



**HAL**  
open science

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

Thierry Paul, Emmanuel Trélat

► **To cite this version:**

Thierry Paul, Emmanuel Trélat. From microscopic to macroscopic scale dynamics: mean field, hydrodynamic and graph limits. 2024. hal-04577170

**HAL Id: hal-04577170**

**<https://hal.science/hal-04577170>**

Preprint submitted on 16 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Thierry Paul\*      Emmanuel Trélat†

## Abstract

Considering deterministic finite particle systems, we elaborate on various ways to pass to the limit as the number of agents tends to infinity, 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.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Setting	3
1.2	Microscopic viewpoint: family of particle systems	5
1.3	Examples	7
1.4	Objectives, and structure of the article	9
1.5	General notations	10
<b>2</b>	<b>From microscopic to macroscopic scale: the continuum / graph limit</b>	<b>13</b>
2.1	Continuum / graph limit equation	13
2.2	Convergence estimates for the continuum / graph limit	15
2.3	Examples	17
<b>3</b>	<b>From microscopic to mesoscopic scale I: mean field through empirical measures (from ODEs to Vlasov)</b>	<b>18</b>
3.1	Vlasov equation	19
3.2	Relationship between the particle system and the Vlasov equation	22
3.3	Examples	24
<b>4</b>	<b>From microscopic to mesoscopic scale II: mean field by lifting the particle system (from Liouville to Vlasov)</b>	<b>25</b>
4.1	Liouville equation	25
4.2	Deriving Vlasov from Liouville by taking marginals, propagation of chaos	26
4.2.1	First way, with $\rho_0^N$ Dirac	28
4.2.2	Second way, with $\rho_0^N$ “semi-Dirac”	30

---

\*CNRS Laboratoire Ypatia des Sciences Mathématiques LYSM, Rome, Italy (thierry.paul@sorbonne-universite.fr).

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

<b>5</b>	<b>From mesoscopic to macroscopic scale: hydrodynamic limit (from Vlasov to Euler)</b>	<b>32</b>
5.1	Averaged dynamical quantities defined on $\Omega$	32
5.2	Moment of order 1, Euler equation	33
5.2.1	Linear Euler equation	33
5.2.2	Open issue: how to obtain a closed equation?	34
5.2.3	The $\nu$ -monokinetic case	34
5.3	Moment of order 2	35
5.4	Generalization: coupled equations of moments	36
5.5	Directly from Liouville to Euler	37
5.6	Hydrodynamic limit in the second-order case	39
<b>6</b>	<b>Summary: relationships between various scales (Euler and continuum / graph limit equations are the same)</b>	<b>41</b>
<b>7</b>	<b>Further comments and perspectives</b>	<b>43</b>
7.1	Approximation of PDEs by finite particle systems	43
7.2	Some open questions	45
<b>A</b>	<b>Appendix</b>	<b>48</b>
A.1	Some general facts on the Wasserstein distance	48
A.1.1	Convexity	49
A.1.2	Symmetrization	50
A.1.3	Marginals	50
A.1.4	Tensor product	51
A.1.5	Diameter of the support	52
A.1.6	Propagation	52
A.1.7	Moment of order one	56
A.2	More precise facts on the marginals of a symmetrization	56
A.2.1	First marginal of the symmetrization	56
A.2.2	Marginals of symmetric measures	57
A.2.3	A combinatorial lemma towards propagation of chaos	58
A.3	Density of empirical measures in the set of probability measures	60
A.4	Convergence of empirical and semi-empirical measures	61
A.4.1	Convergence of empirical measures on $\Omega$	62
A.4.2	Convergence of semi-empirical measures	63
A.5	Discrepancy between empirical and $\nu$ -monokinetic measures	64
A.6	Mean field and variance	65
<b>B</b>	<b>Proofs</b>	<b>68</b>
B.1	Proof of Theorem 2.2	68
B.2	Proof of Theorem 2.3	70
B.3	Proof of Theorem 3.1	71
B.4	Proof of Theorem 4.1	75
B.5	Proof of Theorem 4.2	76

# 1 Introduction

## 1.1 Setting

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

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

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  $i^{\text{th}}$  agent and  $G_{ij}^N : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a mapping modeling the interaction between the  $i^{\text{th}}$  and  $j^{\text{th}}$  agents. Note that, here and throughout the paper, we focus on deterministic systems.

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, in quantum systems, in fluid mechanics (see, e.g., [2, 4, 7, 17, 18, 21, 24, 35, 38, 42, 43, 56, 57, 75, 76], just to mention a few of a vast literature).

Among classes of multi-agent systems, we point out the Hegselmann–Krause (opinion propagation) systems (see [42]) 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  $G_{ij}^N(t, \xi_i, \xi_j) = \sigma_{ij}(\xi_j - \xi_i)$ , i.e., their dynamics is

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

where  $(\sigma_{ij})_{1 \leq i, j \leq N}$  is a  $N$ -by- $N$  matrix whose spectral properties may cause the above-mentioned asymptotic behavior.

The large  $N$  limit of systems (1), (2) has been extensively studied over the last years. In [55] 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 [28] to obtain discrete-to-continuum convergence results with error estimates in the Wasserstein distance. We also mention the recent articles [5, 8, 13, 44]. 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, which we call continuum / graph limit 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 [5, 8, 14] 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  $\sigma(x_i, x_j) = \sigma_{ij}$  for some points  $x_j \in \Omega$ .

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  $G$  may take the form

$$G(\xi_i, \xi_j) = \left( \nabla V \left( \frac{p_i}{q_i - q_j} \right) \right) \quad (3)$$

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

$$G(\xi_i, \xi_j) = \left( a(\|q_i - q_j\|) \frac{p_i}{\|p_j - p_i\|} \right) \quad (4)$$

for some influence function  $a$ .

The difference between general systems (1) and 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 “labels” of the agents. A consequence is that, in the usual literature, the associated evolution equation preserves the *indistinguishability* of the particles, a feature often considered 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 [40]) in the late 20’s for quantum systems, and then Vlasov in the 40’s (see [79]) who derived the eponymous *kinetic equation* (here, we present it in a form suitable for our purposes), called *Vlasov equation*,

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

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  $\mathbb{R}^{dN}$  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  $N \rightarrow +\infty$  the Vlasov equation (5). The last step of this process, called the *hydrodynamic limit*, starts from the observation that (5) preserves the structure  $\mu_t(\xi) = \mu_t(q, p) = \nu(t, q) \delta(p - y(t, q))$  leading to the so-called *Euler system* of equations satisfied by the pair  $(\nu, y)$ .

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 continuum / graph limit equation associated to (1) – a nontrivial fact, even at the conceptual level as discussed in Section 5.5 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 finite 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 equations, showing how to pass from one to another and deriving, under appropriate assumptions, some general 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.

Let us point out a difference of methodology between the present paper and several previous works concerning the case of indistinguishable particles: we do not estimate propagation of couplings but we rather use direct estimates of the particles and Vlasov flows thanks to the use of a convenient Wasserstein distance. This allows us to obtain, as a by-product of our main results, quantitative mean field limits for more general vector fields than in (3) and (4).

## 1.2 Microscopic viewpoint: family of particle systems

Let  $d \in \mathbb{N}^*$  be fixed. Throughout the paper, we consider an arbitrary norm  $\|\cdot\|$  on  $\mathbb{R}^d$ . At the microscopic level, given any  $N \in \mathbb{N}^*$ , we consider a system of  $N$  interacting “particles” or “agents”  $\xi_i^N(t) \in \mathbb{R}^d$ , called the *particle system* (or *multiagent system*), of dynamics

$$\dot{\xi}_i^N(t) = \frac{1}{N} \sum_{j=1}^N G_{ij}^N(t, \xi_i^N(t), \xi_j^N(t)), \quad i \in \{1, \dots, N\} \quad (6)$$

where  $G_{ij}^N : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  stands for the interaction between the particles  $i$  and  $j$ . The dot stands for the time derivative. The most usual case, widely treated in the existing literature, is when  $G_{ij}^N = G$ : in this case, the particles are indistinguishable (or, exchangeable in the probabilistic language), reflecting the fact that the dynamics are invariant under permutations of the  $\xi_i^N$ . We show here that there is no difficulty to treat the more general situation where the particles are distinguishable and the interactions depend on the agents. In (6),  $G_{ij}^N$  depends on  $i, j, N$ .

Throughout the paper, we make the following crucial assumption:

(G) There exist a complete metric space  $(\Omega, d_\Omega)$  and a *continuous* mapping

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

locally Lipschitz with respect to  $(\xi, \xi')$  uniformly with respect to  $(t, x, x')$  on any compact subset of  $\mathbb{R} \times \Omega \times \Omega$ , such that, for every  $N \in \mathbb{N}^*$ , there exist  $x_1^N, \dots, x_N^N$  in  $\Omega$  such that

$$G(t, x_i^N, x_j^N, \xi, \xi') = G_{ij}^N(t, \xi, \xi') \quad \forall t \in \mathbb{R} \quad \forall \xi, \xi' \in \mathbb{R}^d \quad \forall i, j \in \{1, \dots, N\}. \quad (7)$$

Under Assumption (G), for every  $N \in \mathbb{N}^*$  the particle system (6) is equivalently written as

$$\begin{aligned} \dot{x}_i^N(t) &= 0 \\ \dot{\xi}_i^N(t) &= \frac{1}{N} \sum_{j=1}^N G(t, x_i^N, x_j^N, \xi_i^N(t), \xi_j^N(t)), \quad i \in \{1, \dots, N\} \end{aligned} \quad (8)$$

The variables  $x_i^N \in \Omega$  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  $\dot{x}_i^N(t) = 0$  above. For each index  $i$ , the variable  $x_i^N$  can be seen as the “label” (type, name, color) of the agent  $i$ , used to distinguish it from the others.

Assumption (G) (in particular, (7)) is a kind a continuous interpolation of the mappings  $G_{ij}^N$ . The continuity assumption includes the idea of the existence of a limit system as  $N \rightarrow +\infty$ . In some sense, this assumption is unavoidable: indeed, if  $G$  were not required to be continuous, then completely different systems (6) could be considered as  $N$  varies and then no limit (at least, in a strong sense) for large  $N$  could exist. Note anyway that, interestingly, the authors of [44] do not assume (G), but in order to pass to the mean field limit they make another assumption of uniform boundedness on their dynamics in order to have a weak star limit. However at the limit the distinguishability of particles is lost. In contrast, in our paper we want to obtain strong (mean field, hydrodynamic, graph) limits and to preserve distinguishability at the limit.

Note that Assumption (G) implies that the Lipschitz constants of the mappings  $G_{ij}^N$  are uniformly bounded (with respect to  $i, j, N$ ) on any compact.

In Assumption **(G)**, the complete metric space  $\Omega$  used for the parameters  $x_i^N$  is arbitrary. For instance we can take  $\Omega = [0, 1]$ , but we allow for more general sets, in view of deriving on  $\Omega$  some interesting classes of PDEs (see Section 7.1).

The choice of the possible values of the  $x_i^N$  is not imposed in Assumption **(G)**. If one wishes moreover to fix some precise points  $x_i^N$ , such as the natural ones  $x_i^N = \frac{i}{N}$  when  $\Omega = [0, 1]$ , often used in numerical analysis, then having (7) satisfied requires some compatibility conditions on the mappings  $G_{ij}^N$ .

In the above framework, the classical case studied in the existing literature, where particles are indistinguishable, is when the mapping  $G$  does not depend on  $(x, x')$ .

Setting  $X^N = (x_1^N, \dots, x_N^N) \in \Omega^N$ , the system (8) can also be written in the form

$$\boxed{\dot{\Xi}^N(t) = Y^N(t, X^N, \Xi^N(t))} \quad (9)$$

where  $\Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t))$ . Here and in what follows, the time-dependent vector field  $Y^N(t, X, \cdot)$  on  $(\mathbb{R}^d)^N$ , depending on the parameter  $X \in \Omega^N$ , is defined by

$$Y^N(t, X, \cdot) = (Y_1^N(t, X, \cdot), \dots, Y_N^N(t, X, \cdot)) \quad (10)$$

with

$$Y_i^N(t, X, \Xi) = \frac{1}{N} \sum_{j=1}^N G(t, x_i, x_j, \xi_i, \xi_j) \quad \forall i \in \{1, \dots, N\} \quad (11)$$

for all  $t \in \mathbb{R}$ ,  $X = (x_1, \dots, x_N) \in \Omega^N$  and  $\Xi = (\xi_1, \dots, \xi_N) \in (\mathbb{R}^d)^N$ . We denote by  $(\Phi^N(t, X, \cdot))_{t \in I}$  ( $I \subset \mathbb{R}$ ) the local-in-time flow of diffeomorphisms of  $\mathbb{R}^{dN}$  generated by the time-dependent vector field  $Y^N(t, X, \cdot)$ : this flow, called the *particle flow*, is parametrized by  $X \in \Omega^N$ . We have  $\Xi^N(t) = \Phi^N(t, X^N, \Xi^N(0))$  for every  $t \in I$ .

