μ_0 is obtained by Lemma 20 of Appendix A.3.2), which is the preliminary remark to the theorem.

We are going to prove the theorem by using approximation of probability measures by empirical measures, and then use Theorem 2, as follows. Since $(\mu_0)^{SE}_X = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \mu_{0,x_i}$, by Lemma 17 in Appendix A.2, for every $i \in \{1,\ldots,N\}$ the measure $\mu_{0,x_i} \in \mathcal{P}_1(\mathbb{R}^d)$ is the limit, in Wasserstein distance W_1, of a sequence of empirical measures:

$$\lim_{K\to+\infty} W_1(\gamma^K_i, \mu_{0,x_i}) = 0 \qquad \text{where} \quad \gamma^K_i = \frac{1}{K}\sum_{j=1}^K \delta_{\xi^K_{i,j}},$$

where $(\xi^K_{i,1},\ldots,\xi^K_{i,K}) \in (\mathbb{R}^d)^K$ with $\xi^K_{i,j} \in \mathrm{supp}(\mu_{0,x_i})$, for all $K \in \mathbb{N}^*$, $j \in \{1,\ldots,K\}$ and $i \in \{1,\ldots,N\}$. Therefore, on the one hand,

$$\lim_{K\to+\infty} W_1\left(\frac{1}{NK}\sum_{i=1}^N\sum_{j=1}^K \delta_{x_i} \otimes \delta_{\xi^K_{i,j}}, (\mu_0)^{SE}_X\right) = 0, \quad (133)$$

i.e., $(\mu_0)^{SE}_X$ is the limit, in Wasserstein distance W_1, of a sequence of empirical measures. On the other hand, setting

$$\Gamma^K = \bigotimes_{i=1}^N \gamma^K_i = \bigotimes_{i=1}^N \left(\frac{1}{K}\sum_{j=1}^K \delta_{\xi^K_{i,j}}\right) = \frac{1}{K^N}\sum_{(j_1,\ldots,j_N)\in\{1,\ldots,K\}^N} \delta_{\xi^K_{1,j_1}} \otimes \cdots \otimes \delta_{\xi^K_{N,j_N}}$$
$$= \frac{1}{K^N}\sum_{J=(j_1,\ldots,j_N)\in\{1,\ldots,K\}^N} \delta_{\Xi^K_J} \qquad \text{with} \quad \Xi^K_J = (\xi^K_{1,j_1},\ldots,\xi^K_{N,j_N})$$

which is an empirical measure on $\mathbb{R}^{dN}$, and setting $\rho^K_0 = \delta_X \otimes \Gamma^K$, we have $\rho^K_{0,X} = \Gamma^K$ and

$$\lim_{K\to+\infty} W_1(\rho^K_{0,X}, \rho_{0,X}) = 0 \qquad \text{and} \qquad \lim_{K\to+\infty} W_1(\rho^K_0, \rho_0) = 0. \quad (134)$$

Propagating ρ_0^K under the flow $\Phi(t)$ of the particle system (17), since we handle Dirac masses, the unique solution of the Liouville equation (29) such that $\rho^K(0) = \rho_0^K$ is given by $\rho^K(t) = \Phi(t)_*\rho_0^K = \delta_X \otimes \Gamma^K(t)$ with

$$\Gamma^K(t) = \bigotimes_{i=1}^{N} \gamma_i^K(t) \qquad \text{and} \qquad \gamma_i^K(t) = \mu^E_{\Xi^i(t)} = \frac{1}{K}\sum_{j=1}^{K} \delta_{\xi^K_{i,j}(t)}$$

where, for every $j \in \{1, \dots, K\}$, $t \mapsto (\xi^K_{1,j}(t), \dots, \xi^K_{N,j}(t))$ is the unique solution of the particle system (17) such that $\xi^K_{i,j}(0) = \xi^K_{i,j}$ for $i = 1, \dots, N$. By (122) in Lemma 22 of Appendix A.4, we have

$$\rho^K(t)^s_{N:1} = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i} \otimes \gamma_i^K(t) = \frac{1}{NK}\sum_{i=1}^{N}\sum_{j=1}^{K} \delta_{x_i} \otimes \delta_{\xi^K_{i,j}(t)} \tag{135}$$

which is an empirical measure on $\Omega \times \mathbb{R}^d$. Now, by the triangular inequality,

$$W_1\left(\rho(t)^s_{N:1}, \mu(t)\right) \leqslant W_1\left(\rho(t)^s_{N:1}, \rho^K(t)^s_{N:1}\right) + W_1\left(\rho^K(t)^s_{N:1}, \mu(t)\right). \tag{136}$$