**Lemma 1.1.** *[Uniform maximal time] For any compact subset  $K$  of  $\Omega \times \mathbb{R}^d$ , there exists  $T_{\max}(K) \in (0, +\infty]$  such that, for any  $N \in \mathbb{N}^*$ , for any  $(X, \Xi(0)) \in K^N$ ,<sup>1</sup> there exists a unique solution  $t \mapsto \Xi^N(t) = \Phi^N(t, X^N, \Xi^N(0))$  of (9) on  $[0, T_{\max}(K))$ , of parameter  $X^N$  and of initial condition  $\Xi^N(0)$  at  $t = 0$ , and of class  $\mathcal{C}^1$  with respect to  $t$ . Moreover, for any  $T \in [0, T_{\max}(K))$ , the set  $\Phi^N([0, T] \times K^N)$  is contained in a compact subset of  $\mathbb{R}^d$  depending on  $T$  but not on  $N$ .*

Lemma 1.1 shows that, given a compact set  $K$  of initial conditions, the time  $T_{\max}(K)$  is uniform with respect to  $N \in \mathbb{N}^*$ , and that, given any  $T \in (0, T_{\max}(K))$ , any solution of (9) on  $[0, T]$ , starting in  $K$  at  $t = 0$ , is contained in a compact set that depends on  $T$  but not on  $N$ .

Lemma 1.1 straightforwardly follows from the usual proof of the Picard-Lindelöf (Cauchy-Lipschitz) theorem by a fixed point argument (see [39, Chapter II]), using Assumption **(G)**, noting that, for every  $T > 0$ , on  $[0, T] \times K^N$  the vector field  $Y^N$  is uniformly bounded with respect to  $N$  and is Lipschitz with respect to  $\Xi$  uniformly with respect to  $(t, X)$  on any compact, with a Lipschitz constant that is uniform with respect to  $N$ . Note that, for a given  $N \in \mathbb{N}^*$ , the maximal time of definition of the solution  $t \mapsto \Phi^N(t, X, \Xi(0))$  may be larger than  $T_{\max}(K)$ ; what is important in the lemma is the uniform bound below with respect to  $N$ .

Of course, if  $G$  is globally Lipschitz with respect to  $(\xi, \xi') \in \mathbb{R}^d \times \mathbb{R}^d$ , uniformly with respect to  $(t, x, x')$  on any compact subset of  $[0, +\infty) \times \Omega \times \Omega$ , then  $T_{\max}(K) = +\infty$  for any compact  $K \subset \Omega \times \mathbb{R}^d$ . But our framework is more general and allows for superlinearities.

We next give some examples covered by this general framework.

<sup>1</sup>With a slight abuse of notation,  $(X^N, \Xi^N(0)) \in K^N$  means that  $(x_i^N, \xi_i^N(0)) \in K$  for every  $i \in \{1, \dots, N\}$ .

### 1.3 Examples

**First-order systems.** General first-order systems of the form

$$\dot{\xi}_i^N(t) = F_i^N(t, \xi_i^N(t)) + \frac{1}{N} \sum_{j=1}^N K_{ij}^N(t, \xi_i^N(t), \xi_j^N(t)), \quad i \in \{1, \dots, N\}, \quad (12)$$

can be written as (6) with  $G_{ij}^N(t, \xi, \xi') = F_i(t, \xi) + K_{ij}^N(t, \xi, \xi')$ . Assumption **(G)** is satisfied if there exist a set  $\Omega$  and sufficiently regular mappings  $F$  and  $K$  such that  $F(t, x_i^N, x_j^N, \xi, \xi') = F_i^N(t, \xi)$  and  $K(t, x_i^N, x_j^N, \xi, \xi') = K_{ij}^N(t, \xi, \xi')$  as in (7).

– A first meaningful example is the linear Hegselmann–Krause first-order consensus system (see [42]), modeling for instance the propagation of opinions (studied in [14]), of dynamics

$$\dot{\xi}_i^N(t) = \frac{1}{N} \sum_{j=1}^N \sigma_{ij}^N (\xi_j^N(t) - \xi_i^N(t)), \quad i \in \{1, \dots, N\}, \quad (13)$$

with constant interaction coefficients  $\sigma_{ij}^N \geq 0$  (not necessarily symmetric). Assumption **(G)** requires that there exist a set  $\Omega$  (for example, but not necessarily,  $\Omega = [0, 1]$ ) and a continuous function  $\sigma$  on  $\Omega^2$  such that, for every  $N \in \mathbb{N}^*$ , there exist distinct points  $x_1^N, \dots, x_N^N$  in  $\Omega$  such that  $\sigma(x_i^N, x_j^N) = \sigma_{ij}^N$ . The graph interpretation, which is particularly relevant here, will be commented in Section 2. We have then  $G(t, x, x', \xi, \xi') = \sigma(x, x')(\xi' - \xi)$  for all  $(t, x, x', \xi, \xi') \in \mathbb{R} \times \Omega^2 \times (\mathbb{R}^d)^2$ .

More general models can be considered, with interaction coefficients  $\sigma_{ij}$  depending on  $t$  and on the  $\xi_i$  (see the survey [57] and see the recent Transformers particle model studied in [34]).

– A second interesting example is the Kuramoto model

$$\dot{\xi}_i^N(t) = \alpha_i^N + \frac{1}{N} \sum_{j=1}^N \sigma_{ij}^N \sin(\xi_j^N(t) - \xi_i^N(t)), \quad i \in \{1, \dots, N\}, \quad (14)$$

where  $d = 1$ ,  $\xi_i^N(t) \in \mathbb{R}$  is the phase of the oscillator  $i$ ,  $\alpha_i^N \in \mathbb{R}$  is its frequency and  $\sigma_{ij}^N \in \mathbb{R}$  is an interaction coefficient between oscillators  $i$  and  $j$ . This system was introduced in [49] in view of studying synchronization of interacting oscillators. To write the particle system (14) in the form (8), now two parameters (labels) are required for each particle, one standing for the frequency and the other for the interaction as in the previous example. We set  $\Omega = \mathbb{R} \times [0, 1]$  and for every  $x \in \Omega$  we denote by  $x = (\alpha, \beta) \in \Omega$  the two coordinates of  $x$ . Assumption **(G)** is satisfied if there exists a continuous function  $\sigma$  on  $[0, 1]^2$  satisfying  $\sigma(\beta_i^N, \beta_j^N) = \sigma_{ij}^N$  as in (7), and we have then  $G(t, x, x', \xi, \xi') = \alpha + \sigma(\beta, \beta')(\xi' - \xi)$  (where  $x = (\alpha, \beta)$  and  $x' = (\alpha', \beta')$ ).

– Consider again the general system (12), but where now  $F_i^N = F$  and  $K_{ij}^N = K$  do not depend on  $i, j, N$ , with  $F, K \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ . In this case (12) becomes

$$\dot{\xi}_i^N(t) = F(\xi_i^N(t)) + \frac{1}{N} \sum_{j=1}^N K(\xi_i^N(t) - \xi_j^N(t)), \quad i \in \{1, \dots, N\}, \quad (15)$$

which is a much used particle model (see [43]). Assumption **(G)** is satisfied and  $G(t, x, x', \xi, \xi') = F(\xi) + K(\xi - \xi')$ , not depending on  $(t, x, x')$ : this is an indistinguishable case. Often,  $K = -\nabla V$  where  $V$  is an interaction potential (that we consider here to be regular).



**Second-order systems.** Setting  $d = 2r$  and denoting  $\xi = (q, p) \in \mathbb{R}^r \times \mathbb{R}^r$ , general second-order systems of the form

$$\dot{q}_i^N(t) = p_i^N(t), \quad \dot{p}_i^N(t) = \frac{1}{N} \sum_{j=1}^N b_{ij}^N(t, q_i^N(t), p_i^N(t), q_j^N(t), p_j^N(t)), \quad i \in \{1, \dots, N\}, \quad (16)$$

can be written as (6) with  $G_{ij}^N(t, \xi, \xi') = (p, b_{ij}^N(t, \xi, \xi'))$ . Assumption (G) is satisfied if there exist a set  $\Omega$  and a sufficiently regular mapping  $b$  interpolating all mapping  $b_{ij}^N$  as stated in (7), i.e.,  $b(t, x_i^N, x_j^N, \xi, \xi') = b_{ij}^N(t, \xi, \xi')$ .

Here,  $q$  is a position and  $p$  is a momentum. It is important to note that the variable  $q$  should not be confused with the variable  $x \in \Omega$  that is used here to designate the label of a particle.

– A famous example of second-order dynamics is the Cucker–Smale model (see [24])

$$\dot{q}_i^N(t) = p_i^N(t), \quad \dot{p}_i^N(t) = \frac{1}{N} \sum_{j=1}^N a(\|q_j^N(t) - q_i^N(t)\|)(p_j^N(t) - p_i^N(t)), \quad i \in \{1, \dots, N\}, \quad (17)$$

where  $a \in \mathcal{C}^1(\mathbb{R})$ . Assumption (G) is satisfied with  $G = (G_q, G_p)$  where  $G_q(t, x, x', \xi, \xi') = p$  and  $G_p(t, x, x', \xi, \xi') = a(\|q' - q\|)(p' - p)$ , not depending on  $(t, x, x')$ : this is an indistinguishable case.

Many variants of that model are covered by our framework, for instance the potential  $a$  may depend on  $i$  and  $j$ , and other terms can be added to the dynamics of  $p_i$ , for instance self-propulsion and attraction-repulsion forces (like in [17]); in this case, defining a set  $\Omega$  is required.

– Many second-order particle systems studied in the literature, coming from the Newton law for  $N$  particles, are of the form (16) with  $b_{ij}^N(t, \xi, \xi') = K(\xi, \xi')$ , not depending on  $i, j, N$ , with  $K \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ , yielding

$$\dot{q}_i^N(t) = p_i^N(t), \quad \dot{p}_i^N(t) = \frac{1}{N} \sum_{j=1}^N K(q_i^N(t), q_j^N(t)), \quad i \in \{1, \dots, N\}. \quad (18)$$

Assumption (G) is satisfied with  $G(t, x, x', \xi, \xi') = (p, K(q, q'))$ , not depending on  $(t, x, x')$ : this is an indistinguishable case. Note that, when  $K = -\nabla V$  for some potential function, the above particle system stands for the classical  $N$ -body problem in Hamiltonian form (see next for more general Hamiltonian cases), with the Hamiltonian function given by  $H(q_1, p_1, \dots, q_N, p_N) = \frac{1}{2} \sum_{j=1}^N \|p_j\|^2 + \frac{1}{N} \sum_{j=1}^N V(q_i, q_j)$ .

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

$$H^N(q_1, p_1, \dots, q_N, p_N) = \sum_{j=1}^N h_j^N(q_j, p_j) + \frac{1}{N} \sum_{j,k=1}^N h_{jk}^N(q_j, p_j, q_k, p_k) \quad (19)$$

for some  $\mathcal{C}^1$  functions  $h_j^N$  and  $h_{jk}^N$ . The 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  $i \in \{1, \dots, N\}$ , can be written as (6) with

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

where  $\partial_k$  denotes the partial derivative with respect to the  $k^{\text{th}}$ -variable.

Having Assumption **(G)** satisfied requires at least that the Hamiltonians  $h_j^N$  and  $h_{jk}^N$  be uniformly (wrt  $j, k, N$ ) locally Lipschitz. Note that the Hamiltonian  $H^N$  defined by (19) involves sums of “single” (noninteracting) and of “pairwise” Hamiltonians, but not of “triplewise” or more.

Many classical Hamiltonian systems of  $N$  particles are written as above with Hamiltonians not depending on  $j, k, N$ , for instance in quantum mechanics (see [35]) or in differential geometry. An example, where Assumption **(G)** is satisfied, used to model systems of fermions confined in a magnetic field, is when  $h_j^N(q_j, p_j) = V(q_j) + \frac{1}{2}\|p_j - A(q_j)\|^2$  for some confining potential  $V \in \mathcal{C}^1(\mathbb{R}^d)$  and some magnetic potential vector  $A \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ , and  $h_{jk}(q_j, p_j, q_k, p_k) = W(\|q_j - q_k\|)$  for some pairwise interaction potential  $W \in \mathcal{C}^1(\mathbb{R}^d)$ . In this case, we have

$$G(t, x, x', \xi, \xi') = \left( p - A(q), -\nabla V(q) + dA(q) \cdot (p - A(q)) - \partial_1 W(q, q') - \partial_2 W(q', q) \right). \quad (20)$$

**Remark 1.1** (On the wording “indistinguishability”). In the literature, a dynamical system  $\dot{z}(t) = X(t, z(t))$  in  $\mathbb{R}^n$  is said to be “indistinguishable”, or “exchangeable” in the probabilistic wording, if it is invariant under permutations in the following sense: given any  $z_0 \in \mathbb{R}^n$ , denoting by  $t \mapsto z(t, z_0)$  the unique solution on some interval  $I$  of the system such that  $z(0, z_0) = z_0$ , we have  $z(t, \sigma(z_0)) = \sigma(z(t, z_0))$  for every  $t \in I$ , for every permutation  $\sigma \in \mathfrak{S}_n$ . Equivalently, the vector field  $X$  is invariant under the action of the permutation, i.e.,  $\sigma_* X(t, \cdot) = X(t, \cdot)$  for every  $t$ .

(1) General particle systems in  $\mathbb{R}^{dN}$  of the form (6) are not indistinguishable in general because the interaction mapping  $G_{ij}^N$  depends on  $i$  and  $j$  (but they are indistinguishable if  $G_{ij}^N = G$ ): the dynamics are not invariant under permutations  $\sigma \in \mathfrak{S}_{dN}$  acting on  $\Xi = (\xi_1, \dots, \xi_N)$ . This is the standard wording used in the literature to describe the distinguishability or indistinguishability of systems of particles, and we will follow this wording throughout the article.