Let us estimate the first term of the right-hand side of (136). First, it follows from Lemmas 9 and 10 that

$$W_1\left(\rho(t)^s_{N:1}, \rho^K(t)^s_{N:1}\right) \leqslant W_1(\rho(t), \rho^K(t)) = W_1(\rho_X(t), \rho^K_X(t))$$

where the equality at the right-hand side above straightforwardly follows from the fact that $\rho(t)$ and $\rho^K(t)$ have the same marginal δ_X. Now, applying Lemma 15 in Appendix A.1, we have

$$W_1(\rho_X(t), \rho^K_X(t)) \leqslant e^{tL(t)}\, W_1(\rho_{0,X}, \rho^K_{0,X})$$

where $L(t)$ is the maximal Lipschitz constant of the vector field $Y(s, X, \cdot)$ restricted to $\mathrm{supp}(\rho(s))$ for $0 \leqslant s \leqslant t$ (note, indeed, that $\mathrm{supp}(\rho_0^K) \subset \mathrm{supp}(\rho_0)$). Hence we have

$$W_1\left(\rho(t)^s_{N:1}, \rho^K(t)^s_{N:1}\right) \leqslant e^{tL(t)}\, W_1(\rho_0, \rho_0^K). \tag{137}$$

Let us now estimate the second term of the right-hand side of (136). In the case where G is locally Lipschitz with respect to (x, x', ξ, ξ'), since $\rho^K(t)^s_{N:1}$ is the empirical measure given by (135) and thus is solution of the Vlasov equation, it follows from the item **(B)** of Theorem 1 that

$$W_1\left(\rho^K(t)^s_{N:1}, \mu(t)\right) \leqslant C(\mu(t))\, W_1\left(\left(\rho_0^K\right)^s_{N:1}, \mu_0\right) \tag{138}$$

where $C(\mu(t))$ is defined by (31) (same argument as at the end of the proof of Theorem 2). In the case where G is continuous with respect to (x, x', ξ, ξ') and locally Lipschitz with respect to (ξ, ξ'), we will hereafter replace the estimate (138) with the fact that $W_1\left(\left(\rho_0^K\right)^s_{N:1}, \mu_0\right) \to 0$ implies, by the item **($\mathbf{A_1}$)** of Theorem 1, that $W_1\left(\rho^K(t)^s_{N:1}, \mu(t)\right) \to 0$ as $K \to +\infty$.

Finally, when G is locally Lipschitz with respect to (x, x', ξ, ξ'), we infer from (136), (137) and (138) that

$$\begin{aligned} W_1\left(\rho(t)^s_{N:1}, \mu(t)\right) &\leqslant e^{tL(t)}\, W_1(\rho_0, \rho_0^K) + C(\mu(t))\, W_1\left(\left(\rho_0^K\right)^s_{N:1}, \mu_0\right) \\ &\leqslant e^{tL(t)}\, W_1(\rho_0, \rho_0^K) + C(\mu(t))\, W_1\left(\left(\rho_0^K\right)^s_{N:1}, (\mu_0)^{SE}_X\right) + C(\mu(t))\, W_1\left((\mu_0)^{SE}_X, \mu_0\right) \end{aligned}$$

where we have used, again, the triangular inequality. Now, taking the limit $K \to +\infty$, the first at the right-hand side of the above inequality disappears by (134); the second term disappears by

(133) (indeed, by (135), $\left(\rho_0^K\right)^s_{N:1}$ is exactly the empirical measure appearing in (133)). We thus obtain (37) by applying the second item of Lemma 20 of Appendix A.3.2 to the third term.

When G is continuous with respect to (x, x', ξ, ξ') and locally Lipschitz with respect to (ξ, ξ'), the above argument is adapted by saying that $W_1\left((\mu_0)^{SE}_X, \mu_0\right) \to 0$ as $N \to +\infty$ (which follows from the first item of Lemma 20 of Appendix A.3.2) implies that $W_1\left(\rho^K(t)^s_{N:1}, \mu(t)\right) \to 0$, which gives the statement (**A**) of the theorem for $n = 1$.

Let us now establish (38). The nth-order marginal $\rho^K(t)^s_{N:n}$ is given by (123) in Appendix A.4 with $\gamma_i = \gamma_i^K(t)$. By the triangular inequality, we have

$$W_1\left(\rho(t)^s_{N:n}, \mu(t)^{\otimes n}\right) \leqslant W_1\left(\rho(t)^s_{N:n}, \rho^K(t)^s_{N:n}\right) + W_1\left(\rho^K(t)^s_{N:n}, (\rho(t)^s_{N:1})^{\otimes n}\right) + W_1\left((\rho(t)^s_{N:1})^{\otimes n}, \mu(t)^{\otimes n}\right). \quad (139)$$

Let us estimate the three terms at the right-hand side of the inequality (139). For the first term, we infer from Lemmas 9, 10 and 15 in Appendix A.1 that

$$W_1\left(\rho(t)^s_{N:n}, \rho^K(t)^s_{N:n}\right) \leqslant W_1\left(\rho(t), \rho^K(t)\right) \leqslant e^{tL(t)}\, W_1(\rho_0, \rho_0^K). \quad (140)$$

Since $\rho^K(t)^s_{N:n} = (1+\varepsilon_n)\,(\rho(t)^s_{N:1})^{\otimes n} - \varepsilon_n \beta_n$, using (126) in Lemma 22 of Appendix A.4 and then Lemma 12 in Appendix A.1, the second term is estimated by

$$W_1\left(\rho^K(t)^s_{N:n}, (\rho(t)^s_{N:1})^{\otimes n}\right) \leqslant \varepsilon_n W_1\left((\rho(t)^s_{N:1})^{\otimes n}, \beta_n\right) \leqslant \left(e^{\frac{n^2}{2N}} - 1\right)\left(\max\left(1, \|\Xi(t)\|_\infty\right) + \|\text{supp}(\mu(t))\|_\infty\right) \quad (141)$$

where we have also used that $\text{supp}(\rho(t)^s_{N:1}) \subset \text{supp}(\mu(t))$. Finally, for the third term, it follows from Lemma 13 that

$$W_1\left((\rho(t)^s_{N:1})^{\otimes n}, \mu(t)^{\otimes n}\right) \leqslant W_1\left(\rho(t)^s_{N:1}, \mu(t)\right). \quad (142)$$

Therefore, from (139), (140), (141) and (142), we obtain

$$W_1\left(\rho(t)^s_{N:n}, \mu(t)^{\otimes n}\right) \leqslant e^{tL(t)}\, W_1(\rho_0, \rho_0^K) + \left(e^{\frac{n^2}{2N}} - 1\right)\left(\max\left(1, \|\Xi(t)\|_\infty\right) + \|\text{supp}(\mu(t))\|_\infty\right) + W_1\left(\rho(t)^s_{N:1}, \mu(t)\right).$$

Taking the limit $K \to +\infty$, the first term at the right-hand side of the above inequality disappears. Using that $\max\left(1, \|\Xi(t)\|_\infty\right) \leqslant 1 + \|\text{supp}(\mu(t))\|_\infty$, and using the estimate (37) obtained for $n = 1$, the estimate (38) follows.

Finally, the statement (**A**) for any n is obtained by a reasoning that is similar to the case $n = 1$, by adapting the above argument.

B.4 Proof of Corollary 4

Following the proof of Corollary 1, in the indistinguishable case $\bar{\mu}(\cdot)$ is solution of the Vlasov equation (21) (without dependence on x) if and only if $\mu(\cdot) = \bar{\nu} \otimes \bar{\mu}(\cdot)$ is solution of the Vlasov equation (21). Here, $\bar{\nu}$ is an arbitrary probability measure on Ω that is absolutely continuous with respect to a Lebesgue measure. We have then $\mu_{0,x_i} = \bar{\mu}_0$ for every $i \in \{1, \ldots, N\}$. Therefore, the initial condition ρ_0 that is considered in Theorem 3 is $\rho_0 = \delta_X \otimes \bar{\mu}_0^{\otimes N}$. Note that ρ_0 differs from

$\bar\nu \otimes \bar\mu_0^{\otimes N}$ which would have been the embedding considered previously. We thus have something additional to prove here.

With this ρ_0, we have $(\rho_0)^s_{N:n} = (\delta_X)^s_{N:n} \otimes \bar\mu_0^{\otimes n}$ for every $n \in \{1, \dots, N\}$. Note moreover that $(\delta_X)^s_{N:1} = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} = \nu_X^E$ (see Lemma 22 in Appendix A.4). Besides, we have $\rho(t) = \delta_X \otimes \bar\rho(t)$ where $\bar\rho(t) = \Phi(t)_* \bar\mu_0^{\otimes N}$ is the unique solution of the Liouville equation (29) (without dependence on X) such that $\bar\rho(0) = \bar\mu_0^{\otimes N}$. Since $\bar\rho(t)^s_{N:n} = \bar\rho(t)_{N:n}$ because the flow preserves indistinguishability, we infer that $\rho(t)^s_{N:n} = (\delta_X)^s_{N:n} \otimes \bar\rho(t)_{N:n}$.

We now apply the item (B) of Theorem 3, with $L = 0$ because $\mu_{0,x} = \bar\mu_0$ does not depend on x. Since $\mu(t)^{\otimes n} = \bar\nu^{\otimes n} \otimes \bar\mu(t)^{\otimes n}$, the estimate (37) gives

$$W_1(\nu_X^E \otimes \bar\rho(t)_{N:1}, \bar\nu \otimes \bar\mu(t)) \leqslant \frac{C(\bar\mu(t))}{N} \tag{143}$$

and the estimate (38) gives, for every $n \in \{2, \dots, N\}$,

$$W_1((\delta_X)^s_{N:n} \otimes \bar\rho(t)_{N:n}, \bar\nu^{\otimes n} \otimes \bar\mu(t)^{\otimes n}) \leqslant \left(e^{\frac{n^2}{2N}} - 1\right)(1 + 2\|\mathrm{supp}(\bar\mu(t))\|_\infty) + \frac{1}{N}C(\bar\mu(t)). \tag{144}$$

It remains to prove that (143) implies (39) and that (144) implies (40).