(2) In Section 1.2 we have introduced a set of labels  $x \in \Omega$ , distinguishing particles, and we have done the fundamental assumption **(G)**. In this context, the particle system (6) (which is, in general, distinguishable) has been rewritten as (8) or equivalently as (9), by augmenting the state space to  $\Omega^N \times \mathbb{R}^{dN}$ . But then, in this augmented form, the system (9) is *always* indistinguishable in the sense that it is invariant under permutations  $\sigma \in \mathfrak{S}_{dN}$  acting simultaneously on  $X = (x_1, \dots, x_N)$  and on  $\Xi = (\xi_1, \dots, \xi_N)$ . Hence, in some way, we recover indistinguishability in the new state space  $\Omega^N \times \mathbb{R}^{dN}$ .

Despite the slight ambiguity, throughout the paper, we will continue to use the wording described in (1).

## 1.4 Objectives, and structure of the article

When  $N$  becomes larger and larger, we want to pass to the limit in some sense.

A first possibility is to pass to the *continuum limit* in (8) by interpreting the right-hand side of (8) as a Riemann sum, and use the Riemann integral. This procedure is sometimes called *graph limit* and the result is an integro-differential equation that we call the continuum / graph limit equation. This is done in Section 2, where we show how to pass from the microscopic to the macroscopic scale, i.e., to pass “from particle to continuum / graph limit” by graph limit. Theorems 2.2 and 2.3 quantify some convergence estimates in  $L^\infty$  norm as  $N \rightarrow +\infty$ .

Another possibility is to replace the set of particles with a nonnegative Radon measure. Two classical viewpoints are the Lagrangian and the Eulerian one. The *Lagrangian* viewpoint consists of keeping the trajectories of (8), 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 3. The *Eulerian* viewpoint consists of using the flow of diffeomorphisms

of  $Y$  to propagate an initial measure in the big space  $\Omega^N \times \mathbb{R}^{dN}$ , thus obtaining the *Liouville equation* in  $\Omega^N \times \mathbb{R}^{dN}$ . This is done in Section 4.

More precisely, Section 3 focuses on the passage from microscopic to mesoscopic scale: we show how to pass “from particle to Vlasov” by mean field limit. In Theorem 3.1, we establish existence, uniqueness and stability properties for the Vlasov equation for distinguishable particles, generalizing the famous Dobrushin estimates. We explain the relationship between the particle system and the Vlasov equation.

In Section 4, following the Eulerian viewpoint, we lift particles to probability measures on a big product space and we elaborate on the Liouville equation associated with the particle system. We show in Theorems 4.1 and 4.2 how to derive Vlasov from Liouville by taking marginals and passing to the limit (propagation of chaos), with estimates in Wasserstein distance. We also discuss how Euler can be derived from Liouville by taking adequate moments.

Section 5 is devoted to showing how to pass from the mesoscopic to the macroscopic scale: we show how to derive the Euler equation from the Vlasov equation, by the procedure usually called hydrodynamic limit, mainly consisting of taking the moment of order 1 of the solution of the Vlasov equation. Proposition 5.2 in that section is concerned with the well known monokinetic approach, but we also investigate the moment of order 2, yielding some consensus results.

In all sections, we also give corollaries in the indistinguishable case, thus recovering some well known results. But this article is not only a survey: we generalize many known results, sometimes to a wide extent, in a hopefully general and unifying context.

Note that, throughout this paper, the interaction mapping  $G$  is assumed to be continuous; in other words, we treat *regular* interaction mappings. This already covers a large number of examples, although many of the interaction mappings for applications are singular, like the Poisson kernel, point vortices, etc (see, e.g., [43]). For singular kernels, passing to the limit can be a real challenge (see [10, 25, 29, 33, 35, 41, 74]) and requires other mathematical techniques. One of them is to cut off the interaction mapping around its singularities, by introducing an additional cut-off parameter  $\varepsilon$  tending to 0 (a bit like in Section 7.1 where we give the idea of how to approach unbounded operators) but then one has to deal with a double limit,  $\varepsilon \rightarrow 0$  and  $N \rightarrow +\infty$ . This raises other difficulties and challenges, not addressed here. In this article, we only consider regular kernels, showing in a unified framework the various ways to pass to the limit and the relationships between them.

Note also that, in this paper, we consider deterministic finite systems of particles. Stochastic models and their various limits are outside of the scope of the present article.

Section 6 provides a synthetic summarize of all relationships that we have unraveled. In particular, Figure 1 illustrates the various two-ways passages between particle (microscopic) ODEs systems, the continuum / graph limit, Euler (macroscopic) equations, the Vlasov (mesoscopic, mean field) equation, and the Liouville (microscopic again) equation. This section can even be read as a motivating preliminary before going ahead.

In order to state all subsequent results, we recall in Section 1.5 hereafter some notations and concepts that we use throughout, in particular some background on Wasserstein distances, on disintegration of measures, and the concept of tagged partition that is classically used in general 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.

## 1.5 General notations

Let  $E$  be a Polish space, endowed with a distance  $d_E$ .

**Hölder and Lipschitz mappings.** Let  $U$  be a subset of  $E$ . Let  $k \in \mathbb{N}^*$  and let  $\|\cdot\|$  be a norm on  $\mathbb{R}^k$ . Given any  $\alpha \in (0, 1]$ , we denote by  $\mathcal{C}^{0,\alpha}(U, \mathbb{R}^k)$  the set of all continuous mappings  $g \in \mathcal{C}^0(U, \mathbb{R}^k)$  that are  $\alpha$ -Hölder continuous (with respect to the norm  $\|\cdot\|$ ), meaning that

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

When  $\alpha = 1$ , we speak of a Lipschitz mapping and we denote  $\text{Lip}(g) = \text{Hol}_1(g)$ . When  $U$  is compact,  $\mathcal{C}^{0,\alpha}(U, \mathbb{R}^k)$  is a Banach space endowed with the norm

$$\|g\|_{\mathcal{C}^{0,\alpha}(U, \mathbb{R}^k)} = \max_{y \in U} \|g(y)\| + \text{Hol}_\alpha(g).$$

When  $k = 1$  and  $\alpha = 1$ , we denote  $\text{Lip}(U) = \mathcal{C}^{0,1}(U, \mathbb{R})$ .

**Probability Radon measures.** We denote by  $\mathcal{P}(E)$  the set of nonnegative probability Radon measures on  $E$ . We also consider  $\mathcal{P}_c(E)$ ,  $\mathcal{P}^{ac}(E)$ , where the subscript  $c$  means “with compact support” and the superscript  $ac$  means “absolutely continuous with respect to a Lebesgue measure” (in the case where  $E$  is equipped with a Lebesgue measure), and for every  $p \geq 1$  the set  $\mathcal{P}_p(E)$  stands for the set of all  $\mu \in \mathcal{P}(E)$  that have a finite moment of order  $p$ , i.e.,  $\int_E d_E(y_0, y)^p d\mu(y) < +\infty$  where  $y_0 \in E$  is arbitrary. Given any Borel mapping  $\phi : E \rightarrow F$  where  $F$  is another Polish space and given any  $\mu \in \mathcal{P}(E)$ , the image (or pushforward) of  $\mu$  under  $\phi$  is  $\phi_*\mu = \mu \circ \phi^{-1}$ .

We denote by  $\mathcal{C}^0(E)$  the set of continuous functions on  $E$  and by  $\mathcal{C}_c^0(E)$  the set of continuous functions of compact support on  $E$ . When  $E$  is a smooth manifold, we adopt similar notations for the set  $\mathcal{C}^\infty(E)$  of smooth functions on  $E$ . We recall that the topological dual  $(\mathcal{C}_c^0(E))'$  (resp.,  $(\mathcal{C}^0(E))'$ ) is the set of all Radon measures on  $E$  (resp., with compact support). Endowed with the total variation norm  $\|\cdot\|_{TV}$  which is the dual norm, it is a Banach space.

Throughout the paper,  $\delta_\star$  is the Dirac measure at  $\star$ .

**Wasserstein distance.** Given any  $p \geq 1$ , the Wasserstein distance  $W_p(\mu_1, \mu_2)$  of order  $p$  between two probability measures  $\mu_1, \mu_2 \in \mathcal{P}(E)$ , with respect to the distance  $d_E$ , is defined as the infimum of the Monge-Kantorovich cost  $\int_{E^2} d_E(y_1, y_2)^p d\Pi(y_1, y_2)$  over the set of probability measures  $\Pi \in \mathcal{P}(E^2)$  coupling  $\mu_1$  with  $\mu_2$ , i.e., whose marginals on the two copies of  $E$  are  $\mu_1$  and  $\mu_2$ :

$$W_p(\mu_1, \mu_2) = \inf \left\{ \left( \int_{E^2} d_E(y_1, y_2)^p d\Pi(y_1, y_2) \right)^{1/p} \mid \Pi \in \mathcal{P}(E^2), (\pi_1)_*\Pi = \mu_1, (\pi_2)_*\Pi = \mu_2 \right\} \quad (21)$$

where  $\pi_1 : E^2 \rightarrow E$  and  $\pi_2 : E^2 \rightarrow E$  are the canonical projections defined by  $\pi_1(y_1, y_2) = y_1$  and  $\pi_2(y_1, y_2) = y_2$  for all  $(y_1, y_2) \in E \times E$ . Equivalently,

$$W_p(\mu_1, \mu_2) = \inf \left\{ \left( \mathbb{E} d_E(Y_1, Y_2)^p \right)^{1/p} \mid \text{law}(Y_1) = \mu_1, \text{law}(Y_2) = \mu_2 \right\} \quad (22)$$

where the infimum is taken over all possible random variables  $Y_1$  and  $Y_2$  (defined on a same probability space, with values in  $E$ ) having the laws  $\mu_1$  and  $\mu_2$  respectively. Then,  $W_p$  is a distance on  $\mathcal{P}_p(E)$ , which metrizes the weak convergence in  $\mathcal{P}_p(E)$  in the following sense: given  $\mu \in \mathcal{P}_p(E)$  and given a sequence  $(\mu_j)_{j \in \mathbb{N}^*}$  in  $\mathcal{P}_p(E)$ , we have  $W_p(\mu_j, \mu) \rightarrow 0$  as  $j \rightarrow +\infty$  if and only if  $\int_E f d\mu_j \rightarrow \int_E f d\mu$  for every continuous bounded function  $f$  on  $E$  and  $\int_E d_E(y_0, y)^p d\mu_j(y) \rightarrow \int_E d_E(y_0, y)^p d\mu(y)$  as  $j \rightarrow +\infty$  for some (and thus any)  $y_0 \in E$  (see [73, Chapter 5, Section 5.2] or [78, Theorem 6.9]), if and only if  $\int_E f d\mu_j \rightarrow \int_E f d\mu$  for every continuous function  $f$  on  $E$  such

that  $|f(y)| \leq C(1 + d_E(y_0, y)^p)$  for every  $y \in E$ , for some  $C > 0$  and some (and thus any)  $y_0 \in E$  (see [77, Theorem 7.12]). It can be noted that, given any subset  $K \subset E$  of finite diameter, we have

$$1 \leq p_1 \leq p_2 \Rightarrow W_{p_1}(\mu_1, \mu_2) \leq W_{p_2}(\mu_1, \mu_2) \leq \text{diam}_E(K)^{1-p_1/p_2} W_{p_1}(\mu_1, \mu_2)^{p_1/p_2} \quad (23)$$

for all  $\mu_1, \mu_2 \in \mathcal{P}_c(E)$  of compact support contained in  $K$  (see [73, Chapter 5]), where  $\text{diam}_E(K)$  is the supremum of all  $d_E(y, y')$  over all possible  $y, y' \in K$ .

For  $p = 1$ , the duality formula for the Kantorovich-Rubinstein distance (see [78, Chapter 5]) gives the equivalent definition

$$W_1(\mu_1, \mu_2) = \sup \left\{ \int_E f d(\mu_1 - \mu_2) \mid f \in \text{Lip}(E), \text{Lip}(f) \leq 1 \right\}, \quad (24)$$

valid for all  $\mu_1, \mu_2 \in \mathcal{P}_1(E)$ .

For  $p = +\infty$ , we set  $W_\infty(\mu_1, \mu_2) = \lim_{p \rightarrow +\infty} W_p(\mu_1, \mu_2)$  (see [73, Chapter 5, Section 5.5.1]).

Note that the infimum in (21), as well as in (22), is a minimum (i.e., there exists an optimal coupling) and that the supremum in (24) is a maximum (see [78, Chapters 4 and 5] or [73, Chapter 3, Section 3.1.1]).

**Disintegration.** In this paper, we are going to consider measures on  $\Omega \times \mathbb{R}^d$ , for  $d \in \mathbb{N}^*$  (and on  $\Omega^k \times (\mathbb{R}^d)^k$  for  $k \in \mathbb{N}^*$ ), where  $(\Omega, d_\Omega)$  is a complete metric space and  $\mathbb{R}^d$  is endowed with an arbitrary norm  $\|\cdot\|$ . We endow  $\Omega \times \mathbb{R}^d$  with the distance  $d_{\Omega \times \mathbb{R}^d} = d_\Omega + d_{\mathbb{R}^d}$  where  $d_{\mathbb{R}^d}$  is the distance on  $\mathbb{R}^d$  induced by the norm  $\|\cdot\|$ .

Denoting by  $\pi : \Omega \times \mathbb{R}^d \rightarrow \Omega$  the canonical projection, given any  $\mu \in \mathcal{P}(\Omega \times \mathbb{R}^d)$ , in the sequel we will always denote by  $\nu$  the nonnegative probability Radon measure on  $\Omega$  defined as the image (pushforward) of  $\mu$  under  $\pi$ ,

$$\nu = \pi_* \mu = \mu \circ \pi^{-1}, \quad (25)$$

that is also the marginal of  $\mu$  on  $\Omega$ . Note that, since  $\pi$  is continuous,  $\text{supp}(\nu) = \overline{\pi(\text{supp}(\mu))}$ . By disintegration of  $\mu$  with respect to  $\nu$ , there exists a family  $(\mu_x)_{x \in \Omega}$  of probability Radon measures on  $\mathbb{R}^d$  (uniquely defined  $\nu$ -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  $h : \Omega \times \mathbb{R}^d \rightarrow [0, +\infty)$  (see, e.g., [11]). Moreover, we set  $\mu_x = 0$  whenever  $x \in \Omega \setminus \text{supp}(\nu)$ .