Using Lemma 11 in Appendix A.1, the triangular inequality, and then Lemma 11 again, we have

$$\begin{aligned}
W_1(\bar\rho(t)_{N:1}, \bar\mu(t)) &= W_1(\nu_X^E \otimes \bar\rho(t)_{N:1}, \nu_X^E \otimes \bar\mu(t)) \\
&\leqslant W_1(\nu_X^E \otimes \bar\rho(t)_{N:1}, \bar\nu \otimes \bar\mu(t)) + W_1(\bar\nu \otimes \bar\mu(t), \nu_X^E \otimes \bar\mu(t)) \\
&= W_1(\nu_X^E \otimes \bar\rho(t)_{N:1}, \bar\nu \otimes \bar\mu(t)) + W_1(\bar\nu, \nu_X^E)
\end{aligned}$$

and then, using (39) and Lemma 19 in Appendix A.3, we infer (143).

Now, similarly, using Lemma 11 in Appendix A.1, the triangular inequality, then Lemma 11 again, and then (144), we have

$$\begin{aligned}
W_1(\bar\rho(t)_{N:n}, \bar\mu(t)^{\otimes n}) &= W_1((\delta_X)^s_{N:n} \otimes \bar\rho(t)_{N:n}, (\delta_X)^s_{N:n} \otimes \bar\mu(t)^{\otimes n}) \\
&\leqslant W_1((\delta_X)^s_{N:n} \otimes \bar\rho(t)_{N:n}, \bar\nu^{\otimes n} \otimes \bar\mu(t)^{\otimes n}) \\
&\qquad + W_1(\bar\nu^{\otimes n} \otimes \bar\mu(t)^{\otimes n}, (\delta_X)^s_{N:n} \otimes \bar\mu(t)^{\otimes n}) \\
&= W_1((\delta_X)^s_{N:n} \otimes \bar\rho(t)_{N:n}, \bar\nu^{\otimes n} \otimes \bar\mu(t)^{\otimes n}) + W_1(\bar\nu^{\otimes n}, (\delta_X)^s_{N:n}) \\
&\leqslant \left(e^{\frac{n^2}{2N}} - 1\right)(1 + 2\|\mathrm{supp}(\bar\mu(t))\|_\infty) + \frac{1}{N}C(\bar\mu(t)) + W_1(\bar\nu^{\otimes n}, (\delta_X)^s_{N:n})
\end{aligned}$$

Now, to obtain (40), it suffices to prove that $W_1(\bar\nu^{\otimes n}, (\delta_X)^s_{N:n}) \leqslant 2(e^{\frac{n^2}{2N}} - 1)$. Using the same argument as in the proof of Lemma 22 in Appendix A.4, just ignoring the terms γ_j, since $(\delta_X)^s_{N:1} = \nu_X^E$, we have (using (123)) $(\delta_X)^s_{N:n} = (1 + \varepsilon_n)\left(\nu_X^E\right)^{\otimes n} - \varepsilon_n \beta_n$ where $\varepsilon_n \in [0, e^{\frac{n^2}{2N}} - 1]$ is defined by (124) and $\beta_n = \frac{1}{\varepsilon_n}\frac{(N-n)!}{N!}\sum \delta_{x_{i_1}} \otimes \cdots \otimes \delta_{x_{i_n}} \in \mathcal{P}(\mathbb{R}^{dn})$ where the sum is taken over all n-tuples $(i_1, \dots, i_n) \in \{1, \dots, N\}^n$ for which at least two elements are equal. Then, using the estimate (126) of Lemma 22 (Appendix A.4) and then Lemma 12 in Appendix A.1, we get

$$W_1\left((\delta_X)^s_{N:n}, (\nu_X^E)^{\otimes n}\right) \leqslant \varepsilon_n W_1\left((\nu_X^E)^{\otimes n}, \beta_n\right) \leqslant 2\left(e^{\frac{n^2}{2N}} - 1\right).$$

This finishes the proof.

B.5 Proof of Theorem 4

We start by proving the second item. Hence, we assume that G is locally α-Hölder continuous with respect to (x, x', ξ, ξ') (uniformly with respect to t on any compact).

Lemma 25. *Let $x, x' \in \Omega$ be arbitrary. We have*

$$\|y(t,x) - y(t,x')\| \leqslant e^{tL^y(t)} \left(\|y^0(x) - y^0(x')\| + d_\Omega(x,x')^\alpha\right) \qquad \forall t \geqslant 0. \tag{145}$$

Proof of Lemma 25. By definition, we have $\partial_t y(t,z) = \int_\Omega G(t,z,x'',y(t,z),y(t,x''))\,d\nu(x'')$ for every $z \in \Omega$, hence

$$\begin{aligned}\partial_t y(t,x) - \partial_t y(t,x') &= \int_\Omega G(t,x,x'',y(t,x),y(t,x''))\,d\nu(x'') - \int_\Omega G(t,x',x'',y(t,x),y(t,x''))\,d\nu(x'') \\ &+ \int_\Omega G(t,x',x'',y(t,x),y(t,x''))\,d\nu(x'') - \int_\Omega G(t,x',x'',y(t,x'),y(t,x''))\,d\nu(x'')\end{aligned} \tag{146}$$

and using (54) we obtain

$$\|\partial_t(y(t,x) - y(t,x'))\| \leqslant L^y(t)\left(d_\Omega(x,x')^\alpha + \|y(t,x) - y(t,x')\|\right)$$

and (145) follows by integration (noting that $s \mapsto L^y(s)$ is nondecreasing). □

By assumption, $\|y^0(x) - y^0(x')\| \leqslant \mathrm{Hol}_\alpha(y^0) d_\Omega(x,x')^\alpha$ for all $x, x' \in \Omega$, hence, using (145) in Lemma 25 we infer that $y(t,\cdot)$ is α-Hölder continuous and (51) follows.

Let us establish (52). We set $r_i(t) = y(t,x_i) - \xi_i(t)$, for $i = 1, \dots, N$. By definition, we have

$$\dot{r}_i(t) = \frac{1}{N}\sum_{j=1}^N \left(G(t,x_i,x_j,y(t,x_i),y(t,x_j)) - G(t,x_i,x_j,\xi_i(t),\xi_j(t))\right) + \epsilon_i(t) \tag{147}$$

where

$$\epsilon_i(t) = \int_\Omega G(t,x_i,x',y(t,x_i),y(t,x'))\,d\nu(x') - \frac{1}{N}\sum_{j=1}^N G(t,x_i,x_j,y(t,x_i),y(t,x_j)) \tag{148}$$

with $r_i(0) = 0$, for every $i \in \{1, \dots, N\}$. We have on the one part

$$\|G(t,x_i,x_j,y(t,x_i),y(t,x_j)) - G(t,x_i,x_j,\xi_i(t),\xi_j(t))\| \leqslant L(t)(\|r_i(t)\| + \|r_j(t)\|) \tag{149}$$

where $L(t)$ is defined by (54), and on the other part, using (114) in Lemma 19 (see Appendix A.3),

$$\|\varepsilon_i(t)\| \leqslant \frac{C_\Omega^\alpha}{N^\alpha}\,\mathrm{Hol}_\alpha(x' \mapsto G(t,x_i,x',y(t,x_i),y(t,x')))$$

and we estimate

$$\mathrm{Hol}_\alpha(x' \mapsto G(t,x_i,x',y(t,x_i),y(t,x'))) \leqslant L^y(t)(1 + e^{tL^y(t)}(\mathrm{Hol}_\alpha(y^0) + 1)).$$

Indeed, writing for short $g(x', y(t,x')) = G(t,x_i,x',y(t,x_i),y(t,x'))$, we have

$$\begin{aligned}&\|g(x_1', y(t,x_1')) - g(x_2', y(t,x_2'))\| \\ &\leqslant \|g(x_1', y(t,x_1')) - g(x_2', y(t,x_1'))\| + \|g(x_2', y(t,x_1')) - g(x_2', y(t,x_2'))\| \\ &\leqslant L^y(t) d_\Omega(x_1',x_2')^\alpha + L^y(t)\|y(t,x_1') - y(t,x_2')\| \\ &\leqslant L^y(t) d_\Omega(x_1',x_2')^\alpha + L^y(t)\,\mathrm{Hol}_\alpha(y(t,\cdot)) d_\Omega(x_1',x_2')^\alpha\end{aligned}$$

and the estimate follows by using (51). Finally, setting $R(t) = (r_1(t), \dots, r_N(t))$, noting that $L^y(t) \leqslant L(t)$, we infer from (147) that

$$\frac{d}{dt}\|R(t)\|_\infty \leqslant \|\dot{R}(t)\|_\infty \leqslant L(t)\left(2\|R(t)\|_\infty + \frac{C_\Omega^\alpha}{N^\alpha}(1 + e^{tL(t)}(\mathrm{Hol}_\alpha(y^0) + 1))\right)$$

and, using (54) (noting again that $s \mapsto L(s)$ is nondecreasing) and by integration, we obtain (52).