When  $\Omega$  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  $\nu$  is absolutely continuous, of density  $\frac{d\nu}{dx}(x) = \int_{\mathbb{R}^d} f(x, \xi) d\xi$ , and for  $\nu$ -almost every  $x \in \Omega$  the probability measure  $\mu_x$  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  $\nu$  on  $\Omega$ , we define

$$L_\nu^1 W_p(\mu^1, \mu^2) = \int_\Omega W_p(\mu_x^1, \mu_x^2) d\nu(x). \quad (26)$$

Obviously,  $L_\nu^1 W_p$  is a distance on the subset denoted  $\mathcal{P}_p^\nu(\Omega \times \mathbb{R}^d)$  of elements of  $\mathcal{P}_p(\Omega \times \mathbb{R}^d)$  having the same marginal  $\nu$ . Note that  $W_1(\mu^1, \mu^2) \leq L_\nu^1 W_1(\mu^1, \mu^2)$  for all  $\mu^1, \mu^2 \in \mathcal{P}_1^\nu(\Omega \times \mathbb{R}^d)$ .<sup>2</sup>

<sup>2</sup>Indeed,  $\int_{\Omega \times \mathbb{R}^d} f d(\mu^1 - \mu^2) = \int_\Omega \int_{\mathbb{R}^d} f(x, \xi) d(\mu_x^1 - \mu_x^2) d\nu(x) \leq \int_\Omega \text{Lip}(f(x, \cdot)) W_1(\mu_x^1, \mu_x^2) d\nu(x)$  for every  $f \in \text{Lip}(\Omega \times \mathbb{R}^d)$ , and if  $\text{Lip}(f) \leq 1$  then  $\text{Lip}(f(x, \cdot)) \leq 1$  for every  $x \in \Omega$ . Then, take the supremum over all  $f$ .

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

$$\Omega = \bigcup_{i=1}^N \Omega_i^N \quad \text{with} \quad \nu(\Omega_i^N) = \frac{1}{N} \quad \text{and} \quad \text{diam}_\Omega(\Omega_i^N) \leq \frac{C_\Omega}{N^r} \quad \forall i \in \{1, \dots, N\}, \quad (27)$$

for some  $C_\Omega > 0$  and  $r > 0$  not depending on  $N$ , and  $X^N = (x_1^N, \dots, x_N^N)$  is a  $N$ -tuple of points  $x_i^N \in \Omega_i^N$ . Here,  $\text{diam}_\Omega(\Omega_i^N)$  is the supremum of all  $d_\Omega(x, x')$  over all possible  $x, x' \in \Omega_i^N$ .

Families of tagged partitions always exist when  $\Omega$  is a compact  $n$ -dimensional smooth manifold having a boundary or not and  $\nu$  is a Lebesgue measure on  $\Omega$ , with  $r = 1/n$ . For instance, when  $\Omega = [0, 1]$ , we take  $\Omega_i^N = [a_i^N, a_{i+1}^N)$  for some subdivision  $0 = a_1^N < a_2^N < \dots < a_{N+1}^N = 1$  satisfying (27); when  $d\nu(x) = dx$ , a natural choice is  $a_i^N = \frac{i-1}{N}$ , and  $x_i^N = a_i^N$  or  $\frac{a_i^N + a_{i+1}^N}{2}$ , for every  $i \in \{1, \dots, N\}$  (and then  $C_\Omega = 1$  and  $r = 1$  in this case). When  $\Omega$  is a compact domain of  $\mathbb{R}^n$ , a family of tagged partitions is obtained by considering a family of meshes, as classically done in numerical analysis, with  $r = 1/n$ .

The concept of tagged partition is used in Riemann (and more generally, Henstock-Kurzweil) integration theory. We refer to [32] for (much more) general results. A real-valued function  $f$  on  $\Omega$ , of compact support, is said to be  $\nu$ -Riemann integrable if it is bounded,  $\nu$ -measurable, and if, for any family  $(\mathcal{A}^N, X^N)_{N \in \mathbb{N}^*}$  of tagged partitions, we have

$$\sum_{i=1}^N \int_{\Omega_i^N} |f(x) - f(x_i^N)| d\nu(x) = o(1) \quad (28)$$

and thus

$$\int_\Omega f d\nu = \frac{1}{N} \sum_{i=1}^N f(x_i^N) + o(1) \quad (29)$$

as  $N \rightarrow +\infty$ . A function  $f$  of essential compact support on  $\Omega$  is  $\nu$ -Riemann integrable if and only if  $f$  is bounded and continuous  $\nu$ -almost everywhere on  $\Omega$ .

## Acknowledgment

We are indebted to Claude Bardos, Julien Barré, Arnaud Debussche, Nicolas Fournier, Isabelle Gallagher, Thierry Gallay, François Golse, Alain Joye, Benoît Perthame, David Poyato, Mario Pulvirenti, Laure Saint-Raymond, Alain-Sol Sznitman and Eitan Tadmor for useful discussions.

## 2 From microscopic to macroscopic scale: the continuum / graph limit

In this section we 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 [55].

### 2.1 Continuum / graph limit equation

Given any  $\nu \in \mathcal{P}(\Omega)$ , we define the nonlinear operator  $A : \mathbb{R} \times L_\nu^\infty(\Omega, \mathbb{R}^d) \rightarrow L_\nu^\infty(\Omega, \mathbb{R}^d)$  (depending on  $\nu$ ) by

$$A(t, y)(x) = \int_\Omega G(t, x, x', y(x), y(x')) d\nu(x') \quad (30)$$

(recall that  $G$  satisfies Assumption **(G)**) for every  $t \in \mathbb{R}$  and for every  $y \in L^\infty_\nu(\Omega, \mathbb{R}^d)$ . We consider the continuum / graph limit equation

$$\boxed{\partial_t y(t, \cdot) = A(t, y(t, \cdot))} \quad (31)$$

It is a nonlinear (nonlocal) integro-differential equation.

We will see in Section 5 the interpretation of  $y(t, x)$  as a “velocity field” (moment of order 1 of the solution of the Vlasov equation).

**Theorem 2.1** (Existence and uniqueness for the continuum / graph limit equation (31)). *Assume that  $\Omega$  is compact. Let  $\nu \in \mathcal{P}(\Omega)$  and let  $y^0 \in L^\infty_\nu(\Omega, \mathbb{R}^d)$ . We denote by  $K' = \text{ess.im}(y^0)$  its essential range (it is a compact subset of  $\mathbb{R}^d$ ) and we set  $K = \Omega \times K'$  (compact). There exists a unique solution  $t \mapsto y(t, \cdot) \in L^\infty_\nu(\Omega, \mathbb{R}^d)$  on  $[0, T_{\max}(K))$  (where  $T_{\max}(K)$  is given by Lemma 1.1) of the nonlinear continuum / graph limit equation (31) such that  $y(0, \cdot) = y^0(\cdot)$ , of class  $\mathcal{C}^1$  with respect to  $t$ .*

*Moreover, if  $y^0 \in \mathcal{C}^0(\Omega, \mathbb{R}^d)$  then  $y(t, \cdot) \in \mathcal{C}^0(\Omega, \mathbb{R}^d)$  for every  $t \in [0, T_{\max}(K))$ .*

Local-in-time existence and uniqueness for the continuum / graph limit equation (31) straightforwardly follow from the Picard-Lindelöf (Cauchy-Lipschitz) theorem applied in the Banach space  $L^\infty_\nu(\Omega, \mathbb{R}^d)$ , but we prefer to see Theorem 2.1 as a consequence of Theorem 3.1 (existence and uniqueness for Vlasov equations) and of Proposition 5.2 (monokinetic measures), as it will be made precise in Remark 5.2 in Section 5.2.3. For the last statement of Theorem 2.1, we also refer to Theorem 2.2 and to its proof (see Appendix B.1) for possible variants.

**Remark 2.1.** When  $\Omega$  is not compact, the above result remains true provided that there exists a compact subset  $\Omega_1$  of  $\Omega$  such that  $G(t, x, x', \xi, \xi') = 0$  for every  $x \in \Omega \setminus \Omega_1$  and all  $(t, x', \xi, \xi') \in \mathbb{R} \times \Omega \times \mathbb{R}^d \times \mathbb{R}^d$ , and the initial condition  $y^0$  for the Euler equation satisfies  $y^0(x) = 0$  for  $\nu$ -almost every  $x \in \Omega \setminus \Omega_1$ . Indeed, in this case the solution of the continuum / graph limit equation is supported in  $\Omega_1$ . Alternatively, we can also assume that  $\text{supp}(\nu) \subset \Omega_1$ .

**Remark 2.2.** When  $G$  does not depend on  $(x, x')$  (and thus, particles are indistinguishable), it makes sense anyway to consider the continuum limit equation (31), with a solution  $y(t, x)$  depending on  $x \in \Omega$ . Although the particles are indistinguishable, the set of labels  $\Omega$  may be seen as a way to “enforce” distinguishability at the level of the Euler equation, by assigning to each particle a label that is an element of  $\Omega$ . As we will see in Section 2.3, such continuum limit equations do not seem to have been studied in the literature in the indistinguishable case. Note that distinguishability is made possible because we take an initial condition  $y^0(\cdot)$  depending on  $x \in \Omega$  in a nontrivial way. In contrast, if  $y^0(\cdot) \equiv y^0 \in \mathbb{R}^d$  is constant, then  $y(t, \cdot) \equiv y(t) \in \mathbb{R}^d$  does not depend on  $x$  (this follows from Remark B.1 in Appendix B.1) and the continuum limit equation becomes the differential equation  $\dot{y}(t) = G(t, y(t), y(t))$  in  $\mathbb{R}^d$ , which is much less meaningful.

**Remark 2.3** (“Empirically embedding” the particle system to the continuum / graph limit equation). In this remark, we assume that  $\nu = \nu_{X^N}^e = \frac{1}{N} \sum_{j=1}^N \delta_{x_j^N}$ . The operator  $A$  defined by (30) is then given by  $A(t, y)(x) = \frac{1}{N} \sum_{j=1}^N G(t, x, x_j^N, y(x), y(x_j^N))$  for every  $t \in \mathbb{R}$  and every  $y \in L^\infty_\nu(\Omega, \mathbb{R}^d)$ . Consequently:

- If  $t \mapsto \Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t))$  is a solution of the particle system (8) then, defining  $y(t, x) = \xi_i^N(t)$  if  $x = x_i^N$  for  $i \in \{1, \dots, N\}$  and 0 otherwise,  $t \mapsto y(t, \cdot)$  is a solution of the continuum / graph limit equation (31).
- Conversely, if  $t \mapsto y(t, \cdot)$  is a solution of the continuum / graph limit equation (31) then, defining  $\xi_i^N(t) = y(t, x_i^N)$  for  $i \in \{1, \dots, N\}$ ,  $t \mapsto \Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t))$  is a solution of the particle system (8). Note however that we may have  $y(t, x) \neq 0$  for  $x \notin \{x_1^N, \dots, x_N^N\}$ . This is a kind of projection.

The above empirical embedding is not very interesting. It is of much greater interest to fix a probability measure  $\nu \in \mathcal{P}(\Omega)$  and to approach the solutions of the continuum / graph limit equation, in some sense, by solutions of the particle system. The rough idea is to approximate the integral in the continuum / graph limit equation

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

by a Riemann sum, so that

$$\partial_t y(t, x_i^N) \simeq \frac{1}{N} \sum_{j=1}^N G(t, x_i^N, x_j^N, y(t, x_i^N), y(t, x_j^N))$$

for  $N$  sufficiently large, and then, comparing with (8), it is expected that  $\xi_i^N(t) \simeq y(t, x_i^N)$  for every  $i \in \{1, \dots, N\}$ , where  $t \mapsto \Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t))$  is solution of the particle system (8), for appropriate initial conditions. This is done in detail in Section 2.2 hereafter.

The application of the Riemann sum theorem is actually at the core of the notion of *graph limit* used in [55] to pass to the continuum limit in nonlocally coupled dynamical networks (see also the recent papers [5, 8, 13, 14, 28, 44]). Obtaining error estimates is then quite easy by developing standard numerical analysis arguments, which consist of estimating the discrepancy between an integral and approximating Riemann sums. This is the contents of the proofs of Theorems 2.2 and 2.3 hereafter.

The terminology “*graph limit*” refers to the graph interpretation of some classes of particle systems, like, very typically, the opinion propagation model given in Example 13 (see Section 2.3 for its graph limit): in this example, for any  $N \in \mathbb{N}^*$ , to the matrix of coefficients  $\sigma_{ij}^N$  is associated a directed graph whose vertices are the indices  $i \in \{1, \dots, N\}$  and which has an edge from  $i$  to  $j$  if  $\sigma_{ij}^N > 0$ . In this context, under Assumption (G), the function  $\sigma$  which satisfies  $\sigma(x_i^N, x_j^N) = \sigma_{ij}^N$  is referred to as a *graphon* and is the “continuum limit” of the graph as  $N \rightarrow +\infty$ . This is why the continuum / graph limit equation can also be called the graph limit of the system of particles. In [54, 55], for appropriate choices of interaction coefficients, the system (13) is interpreted as a nonlinear heat equation on a graph. The graph interpretation may be particularly relevant when wanting to prove, for instance, consensus results by exploiting the connectivity properties of the graph, as in [14]; we also mention [28] for exploiting the graph structure and [44] for the related mean field context. We stress anyway that, as said above, from the analysis point of view, taking the graph limit mainly consists of taking the limit in a Riemann sum, as in (29). This is thanks to this “numerical analysis” viewpoint that we can easily derive general error estimates, as shown hereafter.