Let us establish (53). For every $x \in \Omega$ there exists $i \in \{1, \dots, N\}$ such that $x \in \Omega_i$, and thus $d_\Omega(x, x_i) \leqslant \mathrm{diam}(\Omega_i) \leqslant \frac{C_\Omega}{N}$ (by (12)). It follows from (51) that

$$\|y(t,x) - y(t,x_i)\| \leqslant \mathrm{Hol}_\alpha(y(t,\cdot))d_\Omega(x,x_i)^\alpha \leqslant \frac{C_\Omega^\alpha}{N^\alpha}e^{tL(t)}\left(\mathrm{Hol}_\alpha(y(0,\cdot)) + 1\right)$$

and, noting that $y_{\Xi(t)}(x) = \xi_i(t)$, (53) follows by the triangular inequality, using (52).

Let us now prove the first item. Starting as in the proof of Lemma 25, by continuity of G, we infer from (146) that, for any $\varepsilon > 0$, if x and x' are sufficiently close then

$$\|\partial_t(y(t,x) - y(t,x'))\| \leqslant L(t)\left(\varepsilon + \|y(t,x) - y(t,x')\|\right)$$

and by integration we obtain

$$\|y(t,x) - y(t,x')\| \leqslant e^{tL(t)}\left(\|y^0(x) - y^0(x')\| + \varepsilon\right). \tag{150}$$

By assumption, y^0 is continuous ν-almost everywhere on Ω. It follows from (150) that, for every $t \geqslant 0$, $y(t,\cdot)$ is continuous ν-almost everywhere on Ω with the same continuity set as y^0 (thus, not depending on t).

Let us finally establish (50). By the Riemann integration theorem (see (13)), we have $\varepsilon_i(t) = \mathrm{o}(1)$ (where $\varepsilon_i(t)$ is defined by (148)) as $N \to +\infty$, uniformly with respect to t on every compact. Besides, we still have the inequality (149), but with $L(t)$ replaced by

$$\max_{\substack{x,x' \in \Omega \\ 0 \leqslant s \leqslant t}} \mathrm{Lip}(G(s,x,x',\cdot,\cdot)_{|S(s)\times S(s)}.$$

With this substitution, we obtain

$$\frac{d}{dt}\|R(t)\|_\infty \leqslant \|\dot{R}(t)\|_\infty \leqslant L(t)\left(2\|R(t)\|_\infty + \mathrm{o}(1)\right)$$

and integrating we get $\|R(t)\|_\infty \leqslant e^{2tL(t)}\mathrm{o}(1)$, which yields (50). Then, (55) follows by the triangular inequality, using the ν-almost everywhere continuity of $y(t,\cdot)$.

B.6 Proof of Theorem 5

The proof is a slight adaptation of the proof of Theorem 4. We start by establishing (56). Hence, we assume that G is locally α-Hölder continuous with respect to (x, x', ξ, ξ') (uniformly with respect to t on any compact).

Lemma 26. *Let $i \in \{1, \dots, N\}$ and $x, x' \in \Omega_i$ be arbitrary. We have*

$$\|y^N(t,x) - y^N(t,x')\| \leqslant e^{tL(t)}d_\Omega(x,x')^\alpha \tag{151}$$

where $y^0(\cdot) = y(0,\cdot)$.

Proof. Following the proof of Lemma 25, we arrive at

$$\|\partial_t(y^N(t,x)-y^N(t,x'))\| \leqslant L(t)\left(d_\Omega(x,x')^\alpha + \|y^N(t,x)-y^N(t,x')\|\right)$$

and (151) follows by integration, noting that $y^N(0,x)-y^N(0,x')=0$ if $x,x'\in\Omega_i$. □

It follows from Lemma 25 that $y^N(t,\cdot)$ is α-Hölder continuous in each Ω_i, with Hölder constant $e^{tL(t)}$.

We set $r(t,x)=y^N(t,x)-y_{\Xi(t)}(x)$ for every $x\in\Omega$. By definition, if $x\in\Omega_i$ then $y_{\Xi(t)}(x)=\xi_i(t)$ and thus

$$\partial_t r(t,x) = \frac{1}{N}\sum_{j=1}^N \left(G(t,x,x_j,y^N(t,x),y^N(t,x_j)) - G(t,x_i,x_j,\xi_i(t),\xi_j(t))\right) + \epsilon(t,x)$$

where

$$\epsilon(t,x) = \int_\Omega G(t,x,x',y^N(t,x),y^N(t,x'))\,d\nu(x') - \frac{1}{N}\sum_{j=1}^N G(t,x,x_j,y^N(t,x),y^N(t,x_j))$$

with $r(0,x)=0$. We have on the one part, for every $x\in\Omega_i$,

$$\begin{aligned}\big\|G(t,x,x_j,y^N(t,x),y^N(t,x_j)) - G(t,x_i,x_j,\xi_i(t),\xi_j(t))\big\| \\ \leqslant L(t)(d_\Omega(x,x_i)^\alpha + \|r(t,x)\| + \|r(t,x_j)\|)\end{aligned}$$

and on the other part, proceeding like in the proof of Theorem 4, for every $x\in\Omega_i$,

$$\|\varepsilon(t,x)\| \leqslant \frac{C_\Omega^\alpha}{N^\alpha}L(t)(1+e^{tL(t)})$$