## 2.2 Convergence estimates for the continuum / graph limit

Throughout this section, we assume that  $\Omega$  is compact. Let  $\nu \in \mathcal{P}(\Omega)$ . We consider the general nonlinear continuum / graph limit equation (31), with the nonlinear operator  $A$  defined by (30). Recall that  $G$  satisfies Assumption (G).

We also assume that there exists a family  $(\mathcal{A}^N, X^N)_{N \in \mathbb{N}^*}$  of tagged partitions associated with  $\nu$  satisfying (27) (see Section 1.5), with  $\mathcal{A}^N = (\Omega_1^N, \dots, \Omega_N^N)$  and  $X^N = (x_1^N, \dots, x_N^N)$ . We have the following two theorems.

**Theorem 2.2.** *Let  $y^0$  be a bounded and  $\nu$ -almost everywhere continuous function on  $\Omega$  (thus,  $\nu$ -Riemann integrable), with values in  $\mathbb{R}^d$ .*

*On the one part, we consider the unique solution  $t \mapsto y(t, \cdot) \in L^\infty(\Omega, \mathbb{R}^d)$  on  $[0, T_{\max}(K))$  of the (nonlinear) continuum / graph limit equation (31) such that  $y(0, \cdot) = y^0(\cdot)$ , where  $K = \Omega \times$*



$\text{ess.im}(y^0)$  (compact) and  $\text{ess.im}(y^0) \subset \mathbb{R}^d$  is the essential range of  $y^0$ .

On the other, for any  $N \in \mathbb{N}^*$ , we consider the unique solution  $t \mapsto \Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t)) \in \mathbb{R}^{dN}$  on  $[0, T_{\max}(K))$  of the particle system (8) such that  $\xi_i^N(0) = y^0(x_i^N)$  for every  $i \in \{1, \dots, N\}$ , and we set

$$y^N(t, x) = \sum_{i=1}^N \xi_i^N(t) \mathbf{1}_{\Omega_i^N}(x) \quad \forall (t, x) \in \mathbb{R} \times \Omega \quad (32)$$

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

- For every  $t \in [0, T_{\max}(K))$ ,  $y(t, \cdot)$  is bounded and continuous  $\nu$ -almost everywhere on  $\Omega$ , with the same continuity set as  $y^0$ , and

$$\|y(t, \cdot) - y^N(t, \cdot)\|_{L^\infty(\Omega, \mathbb{R}^d)} = o(1) \quad (33)$$

as  $N \rightarrow +\infty$ , where the remainder term  $o(1)$  is uniform with respect to  $t$  on any compact interval of  $[0, T_{\max}(K))$ . In particular,

$$\max_{i \in \{1, \dots, N\}} \|y(t, x_i^N) - \xi_i^N(t)\| = o(1). \quad (34)$$

- Assume that there exists  $\alpha \in (0, 1]$  such that  $y^0 \in \mathcal{C}^{0, \alpha}(\Omega, \mathbb{R}^d)$  and  $G$  is locally  $\alpha$ -Hölder continuous with respect to  $(x, x', \xi, \xi')$  (uniformly with respect to  $t$  on any compact). Then, for every  $t \in [0, T_{\max}(K))$ , we have  $y(t, \cdot) \in \mathcal{C}^{0, \alpha}(\Omega, \mathbb{R}^d)$  with

$$\text{Hol}_\alpha(y(t, \cdot)) \leq e^{tL_y(t)} (1 + \text{Hol}_\alpha(y(0, \cdot))) \quad (35)$$

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

$$\max_{i \in \{1, \dots, N\}} \|y(t, x_i^N) - \xi_i^N(t)\| \leq \frac{C_\Omega^\alpha}{N^{\tau\alpha}} (1 + \text{Hol}_\alpha(y^0)) e^{2tL_y^N(t)} \quad (36)$$

and actually,

$$\|y(t, \cdot) - y^N(t, \cdot)\|_{L^\infty(\Omega, \mathbb{R}^d)} \leq 2 \frac{C_\Omega^\alpha}{N^{\tau\alpha}} (1 + \text{Hol}_\alpha(y^0)) e^{2tL_y^N(t)} \quad (37)$$

where  $C_\Omega$  is given by (27). The constant  $L_y^N(t)$  in (36) and (37) is defined by

$$L_y^N(t) = \max_{0 \leq \tau \leq t} \text{Hol}_\alpha(G(\tau, \cdot, \cdot, \cdot, \cdot)_{|\Omega^2 \times S_y^N(\tau)^2}) + \max_{\substack{x, x' \in \Omega \\ 0 \leq \tau \leq t}} \text{Lip}(G(\tau, x, x', \cdot, \cdot)_{|S_y^N(\tau)^2}) \quad (38)$$

where  $S_y^N(\tau) \subset \mathbb{R}^d$  is the (compact) convex closure of all  $y(\tau, x)$  for  $x \in \Omega$  and all  $\xi_i^N(\tau)$  for  $i \in \{1, \dots, N\}$ . The constant  $L_y(t)$  in (35) is defined as  $L_y^N(t)$  but with  $S_y^N(\tau)$  replaced by  $S_y(\tau)$  that is the convex closure of all  $y(\tau, x)$  for  $x \in \Omega$ , i.e., like  $S_y^N(\tau)$  but without the  $\xi_i^N(\tau)$ . We have  $L_y(t) \leq L_y^N(t)$ .

Theorem 2.2 is proved in Appendix B.1. Note that, by Lemma 1.1, given any  $T \in [0, T_{\max}(K))$ , the sets  $S_y^N(t)$  and thus the scalars  $L_y^N(t)$  are uniformly bounded with respect to  $t \in [0, T]$  and to  $N \in \mathbb{N}^*$ .

**Remark 2.4.** Having in mind Remark 2.1, Theorem 2.2 can be extended to the case where  $\Omega$  is not compact, under the following additional assumptions:

- the family of tagged partitions is such that the points  $x_i^N$  remain in a compact subset of  $\Omega$ ;

- the initial condition  $y^0$  is of compact essential support;
- the set  $S_y^N(\tau) \subset \Omega \times \mathbb{R}^d$  is defined as the compact closure of all  $(x, y(\tau, x))$  for  $x \in \text{ess supp}(y(\tau, \cdot))$  (essential support) and all  $(x_i^N, \xi_i^N(\tau))$  for  $i \in \{1, \dots, N\}$ .

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

**Theorem 2.3.** *Let  $K'$  be a compact subset of  $\mathbb{R}^d$ . Given any  $N \in \mathbb{N}^*$ , let  $\Xi_0^N \in (K')^N$ . We set  $K = \Omega \times K'$ .*

*On the one part, we consider the unique solution  $t \mapsto \Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t)) \in \mathbb{R}^{dN}$  on  $[0, T_{\max}(K))$  of the particle system (8) such that  $\Xi^N(0) = \Xi_0^N$ , and we define  $y^N(t, x)$  by (32).*

*On the other part, we consider the unique solution  $t \mapsto y_N(t, \cdot) \in L^\infty(\Omega, \mathbb{R}^d)$  on  $[0, T_{\max}(K))$  of the continuum / graph limit equation (31) such that  $y_N(0, \cdot) = y^N(0, \cdot)$  (i.e.,  $y_N(0, x) = \xi_i^N(0)$  if  $x \in \Omega_i^N$ ). Then, for every  $t \in [0, T_{\max}(K))$ ,*

$$\|y_N(t, \cdot) - y^N(t, \cdot)\|_{L^\infty(\Omega, \mathbb{R}^d)} = o(1) \quad (39)$$

*as  $N \rightarrow +\infty$ , where the remainder term  $o(1)$  is uniform with respect to  $t$  on any compact interval of  $[0, T_{\max}(K))$ . If moreover  $G$  is locally  $\alpha$ -Hölder continuous with respect to  $(x, x', \xi, \xi')$  (uniformly with respect to  $t$  on any compact), then, for every  $N \in \mathbb{N}^*$  and every  $t \in [0, T_{\max}(K))$ ,*

$$\|y_N(t, \cdot) - y^N(t, \cdot)\|_{L^\infty(\Omega, \mathbb{R}^d)} \leq 2 \frac{C_\Omega^\alpha}{N^{r\alpha}} e^{2tL_{y_N}^N(t)} \quad \forall t \geq 0, \quad (40)$$

where  $L_{y_N}^N(t)$  is defined by (38) (with  $y$  replaced by  $y_N$ ).

Theorem 2.3 is proved in Appendix B.2. Note that, by Lemma 1.1, given any  $T \in [0, T_{\max}(K))$ , the scalars  $L_{y_N}^N(t)$  are uniformly bounded with respect to  $t \in [0, T]$  and to  $N \in \mathbb{N}^*$ .

Note that, in particular, taking  $x = x_i$  in (40), we have

$$\max_{i \in \{1, \dots, N\}} \|y_N(t, x_i^N) - \xi_i^N(t)\| \leq 2 \frac{C_\Omega^\alpha}{N^{r\alpha}} e^{2tL_{y_N}^N(t)},$$

which improves the estimates obtained in [5].

**Remark 2.5.** The trivial case where  $y^0(\cdot) \equiv y^0 \in \mathbb{R}^d$ , mentioned in Remark 2.2, corresponds in the framework of Theorems 2.2 and 2.3 to taking  $\xi_i^N(0)$  not depending on  $i$  (equivalently,  $\xi_i^N(t)$  not depending on  $i$ , for every  $t$ ), which is the case where all particles coincide.

## 2.3 Examples

Let us follow some of the examples of particle systems given in Section 1.3 and give their continuum / graph limit equation.

– For the Hegselmann–Krause (opinion propagation) model (13), under Assumption (G) we have  $G(t, x, x', \xi, \xi') = \sigma(x, x')(\xi' - \xi)$  and the continuum / graph limit equation is

$$\partial_t y(t, x) = \int_\Omega \sigma(x, x')(y(t, x') - y(t, x)) d\nu(x'). \quad (41)$$

In this case, the operator  $A$  (defined by (30)) is linear and is given by  $(Ay)(x) = \int_\Omega \sigma(x, x')(y(x') - y(x)) d\nu(x')$  for every  $y \in L_\nu^\infty(\Omega, \mathbb{R}^d)$  (see [8]). Extended to  $L_\nu^2(\Omega, \mathbb{R}^d)$ , it is a Hilbert-Schmidt operator. The spectral study of  $A$  has been done in [14] (with  $\nu$  the Lebesgue measure) in view of

deriving consensus results. The graph interpretation is particularly meaningful on this example, and the function  $\sigma$  (assumed to exist) is called a graphon (see [6] for a recent study and see [13] for a time-varying case).

– For the Kuramoto particle system (14), under Assumption (G) we have  $\Omega = \mathbb{R} \times [0, 1]$  and  $G(t, x, x', \xi, \xi') = \alpha + \sigma(\beta, \beta')(\xi' - \xi)$  (where  $x = (\alpha, \beta)$  and  $x' = (\alpha', \beta')$ ), and the continuum / graph limit equation is

$$\partial_t y(t, x) = \alpha + \int_{\Omega} \sigma(\beta, \beta') \sin(y(t, x') - y(t, x)) d\nu(x').$$

The graph limit operator at the right-hand side of the above equation is introduced in [22] although, in that reference, the authors focus on the study of the mean field limit (see Section 3.3).

– For the first-order system (15), we have  $G(t, x, x', \xi, \xi') = F(\xi) + K(\xi - \xi')$  and the continuum / graph limit equation is

$$\partial_t y(t, x) = F(y(t, x)) + \int_{\Omega} K(y(t, x) - y(t, x')) d\nu(x'). \quad (42)$$

In this case, the operator  $A$  (defined by (30)) is nonlinear, nonlocal, and does not depend on  $t$ . The equation (42) seems to be new and has not been studied in the literature.

– For the Cucker–Smale dynamics (17), setting  $y = (y_1, y_2) \in \mathbb{R}^r \times \mathbb{R}^r$ , the continuum / graph limit equation is

$$\begin{aligned} \partial_t y_1(t, x) &= y_2(t, x), \\ \partial_t y_2(t, x) &= \int_{\Omega} a(\|y_1(t, x') - y_1(t, x)\|)(y_2(t, x') - y_2(t, x)) d\nu(x'). \end{aligned} \quad (43)$$

As before, the equation (43) seems to be new. As discussed in Remark 2.2, thanks to the set  $\Omega$  we have in some sense “enforced” distinguishability. If one takes an initial condition that is constant with respect to  $x$  then  $y_1(t, x) = y_1(t)$  and  $y_2(t, x) = y_2(t)$  do not depend on  $x$  and we have  $\dot{y}_1(t) = y_2(t)$  and  $\dot{y}_2(t) = 0$ , which is the familiar fact that the derivative of the average of positions is the average velocity, and the average velocity is constant.

The Cucker–Smale dynamics (17) is a second-order system. For second-order dynamics, we will see in Section 5.6 later a different definition of Euler equation which gives rise to interesting (and already known and studied) dynamics.

– For the (indistinguishable) second-order dynamics (18), similarly to the Cucker–Smale example, the continuum limit equation is

$$\begin{aligned} \partial_t y_1(t, x) &= y_2(t, x), \\ \partial_t y_2(t, x) &= \int_{\Omega} K(y_1(t, x), y_1(t, x')) d\nu(x'). \end{aligned}$$

### 3 From microscopic to mesoscopic scale I: mean field through empirical measures (from ODEs to Vlasov)

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  $N \rightarrow +\infty$ , 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  $f(t, x, \xi)$  is the density of particles having the parameter  $x$  at  $\xi$  at time  $t$ .

### 3.1 Vlasov equation

Given any  $\mu \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ , we define  $\nu = \pi_*\mu$  by (25) (marginal of  $\mu$  on  $\Omega$ ) 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

$$\begin{aligned} \mathcal{X}[\mu](t, x, \xi) &= \int_{\Omega \times \mathbb{R}^d} G(t, x, x', \xi, \xi') d\mu(x', \xi') \\ &= \int_{\Omega} \int_{\mathbb{R}^d} G(t, x, x', \xi, \xi') d\mu_{x'}(\xi') d\nu(x') \quad \forall (t, x, \xi) \in \mathbb{R} \times \Omega \times \mathbb{R}^d \end{aligned} \quad (44)$$

(recall that  $G$  satisfies Assumption (G)). Note that  $\mathcal{X}[\mu](t, x, \xi)$  is the expectation of  $G(t, x, x', \xi, \xi')$  with respect to the measure  $\mu$ , performed with respect to  $(x', \xi') \in \Omega \times \mathbb{R}^d$  (see Appendix A.6 for more details and consequences of that definition).

We consider the *Vlasov (or continuity) equation*

$$\partial_t \mu + \operatorname{div}_{\xi}(\mathcal{X}[\mu]\mu) = 0 \quad (45)$$

where the divergence<sup>3</sup> acts only with respect to  $\xi$ . It is a nonlocal transport equation because the velocity field  $\mathcal{X}[\mu]$  defined by (44) is nonlocal.

**Remark 3.1.** Given any solution  $t \mapsto \mu(t)$  on  $[0, T]$  of the Vlasov equation (45) (see further for the rigorous definition of a solution), 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  $t \in [0, T]$ . 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  $\xi$ , 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  $\Omega$  (which does not depend on  $t$  by Remark 3.1), by uniqueness  $\nu$ -almost everywhere of the disintegration, (45) is equivalent to

$$\partial_t \mu_{t,x} + \operatorname{div}_{\xi}(\mathcal{X}[\mu_t](t, x, \cdot)\mu_{t,x}) = 0 \quad (46)$$

for  $\nu$ -almost every  $x \in \Omega$ . Note that the time evolution of  $\mu_{t,x}$  depends on the whole  $\mu_0$  and not only on  $\mu_{0,x}$ , since  $\mathcal{X}[\mu_t]$  involves an integral over all possible  $x' \in \Omega$ .

Therefore, the Vlasov equation (45) can be thought of as an infinite number (if  $\Omega$  has an infinite number of elements) of coupled Vlasov equations (46). The most standard case studied in the literature corresponds to a measure  $\mu$  not depending on  $x$ .

Given any interval  $I \subset \mathbb{R}$ , let  $\mathcal{C}^0(I, \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  be the Banach space of continuous mappings  $t \in I \mapsto \mu(t) \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ , with  $\mathcal{P}_c(\Omega \times \mathbb{R}^d)$  endowed with the weak topology (metrized by the Wasserstein distance  $W_p$ , for any  $p \in [1, +\infty)$ , as recalled in Section 1.5).

We define  $\mathcal{C}_{\text{comp}}^0(I, \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  as the set of all  $\mu \in \mathcal{C}^0(I, \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  that are *equi-compactly supported* on any compact interval of  $I$ , meaning that for any  $t_1, t_2 \in I$ , there exists a compact subset  $K \subset \Omega \times \mathbb{R}^d$  such that  $\operatorname{supp}(\mu(t)) \subset K$  for every  $t \in [t_1, t_2]$ . There exist elements of

<sup>3</sup>Recall that  $\operatorname{div}(\mathcal{X}\mu) = L_{\mathcal{X}}\mu$  (Lie derivative of the measure  $\mu$ ) is the measure defined by  $\langle L_{\mathcal{X}}\mu, f \rangle = -\langle \mu, L_{\mathcal{X}}f \rangle = -\int_{\mathbb{R}^d} \mathcal{X} \cdot \nabla f d\mu$  for every  $f \in \mathcal{C}_c^{\infty}(\mathbb{R}^d)$ .

$\mathcal{C}^0(I, \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  that are not equi-compactly supported on any compact interval of  $I$  (for instance, if  $I = [0, T]$ , take  $\mu(t) = (1 - e^{-1/t})\delta_0 + e^{-1/t}\delta_{1/t}$ ).

In view of obtaining existence and uniqueness of solutions of the Vlasov equation (45), we consider the following concept of solution. Assuming that  $0 \in I$ , by definition, a solution  $t \mapsto \mu(t)$  of (45) on  $I$  such that  $\mu(0) = \mu_0 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$  is an element  $\mu \in \mathcal{C}_{\text{comp}}^0(I, \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  such that, denoting  $\mu_t = \mu(t)$ ,<sup>4</sup> for every  $g \in \mathcal{C}_c^\infty(\Omega \times \mathbb{R}^d)$ , the function  $t \mapsto \int g d\mu_t$  is absolutely continuous on  $I$  and

$$\begin{aligned} \int_{\Omega \times \mathbb{R}^d} g(x, \xi) d\mu_t(x, \xi) &= \int_{\Omega \times \mathbb{R}^d} g(x, \xi) d\mu_0(x, \xi) \\ &+ \int_0^t \int_{\Omega \times \mathbb{R}^d} \int_{\Omega \times \mathbb{R}^d} \langle \nabla_\xi g(x, \xi), G(\tau, x, x', \xi, \xi') \rangle d\mu_\tau(x', \xi') d\mu_\tau(x, \xi) d\tau \end{aligned} \quad (47)$$

for almost every  $t \in I$ .

**Theorem 3.1.** *[Existence, uniqueness and stability properties for the Vlasov equation (45)]* Recalling Assumption (G), let  $p \in [1, +\infty)$  be arbitrary.

(A) Given any  $\mu_0 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ , setting  $T_0 = T_{\max}(\text{supp}(\mu_0))$  (given by Lemma 1.1), there exists a unique solution  $\mu \in \mathcal{C}_{\text{comp}}^0([0, T_0], \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  of the Vlasov equation (45) (in the sense (47)) such that  $\mu(0) = \mu_0$ . Moreover,  $t \mapsto \mu(t)$  is locally Lipschitz with respect to  $t$  for the distance  $W_p$ , and we have

$$\mu(t) = \varphi_{\mu_0}(t) * \mu_0, \quad (48)$$

which is a notation meaning that  $\mu_{t,x} = \varphi_{\mu_0}(t, x, \cdot) * \mu_{0,x}$  for every  $t \in [0, T_0]$  and  $\nu$ -almost every  $x \in \Omega$ , and where  $t \mapsto \varphi_{\mu_0}(t, x, \cdot)$  is the unique solution (Vlasov flow) of

$$\partial_t \varphi_{\mu_0}(t, x, \cdot) = \mathcal{X}[\mu(t)](t, x, \cdot) \circ \varphi_{\mu_0}(t, x, \cdot) \quad (49)$$

such that  $\varphi_{\mu_0}(0, x, \cdot) = \text{id}_{\mathbb{R}^d}$  for  $\nu$ -almost every  $x \in \Omega$ . Moreover, if  $\mu_0 \in \mathcal{P}_c^{\text{ac}}(\Omega \times \mathbb{R}^d)$  then  $\mu(t) \in \mathcal{P}_c^{\text{ac}}(\Omega \times \mathbb{R}^d)$  for every  $t \in [0, T_0]$ . Furthermore:

(A<sub>1</sub>) Any solution of (45) depends continuously on its initial condition  $\mu(0) \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$  for the weak topology in the following sense: given any compact subset  $K \subset \Omega \times \mathbb{R}^d$ , given any  $\mu(0) \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$  such that  $\text{supp}(\mu(0)) \subset K$ , given any (equi-compactly supported) sequence of measures  $\mu^k(0) \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$  such that  $\text{supp}(\mu^k(0)) \subset K$  for every  $k \in \mathbb{N}^*$ , if  $\mu^k(0)$  converges weakly to  $\mu(0)$  (equivalently,  $W_p(\mu^k(0), \mu(0)) \rightarrow 0$ ) as  $k \rightarrow +\infty$ , then  $\mu^k(t)$  converges weakly to  $\mu(t)$  (equivalently,  $W_p(\mu^k(t), \mu(t)) \rightarrow 0$ ) as  $k \rightarrow +\infty$ , uniformly on any compact interval of  $[0, T_{\max}(K)]$ .

(A<sub>2</sub>) For all solutions  $\mu^1, \mu^2 \in \mathcal{C}_{\text{comp}}^0([0, T], \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  of (45) (for some  $T > 0$ ) such that  $\mu^1(0), \mu^2(0) \in \mathcal{P}_c^\nu(\Omega \times \mathbb{R}^d)$  have the same marginal  $\nu$  on  $\Omega$ , setting<sup>5</sup>

$$S_{\mu^1, \mu^2}(\tau) = \text{supp}(\nu) \times (\varphi_{\mu_0^1}(\tau, \text{supp}(\mu_0^1)) \cup \text{supp}(\mu_0^2)) \cup \text{supp}(\mu^2(\tau))$$

and defining

$$C_{\mu^1, \mu^2}(t) = \exp \left( 2 \int_0^t \text{ess sup}_{(x, \xi), (x', \xi') \in S_{\mu^1, \mu^2}(\tau)} \|(\partial_\xi G, \partial_{\xi'} G)(\tau, x, x', \xi, \xi')\| d\tau \right), \quad (50)$$

<sup>4</sup>Note that, seeing  $\mu$  as a measure on  $I \times \Omega \times \mathbb{R}^d$ , the marginal of  $\mu$  on  $I$  is the Lebesgue measure and the disintegration of  $\mu$  is  $\mu = \int_I \mu_t dt$ .

<sup>5</sup>Note that  $S_{\mu^1, \mu^2}(t)$  is compact, that  $\varphi_{\mu_0^1}(t, \text{supp}(\mu_0^1)) = \text{supp}(\mu_1^1)$  and  $\text{supp}(\mu^1(t)) \cup \text{supp}(\mu^2(t)) \subset S_{\mu^1, \mu^2}(t)$ .

we have

$$L_\nu^1 W_p(\mu^1(t), \mu^2(t)) \leq C_{\mu^1, \mu^2}(t) L_\nu^1 W_p(\mu^1(0), \mu^2(0)) \quad \forall t \in [0, T] \quad (51)$$

(where  $L_\nu^1 W_p$  is defined by (26)).

(B) Assume moreover that  $G$  is locally Lipschitz with respect to  $(x, x', \xi, \xi')$  uniformly with respect to  $t$  on any compact interval. For all solutions  $\mu^1(\cdot), \mu^2(\cdot) \in \mathcal{C}_{\text{comp}}^0([0, T], \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  of (45) (for some  $T > 0$ ), setting

$$S_{\mu^1, \mu^2}(\tau) = (\text{supp}(\nu_1) \cup \text{supp}(\nu_2)) \times (\varphi_{\mu_0^1}(\tau, \text{supp}(\mu_0^1) \cup \text{supp}(\mu_0^2)) \cup \text{supp}(\mu^2(\tau)))$$

and defining

$$C_{\mu^1, \mu^2}(t) = \exp\left(2 \int_0^t \text{Lip}(G(\tau, \cdot, \cdot, \cdot, \cdot))|_{S_{\mu^1, \mu^2}(\tau)^2} d\tau\right), \quad (52)$$

we have

$$W_p(\mu^1(t), \mu^2(t)) \leq C_{\mu^1, \mu^2}(t) W_p(\mu^1(0), \mu^2(0)) \quad \forall t \in [0, T]. \quad (53)$$

Theorem 3.1 is proved in Appendix B.3. The statement (B) of Theorem 3.1 is a slight extension, with parameter  $x$ , of [68, Theorem 2.3] (see also [65, 66, 67]) where it is assumed that  $G$  is globally Lipschitz. Without parameter  $x$ , we recover the famous stability estimate obtained by Dobrushin in [27] (see Corollary 3.1 further). The statement (A) seems to be new. Note that, in (A<sub>2</sub>), 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  $t$ ). On the contrary, in (A<sub>1</sub>) and in (B), the measures under consideration are not assumed to have the same marginal. In (A<sub>1</sub>), 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  $\mu_x^k(0) \rightharpoonup \mu_x(0)$  for  $x \in \Omega$ .

In the statement (B), the assumption that  $G$  is locally Lipschitz with respect to  $(x, x', \xi, \xi')$  is much stronger than (G): in Example 13 (resp., Example 17) this requires  $\sigma$  (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 (8). Note that, under this stronger assumption, the unique solution  $\mu(\cdot)$  in (A) is locally Lipschitz with respect to  $t$  for the Wasserstein distance  $W_1$ .

Finally, in the usual statements existing in the literature,  $G$  is assumed to be globally Lipschitz. Here, under the weaker assumption (G), we have a maximal time of definition of  $\mu$  depending on the compact support of  $\mu_0$ , according to Lemma 1.1. Note that, when  $G$  is bounded, we can consider in Theorem 3.1 measures that are not of compact support.