Using that $d_\Omega(x,x_i)\leqslant \mathrm{diam}(\Omega_i)\leqslant \frac{C_\Omega}{N}$ (see (12)), we finally obtain

$$\frac{d}{dt}\|r(t,\cdot)\|_{L^\infty(\Omega)} \leqslant \|\partial_t r(t,\cdot)\|_{L^\infty(\Omega)} \leqslant L(t)\left(2\|r(t,\cdot)\|_{L^\infty(\Omega)} + \frac{C_\Omega^\alpha}{N^\alpha}(2+e^{tL(t)})\right)$$

and by integration, noting that $s\mapsto L(s)$ is nondecreasing, (56) follows.

Finally, (55) is established similarly as in the proof of Theorem 4.

References

[1] G. Albi, L. Pareschi, *Modeling of self-organized systems interacting with a few indi- viduals: from microscopic to macroscopic dynamics*, Applied Mathematics Letters **401** (2013), no. 26, 397–401.

[2] V. Arnol'd, *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits*, Ann. Inst. Fourier (Grenoble) **16** (1966), fasc. 1, 319–361.

[3] A. Aydogdu, M. Caponigro, S. McQuade, B. Piccoli, N. Pouradier Duteil, F. Rossi, E. Trélat, *Interaction network, state space, and control in social dynamics*, Active particles. Vol. 1. Advances in theory, models, and applications, 99–140, Model. Simul. Sci. Eng. Technol., Birkhäuser/Springer, Cham, 2017.

[4] N. Ayi, N. Pouradier Duteil, *Mean-field and graph limits for collective dynamics models with time-varying weights*, J. Differential Equations **299** (2021), 65–110.

[5] M. Bauer, M. Bruveris, P. Michor, *Overview of the geometries of shape spaces and diffeomorphism groups*, J. Math. Imaging Vision **50** (2014), no. 1-2, 60–97.

[6] N. Bellomo, A. Bellouquid, J. Nieto, J. Soler, *On the multiscale modeling of vehicular traffic: from kinetic to hydrodynamics*, Discrete Contin. Dyn. Syst. Ser. B **19** (2014), no. 7, 1869–1888.

[7] U. Biccari, D. Ko, E. Zuazua, *Dynamics and control for multi-agent networked systems: a finite-difference approach*, Math. Models Methods Appl. Sci. **29** (2019), no. 4, 755–790.

[8] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., Wiley, 1999.

[9] T. Bodineau, I. Gallagher, L. Saint-Raymond, *The Brownian motion as the limit of a deterministic system of hard-spheres*, Invent. Math. **203** (2016), no. 2, 493–553.

[10] V.I. Bogachev, *Measure theory*, Vol. I, II. Springer-Verlag, Berlin, 2007. Vol. I: xviii+500 pp., Vol. II: xiv+575 pp.

[11] B. Bonnet, N. Pouradier Duteil, M. Sigalotti, *Consensus formation in first-order graphon models with time-varying topologies*, Preprint arXiv:2111.03900v1 (2021).

[12] L. Boudin, F. Salvarani, E. Trélat, *Exponential convergence towards consensus for non-symmetric linear first-order systems in finite and infinite dimensions*, SIAM J. Math. Anal. **54** (2022), no. 3, 2727–2752.

[13] J.A. Carrillo, Y.-P. Choi, *Mean-field limits: from particle descriptions to macroscopic equations*, Arch. Ration. Mech. Anal. **241** (2021), no. 3, 1529–1573.

[14] J.A. Carrillo, Y.-P. Choi, M. Hauray, *The derivation of swarming models: mean-field limit and Wasserstein distances*, Collective dynamics from bacteria to crowds, 1–46, CISM Courses and Lect., 553, Springer, Vienna, 2014.

[15] J.A. Carrillo, M. Fornasier, G. Toscani, F. Vecil, *Mathematical modeling of collective behavior in socio-economic and life sciences*, Particle, kinetic, and hydrodynamic models of swarming, 297–336. Birkhäuser Boston, Boston, 2010.

[16] F. Cucker, S. Smale, *Emergent behavior in flocks*, IEEE Trans. Automat. Control **52** (2007), no. 5, 852–862.

[17] M. Di Francesco, M.D. Rosini, *Rigorous derivation of nonlinear scalar conservation laws from follow-the-leader type models via many particle limit*, Arch. Ration. Mech. Anal. **217** (2015), no. 3, 831–871.

[18] M. Di Francesco, S. Fagioli, M.D. Rosini, *Deterministic particle approximation of scalar conservation laws*, Boll. Unione Mat. Ital. **10** (2017), no. 3, 487–501.

[19] M. Di Francesco, S. Fagioli, E. Radici, *Deterministic particle approximation for nonlocal transport equations with nonlinear mobility*, J. Differential Equations **266** (2019), no. 5, 2830–2868.

[20] K.-J. Engel, R. Nagel, *One-parameter semigroups for linear evolution equations*, Graduate Texts Math. **194**, Springer-Verlag, 2000.

[21] A. Esposito, F.S. Patacchini, A. Schlichting, D. Slepcev, *Nonlocal-interaction equation on graphs: gradient flow structure and continuum limit*, Arch. Ration. Mech. Anal. **240** (2021), no. 2, 699–760.

[22] A. Figalli, M.-J. Kang, *A rigorous derivation from the kinetic Cucker-Smale model to the pressureless Euler system with nonlocal alignment*, Anal. PDE. **12** (2019), 843–866.

[23] N. Fournier, A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probab. Theory Related Fields **162** (2015), no. 3-4, 707–738.

[24] D. H. Fremlin, *Measure theory, Vol. 2, Broad foundations*, Corrected second printing of the 2001 original. Torres Fremlin, Colchester, 2003. 563+12 pp.

[25] I. Gallagher, *From Newton to Navier-Stokes, or how to connect fluid mechanics equations from microscopic to macroscopic scales*, Bull. Amer. Math. Soc. (N.S.) **56** (2019), no. 1, 65–85.

[26] F. Golse, C. Mouhot, T. Paul, *On the mean-field and classical limits of quantum mechanics*, Comm. Math. Physics **343** (2016), 165–205.

[27] S.-Y. Ha, E. Tadmor, *From particle to kinetic and hydrodynamic description of flocking*, Kinet. Relat. Models **1** (2008), 415–435.

[28] D.R. Hartree, *The wave mechanics of an atom with a non-Coulomb central field. Part II. Some results and discussions*, Proc. Cambridge Phil. Soc. **24** (1928), no. 1, 111–132.

[29] R. Hegselmann, U. Krause, *Opinion dynamics and bounded confidence: models, analysis and simulation*, J. Artif. Soc. Soc. Sim. **5** (2002), no. 3.

[30] K. Ito, F. Kappel, *The Trotter-Kato theorem and approximation of PDEs*, Math. Comp. **67** (1998), no. 221, 21–44.

[31] P.-E. Jabin, *A review of the mean field limits for Vlasov equations*, Kinet. Relat. Models **7** (2014), no. 4, 661–711.

[32] R.J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J. **80** (1995), no. 2, 309–323.

[33] G.S. Medvedev, *The nonlinear heat equation on dense graphs and graph limits* SIAM J. Math. Anal. **46** (2014), no. 4, 2743–2766.

[34] R. Natalini, T. Paul, *On the mean field limit for Cucker-Smale models*, Discr. Contin. Dynam. Syst. Ser. B **27** (2022), no. 27, 2873–2889.

[35] R. Natalini, T. Paul, *The mean-field limit for hybrid models of collective motions with chemotaxis*, preprint hal-03310547 (2021).

[36] A. Pazy, *Semigroups of linear operators and applications to partial differential equations*, Applied Mathematical Sciences, 44, Springer-Verlag, New York, 1983. viii+279 pp.

[37] B. Piccoli, F. Rossi, *Transport equation with nonlocal velocity in Wasserstein spaces: convergence of numerical schemes*, Acta Appl. Math. **124** (2013), 73–105.

[38] B. Piccoli, F. Rossi, *Generalized Wasserstein distance and its application to transport equations with source*, Arch. Ration. Mech. Anal. **211** (2014), no. 1, 335–358.

[39] B. Piccoli, F. Rossi, *On properties of the generalized Wasserstein distance* Arch. Ration. Mech. Anal. **222** (2016), no. 3, 1339–1365.

[40] B. Piccoli, F. Rossi, E. Trélat, *Control to flocking of the kinetic Cucker-Smale model*, SIAM J. Math. Anal. **47** (2015), no. 6, 4685–4719.

[41] F. Santambrogio, *Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling*, Progress in Nonlinear Differential Equations and their Applications, 87. Birkhäuser/Springer, Cham, 2015. xxvii+353 pp.

[42] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Mathematical Series, No. 30, Princeton University Press, Princeton, N.J. 1970 xiv+290 pp.

[43] L. Tartar, *An introduction to Sobolev spaces and interpolation spaces*, Lecture Notes of the Unione Matematica Italiana, 3. Springer, Berlin; UMI, Bologna, 2007. xxvi+218 pp.

[44] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003. xvi+370 pp.

[45] C. Villani, *Optimal transport. Old and new*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 338. Springer-Verlag, Berlin, 2009. xxii+973 pp.

[46] A.A. Vlasov, *On the kinetic theory of an assembly of particles with collective interaction*, Acad. Sci. USSR. J. Phys. **9** (1945), 25–40.