**Particular case where  $G$  does not depend on  $(x, x')$ .** When  $G$  does not depend on  $(x, x')$ , particles are indistinguishable: this is the classical case that has been much studied in the existing literature. Let us show how this can be recovered from our more general framework. Given any measure  $\mu \in \mathcal{P}(\Omega \times \mathbb{R}^d)$ , we define  $\bar{\mu} \in \mathcal{P}(\mathbb{R}^d)$  as the image of  $\mu$  under the projection of  $\Omega \times \mathbb{R}^d$  onto  $\mathbb{R}^d$ , that is,

$$\int_{\mathbb{R}^d} f(\xi) d\bar{\mu}(\xi) = \int_{\Omega \times \mathbb{R}^d} f(\xi) d\mu(x, \xi) = \int_{\Omega} \int_{\mathbb{R}^d} f(\xi) d\mu_x(\xi) d\nu(x) \quad (54)$$

for every Borel measurable function  $f : \mathbb{R}^d \rightarrow [0, +\infty)$ . Since  $G$  does not depend on  $(x, x')$ , the mean field  $\mathcal{X}[\mu]$  defined by (44) does not depend on  $x$  and we have  $\mathcal{X}[\mu](t, x, \xi) = \bar{\mathcal{X}}[\bar{\mu}](t, \xi)$  where the mean field  $\bar{\mathcal{X}}[\bar{\mu}]$  is defined by

$$\bar{\mathcal{X}}[\bar{\mu}](t, \xi) = \int_{\mathbb{R}^d} G(t, \xi, \xi') d\bar{\mu}(\xi) \quad \forall (t, \xi) \in \mathbb{R} \times \mathbb{R}^d. \quad (55)$$

Accordingly, since the projection onto  $\mathbb{R}^d$  commutes with  $\partial_t$  and with the divergence with respect to  $\xi$ , it follows that, if  $t \mapsto \mu(t)$  is a solution of the Vlasov equation (45) then  $t \mapsto \bar{\mu}(t)$  is the solution of the Vlasov equation (without dependence on  $x$ )

$$\partial_t \bar{\mu} + \operatorname{div}(\bar{\mathcal{X}}[\bar{\mu}]\bar{\mu}) = 0. \quad (56)$$

We have the following corollary of Theorem 3.1, already well known in the existing literature (famous Dobrushin estimate, see [27]).

**Corollary 3.1.** *Let  $p \in [1, +\infty)$  be arbitrary. Given any  $\bar{\mu}_0 \in \mathcal{P}_c(\mathbb{R}^d)$ , there exists a unique solution  $\bar{\mu} \in \mathcal{C}_{\text{comp}}^0([0, T_{\max}(\operatorname{supp}(\bar{\mu}_0))], \mathcal{P}_c(\mathbb{R}^d))$  of the Vlasov equation (56), locally Lipschitz with respect to  $t$  for the distance  $W_p$ , such that  $\bar{\mu}(0) = \bar{\mu}_0$ , and we have*

$$\bar{\mu}(t) = \varphi_{\bar{\mu}_0}(t, \cdot) * \bar{\mu}_0$$

where  $t \mapsto \varphi_{\bar{\mu}_0}(t, \cdot)$  is the unique solution of  $\partial_t \varphi_{\bar{\mu}_0}(t, \cdot) = \bar{\mathcal{X}}[\mu(t)](t, \cdot) \circ \varphi_{\bar{\mu}_0}(t, \cdot)$  such that  $\varphi_{\bar{\mu}_0}(0, \cdot) = \operatorname{id}_{\mathbb{R}^d}$ . Moreover, if  $\bar{\mu}_0 \in \mathcal{P}_c^{\text{ac}}(\mathbb{R}^d)$  then  $\bar{\mu}(t) \in \mathcal{P}_c^{\text{ac}}(\mathbb{R}^d)$  for every  $t \in \mathbb{R}$ . Furthermore, we have

$$W_p(\bar{\mu}^1(t), \bar{\mu}^2(t)) \leq C_{\bar{\mu}^1, \bar{\mu}^2}(t) W_p(\bar{\mu}^1(0), \bar{\mu}^2(0)) \quad \forall t \in [0, T] \quad (57)$$

for all solutions  $\bar{\mu}^1(\cdot)$  and  $\bar{\mu}^2(\cdot)$  of (56) on  $[0, T]$  (for some  $T > 0$ ) such that  $\bar{\mu}^1(0), \bar{\mu}^2(0) \in \mathcal{P}_c(\mathbb{R}^d)$ . Here,  $C_{\bar{\mu}^1, \bar{\mu}^2}(t)$  is defined by (50) or (52) (without dependence on  $x$ ).

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

This embedding allows us to recover Corollary 3.1 as a consequence of Theorem 3.1. Indeed, obviously,  $\bar{\mu}(\cdot)$  is solution of the Vlasov equation (56) without dependence on  $x$  if and only if  $\mu(\cdot) = \bar{\nu} \otimes \bar{\mu}(\cdot)$  is solution of the Vlasov equation (45). This gives the first part of the corollary.

To obtain (57), it suffices to take  $\bar{\nu} = \delta_{\bar{x}}$  for some  $\bar{x} \in \Omega$  and to note that  $W_p(\bar{\mu}^1, \bar{\mu}^2) = W_p(\bar{\nu} \otimes \bar{\mu}^1, \bar{\nu} \otimes \bar{\mu}^2)$ . Then, (57) follows from (51) or from (53).  $\square$

### 3.2 Relationship between the particle system and the Vlasov equation

For every  $N \in \mathbb{N}^*$ , given any  $X^N = (x_1^N, \dots, x_N^N) \in \Omega^N$  and any  $\Xi^N = (\xi_1^N, \dots, \xi_N^N) \in \mathbb{R}^{dN}$ , we define the *empirical measure*  $\mu_{(X^N, \Xi^N)}^e \in \mathcal{P}(\Omega \times \mathbb{R}^d)$  corresponding to  $(X^N, \Xi^N)$  by

$$\mu_{(X^N, \Xi^N)}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N}. \quad (58)$$

The disintegration of  $\mu_{(X^N, \Xi^N)}^e$  with respect to its marginal  $\nu_{\Xi^N}^e = \pi_* \mu_{(X^N, \Xi^N)}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N}$  on  $\Omega$  (that is itself an empirical measure corresponding to  $X^N$ ) gives the family of conditional measures defined by  $(\mu_{(X^N, \Xi^N)}^e)_x = \delta_{\xi_i^N}$  if  $x = x_i^N$  and 0 otherwise.

The relationship between the particle system (8) and the Vlasov equation (45) is given by the result below. To simplify the notation, hereafter we denote  $\mu_N^e = \mu_{(X^N, \Xi^N)}^e$  and  $\nu_N^e = \nu_{\Xi^N}^e$ .

**Proposition 3.1.** *If  $t \mapsto \Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t)) \in \mathbb{R}^{dN}$  is a solution on  $[0, T]$  (for some  $T > 0$ ) of the particle system (9) with parameter  $X^N = (x_1^N, \dots, x_N^N) \in \Omega^N$ , then*

$$t \mapsto \mu_N^e(t) = \varphi_{\mu_N^e(0)}(t) * \mu_N^e(0) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N(t)}$$

is a solution of the Vlasov equation (45) on  $[0, T]$ . The converse is true if all  $x_i^N$  are distinct and all  $\xi_i^N(t)$  are distinct.

Actually,  $t \mapsto \Xi^N(t)$  is solution on  $[0, T]$  of (9) with parameter  $X^N$  if and only if

$$\xi_i^N(t) = \varphi_{\mu_{\bar{x}_i^N}^e}(t, x_i^N, \xi_i^N(0)) \quad \forall t \in [0, T] \quad \forall i \in \{1, \dots, N\}. \quad (59)$$

*Proof.* The Vlasov equation (45) is written as  $\partial_t \mu + L_{\mathcal{X}[\mu]} \mu = 0$  with the Lie derivative acting with respect to the variable  $\xi$ . Hence, setting  $X^N(t) = (x_1^N(t), \dots, x_N^N(t))$  and  $\Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t))$ , the mapping  $t \mapsto \mu_{\bar{x}_i^N}^e(t)$  is a solution of the Vlasov equation (45) if and only if, for any  $g \in \mathcal{C}_c^\infty(\Omega \times \mathbb{R}^d)$ , we have  $\langle \partial_t \mu_{\bar{x}_i^N}^e + L_{\mathcal{X}[\mu_{\bar{x}_i^N}^e]} \mu_{\bar{x}_i^N}^e, g \rangle = 0$ , i.e.,

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

which is satisfied if  $t \mapsto (X^N, \Xi^N(t))$  is solution of (8). If all  $x_i^N$  are distinct and all  $\xi_i^N(t)$  are distinct, the converse is obtained by taking  $g$  localized around  $(x_i^N, \xi_i^N(t))$ .

To obtain the second part of the proposition, we note that

$$\mathcal{X}[\mu_N^e](t, x, \xi) = \frac{1}{N} \sum_{j=1}^N G(t, x, x_j^N, \xi, \xi_j^N)$$

and thus  $\mathcal{X}[\mu_N^e](t, x_i^N, \xi_i^N) = Y_i(t, X^N, \Xi^N)$  for every  $i \in \{1, \dots, N\}$ . Therefore, (9) is equivalent to  $\xi_i^N(t) = \mathcal{X}[\mu_N^e](t, x_i^N, \xi_i^N(t))$  for every  $i \in \{1, \dots, N\}$ . Besides, by definition of  $t \mapsto \varphi_{\mu_{\bar{x}_i^N}^e}(t, x_i^N, \cdot)$  (given in (A) in Theorem 3.1), we have

$$\partial_t \varphi_{\mu_{\bar{x}_i^N}^e}(t, x_i^N, \xi_i^N(0)) = \mathcal{X}[\mu_N^e](t, x_i^N, \varphi_{\mu_{\bar{x}_i^N}^e}(t, x_i^N, \xi_i^N(0)))$$

with  $\varphi_{\mu_{\bar{x}_i^N}^e}(0, x_i^N, \xi_i^N(0)) = \xi_i^N(0)$ . Then, (59) follows by Cauchy uniqueness.  $\square$

As a consequence of the statements (A<sub>1</sub>) and (B) of Theorem 3.1 and of Proposition 3.1, we have the following corollary (the last part of which is already well known in the indistinguishable case).

**Corollary 3.2.** *Let  $K$  be a compact subset of  $\Omega \times \mathbb{R}^d$ . Let  $p \in [1, +\infty)$  be arbitrary. Let  $\mu_0 \in \mathcal{P}_c(K)$  and let  $t \mapsto \mu(t) = \varphi_{\mu_0}(t, \cdot, \cdot) * \mu_0$  be the solution on  $[0, T_{\max}(K))$  of the Vlasov equation (45) such that  $\mu(0) = \mu_0$ . Besides, for every  $N \in \mathbb{N}^*$ , let  $(X^N, \Xi_0^N) \in K^N$  be such that the empirical measure  $\mu_N^e(0) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N(0)}$  converges weakly (equivalently, in Wasserstein distance  $W_p$ ) to  $\mu_0$  as  $N \rightarrow +\infty$  (see Appendix A.3 for general results). For every  $N \in \mathbb{N}^*$ , let  $t \mapsto \Xi^N(t)$  be the solution on  $[0, T_{\max}(K))$  of the particle system (9) with parameter  $X^N$  such that  $\Xi^N(0) = \Xi_0^N$ .*

*Then, the empirical measure  $\mu_N^e(t) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N(t)}$  converges weakly (equivalently, in Wasserstein distance  $W_p$ ) to  $\mu(t)$  as  $N \rightarrow +\infty$ , uniformly with respect to  $t$  on any compact interval of  $[0, T_{\max}(K))$ .*

*If moreover  $G$  is locally Lipschitz with respect to  $(x, x', \xi, \xi')$  (uniformly with respect to  $t$  on any compact), then*

$$W_p(\mu(t), \mu_N^e(t)) \leq C_{\mu, \mu_{\bar{x}_i^N}^e}(t) W_p(\mu_0, \mu_N^e(0))$$

*for every  $t \in [0, T_{\max}(K))$  (with  $C_{\mu, \mu_{\bar{x}_i^N}^e}(t)$  defined by (50)).*

Lemmas A.15 and A.16 in Appendix A.3 provide general results ensuring that  $W_p(\mu_0, \mu_N^e(0)) \rightarrow 0$  as  $N \rightarrow +\infty$ , and Lemma A.17 gives an estimate of convergence, at rate  $\frac{1}{N^{r/p}}$ , within the framework of tagged partitions.



**Remark 3.2.** Alternatively, instead of empirical measures, we may also consider *semi-empirical measures*: setting

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

the unique solution  $t \mapsto \tilde{\mu}^N(t) = \varphi_{(\mu_0)_{X^N}^{se}}(t, \cdot, \cdot)_*(\mu_0)_{X^N}^{se}$  of the Vlasov equation (45) such that  $\tilde{\mu}^N(0) = (\mu_0)_{X^N}^{se}$  is of the form  $\tilde{\mu}^N(t) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \tilde{\mu}_{t,x_i^N}^N$  (it differs from the semi-empirical measure  $\mu(t)_{X^N}^{se} = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \mu_{t,x_i^N}$ ). Its marginal on  $\Omega$  is the empirical measure  $\nu_{X^N}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N}$ .

Lemma A.19 in Appendix A.4.2 provides results on the convergence of  $W_p(\mu_0, (\mu_0)_{X^N}^{se})$  to 0, as well as estimates with a rate of convergence under appropriate assumptions.

### 3.3 Examples

We follow up with examples given in Sections 1.3 (particle systems) and 2.3 (Euler equation).

– For the Hegselmann–Krause (opinion propagation) system (13), under Assumption (G) the mean field (not depending on  $t$ ) is

$$\mathcal{X}[\mu](x, \xi) = \int_{\Omega \times \mathbb{R}^d} \sigma(x, x')(\xi' - \xi) d\mu(x', \xi') \quad \forall (x, \xi) \in \Omega \times \mathbb{R}^d. \quad (60)$$

The Vlasov equation has been derived and studied in [8, 64] (see also [14, Section 5.2]).

– For the Kuramoto particle system (14), under Assumption (G) the mean field (not depending on  $t$ ) is given by

$$\mathcal{X}[\mu](x, \xi) = \alpha + \int_{\Omega \times \mathbb{R}^d} \sigma(x, x') \sin(\xi' - \xi) d\mu(x', \xi') \quad \forall (x, \xi) \in \Omega \times \mathbb{R}^d.$$

The corresponding Vlasov equation was proposed in [72] as being a formal mean field limit of (14). The rigorous mean field limit, called the Kuramoto–Sakaguchi equation, was established in [50] in the case where  $\sigma$  is constant, by following the classical fixed point arguments of [61, 75]. The general (network) case is treated in [22, 48] and the Vlasov equation associated to the above mean field with the general function  $\sigma$ , is studied in that reference in view of extending the synchronization theory to spatially structured networks. The Vlasov equation [22, Eq. (16)] is of the form (46), i.e., it consists of an infinite number of coupled Vlasov equations, parametrized (and coupled) by  $x = (\alpha, \beta)$ .

– For the first-order system (15), the mean field does not depend on  $(t, x)$  and is given by

$$\begin{aligned} \mathcal{X}[\mu](x, \xi) &= F(\xi) + \int_{\Omega} \int_{\mathbb{R}^d} K(\xi - \xi') d\mu_{x'}(\xi') d\nu(x') = F(\xi) + \int_{\Omega} K \star \mu_{x'}(\xi) d\nu(x') \\ &= F(\xi) + \int_{\mathbb{R}^d} K(\xi - \xi') d\bar{\mu}(\xi) = F(\xi) + K \star \bar{\mu}(\xi) = \bar{\mathcal{X}}[\bar{\mu}](\xi) \quad \forall (x, \xi) \in \Omega \times \mathbb{R}^d \end{aligned}$$

where  $\bar{\mu}$  is defined by (54) and  $\bar{\mathcal{X}}[\bar{\mu}]$  by (55). The Vlasov equation  $\partial_t \bar{\mu} + \operatorname{div}((F + K \star \bar{\mu})\bar{\mu}) = 0$  is used in mathematical biology to model aggregation phenomena (see [18, 19, 26]), in the study of neural networks (see [71]) or, when  $K$  is a singular kernel, in fluid mechanics (see [46, 74]).

– For the Cucker–Smale model (17), the mean field does not depend on  $(t, x)$  and is given by

$$\begin{aligned}\mathcal{X}[\mu](x, \xi) &= \left( \int_{\Omega \times \mathbb{R}^r \times \mathbb{R}^r} a(\|q - q'\|) \frac{p}{\|p' - p\|} d\mu(x', \xi') \right) \\ &= \left( \int_{\mathbb{R}^r \times \mathbb{R}^r} a(\|q - q'\|) \frac{p}{\|p' - p\|} d\bar{\mu}(\xi') \right) = \bar{\mathcal{X}}[\bar{\mu}](\xi) \quad \forall (x, \xi) \in \Omega \times \mathbb{R}^{2r}\end{aligned}$$

where we recall that  $\xi = (q, p)$  and  $\xi' = (q', p')$ . The kinetic Cucker–Smale equation satisfied by  $\bar{\mu}$  has been derived in [38, 58]. Convergence to flocking has been studied in [20, 37].

– For the second-order model (18), similarly to the Cucker–Smale example, is given by

$$\bar{\mathcal{X}}[\bar{\mu}](\xi) = \left( \int_{\mathbb{R}^r \times \mathbb{R}^r} K(q, q') d\bar{\mu}(\xi') \right) \quad \forall \xi \in \mathbb{R}^{2r}.$$

– As an example of a Hamiltonian system, the mean field associated with the mapping (20) is

$$\bar{\mathcal{X}}[\bar{\mu}](\xi) = \left( -\nabla V(q) + dA(q) \cdot (p - A(q)) - \int_{\mathbb{R}^r \times \mathbb{R}^r} (\partial_1 W(q, q') + \partial_2 W(q', q)) d\bar{\mu}(\xi') \right)$$

for every  $\xi \in \mathbb{R}^{2r}$ .

## 4 From microscopic to mesoscopic scale II: mean field by lifting the particle system (from Liouville to Vlasov)

### 4.1 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  $\Phi^N(t, X, \cdot)$  of  $\mathbb{R}^{dN}$  generated by the time-dependent vector field  $Y^N(t, X, \cdot)$  defined by (10).

Given  $N \in \mathbb{N}^*$  fixed, we consider the ( $N$ -body) *Liouville equation* associated with the time-dependent vector field  $Y^N$  defined by (10), depending on the parameter  $X^N \in \Omega^N$ , given by

$$\boxed{\partial_t \rho^N + \operatorname{div}_{\Xi}(Y^N \rho^N) = 0} \quad (61)$$

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

**Proposition 4.1.** *Let  $K$  be a compact subset of  $\Omega \times \mathbb{R}^d$ . Let  $\rho_0^N \in \mathcal{P}_c(\Omega^N \times \mathbb{R}^{dN})$  be such that all marginals of  $\rho_0^N$  on any copy of  $\Omega \times \mathbb{R}^d$  are supported in the same compact  $K$ . There exists a unique solution  $t \mapsto \rho^N(t)$  of the Liouville equation (61) in  $\mathcal{C}^0([0, T_{\max}(K)], \mathcal{P}_c(\Omega^N \times \mathbb{R}^{dN}))$ , locally Lipschitz with respect to  $t$  for the distance  $L_\theta^1 W_1$  (where  $\theta$  is defined below), such that  $\rho^N(0) = \rho_0^N$ , given by*

$$\rho^N(t) = \Phi^N(t)_* \rho_0^N \quad (62)$$

*i.e.,  $\rho^N(t)$  is the image (pushforward) of  $\rho_0^N$  under the particle flow.*

The notation (62) is slightly abusive. To explain it, 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  $\pi^{\otimes N} : \Omega^N \times \mathbb{R}^{dN} \rightarrow \Omega^N$  the canonical projection, we will always denote by  $\theta$  the probability Radon measure on  $\Omega^N$  given by  $\theta = (\pi^{\otimes N})_* \rho$  (image of  $\rho$  under  $\pi^{\otimes N}$ ), that is the marginal of  $\rho$  on  $\Omega^N$ . By disintegration of  $\rho$  with respect to  $\theta$ , 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,  $\rho_t^N = \rho^N(t)$  is disintegrated as  $\rho_t^N = \int_{\Omega^N} \rho_{t,X}^N d\theta^N(X)$  with respect to its marginal  $\theta^N = (\pi^{\otimes N})_* \rho^N(t)$  on  $\Omega^N$ . The marginal  $\theta^N$  does not depend on  $t$  because (61) can be written as  $\partial_t \rho^N + L_{Y^N} \rho^N = 0$ , with the Lie derivative acting with respect to the variable  $\xi$ , and we have  $(\pi^{\otimes N})_* L_{Y^N} = 0$ . Finally, (62) means that

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

for every  $t \in [0, T_{\max}(K))$  and for  $\theta^N$ -almost every  $X \in \Omega^N$ .

**Remark 4.1.** If  $\rho_0^N = \delta_{X^N} \otimes \delta_{\Xi_0^N}$  for some  $(X^N, \Xi_0^N) \in K^N$  then  $\rho^N(t) = \delta_{X^N} \otimes \delta_{\Xi^N(t)}$  where  $t \mapsto \Xi^N(t)$  is the solution on  $[0, T_{\max}(K))$  of the particle system (9) with parameter  $X^N$  such that  $\Xi^N(0) = \Xi_0^N$ . 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 (9), instead of taking a *deterministic* initial condition  $\Xi^N(0) = \Xi_0^N \in \mathbb{R}^{dN}$ , 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  $\Xi_0^N$  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 (61) has a *probabilistic* interpretation with respect to the particle system (8).

If the probability measure  $\rho^N(t)$  on  $\Omega^N \times \mathbb{R}^{dN}$  has a density  $f^N$ , then  $f^N(t, X, \Xi)$  represents the density of particles having the positions  $X = (x_1, \dots, x_N) \in \Omega^N$  and respective momenta  $\Xi = (\xi_1, \dots, \xi_N) \in \mathbb{R}^{dN}$ . 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 derive Vlasov from Liouville by taking marginals.

## 4.2 Deriving Vlasov from Liouville by taking marginals, propagation of chaos

Compared with  $\mu(t)$  that is a probability measure on  $\Omega \times \mathbb{R}^d$ ,  $\rho^N(t)$  is a probability measure on  $(\Omega \times \mathbb{R}^d)^N \simeq \Omega^N \times \mathbb{R}^{dN}$ . It is thus tempting to search for a relationship between  $\mu(t)$  and  $\rho^N(t)$  by taking marginals of  $\rho^N(t)$ . This is what has been done in [76], in [43, 46] or in [35, 36] in the different context of quantum mechanics. Adapted to the present situation, the method developed in [36], which provides an explicit rate of convergence, consists of proving that the marginals of the solutions  $\rho(t)$  of (61) are close, in Wasserstein topology, to solutions  $\mu(t)$  of the Vlasov equation (45), as established hereafter.

As we are going to see, this can be done by taking adequate initial conditions  $\rho_0^N$  for the Liouville equation (61). We have to perform a symmetrization under permutations for the initial condition  $\rho_0^N$  and also for the corresponding solution  $\rho^N(t)$ , not only with respect to  $\Xi$  but also with respect to the parameter variable  $X$ . Note that the symmetrization is not preserved by the flow, so we have to consider the symmetrization  $\rho^N(t)^s$  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 sym-

metrization under permutations of  $\rho$  (see Appendix A.1.2), by

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

for every  $f \in \mathcal{C}_c^0(\Omega^N \times \mathbb{R}^{dN})$ , where  $\sigma \cdot X = (x_{\sigma(1)}, \dots, x_{\sigma(N)})$  and  $\sigma \cdot \Xi = (\xi_{\sigma(1)}, \dots, \xi_{\sigma(N)})$  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  $k \in \{1, \dots, N\}$ , we denote by  $\rho_{N:k}^s$  the  $k^{\text{th}}$ -order marginal of  $\rho^s$  (not to be confused with the symmetrization under permutations of the marginal, which we do not use), which is, by definition, the image of  $\rho^s$  under the projection of  $\Omega^N \times \mathbb{R}^{dN}$  onto the product  $\Omega^k \times \mathbb{R}^{dk}$  of the  $k$  first copies of  $\Omega$  with the  $k$  first copies of  $\mathbb{R}^d$ .

Since we are going to compute Wasserstein distances in  $(\Omega \times \mathbb{R}^d)^k \simeq \Omega^k \times \mathbb{R}^{dk}$ , we have to choose a distance in that space. Recall that  $\Omega \times \mathbb{R}^d$  is equipped with the distance  $d_{\Omega \times \mathbb{R}^d} = d_\Omega + d_{\mathbb{R}^d}$  where  $d_{\mathbb{R}^d}$  is the distance on  $\mathbb{R}^d$  induced by the norm  $\|\cdot\|$  (which is arbitrary). Let  $q \in [1, +\infty]$  be arbitrary. Given any  $k \in \mathbb{N}^*$ , we endow  $(\Omega \times \mathbb{R}^d)^k$  with the  $\ell^q$  distance based on  $d_{\Omega \times \mathbb{R}^d}$ , defined by

$$\begin{aligned} d_{(\Omega \times \mathbb{R}^d)^k}^{[q]}((X, \Xi), (X', \Xi')) &= \|(d_{\Omega \times \mathbb{R}^d}((x_1, \xi_1), (x'_1, \xi'_1)), \dots, d_{\Omega \times \mathbb{R}^d}((x_k, \xi_k), (x'_k, \xi'_k)))\|_{\ell^q} \\ &= \begin{cases} \left( \sum_{i=1}^k (d_\Omega(x_i, x'_i) + \|\xi_i - \xi'_i\|)^q \right)^{1/q} & \text{if } q \in [1, +\infty) \\ \max_{1 \leq i \leq k} (d_\Omega(x_i, x'_i) + \|\xi_i - \xi'_i\|) & \text{if } q = +\infty \end{cases} \end{aligned} \quad (63)$$

for all  $X = (x_1, \dots, x_k)$  and  $X' = (x'_1, \dots, x'_k)$  in  $\Omega^k$  and for all  $\Xi = (\xi_1, \dots, \xi_k)$  and  $\Xi' = (\xi'_1, \dots, \xi'_k)$  in  $\mathbb{R}^{dk}$ . Note that, when  $k = 1$ , we have  $d_{\Omega \times \mathbb{R}^d}^{[q]} = d_{\Omega \times \mathbb{R}^d}^{[1]} = d_\Omega + d_{\mathbb{R}^d}$ .

Given any  $p, q \in [1, +\infty]$ , we denote by  $W_p^{[q]}$  the Wasserstein distance  $W_p$  on  $\mathcal{P}(\Omega^k \times (\mathbb{R}^d)^k)$  with respect to the distance  $d_{(\Omega \times \mathbb{R}^d)^k}^{[q]}$ .

We refer to the beginning of Appendix A.1 and in particular to Remark A.1 for comments on the importance of choosing a distance on the product space  $(\Omega \times \mathbb{R}^d)^k$  and for remarks on the Wasserstein distance  $W_p^{[q]}$ . In particular, by (88), we have  $W_p^{[q_2]} \leq W_p^{[q_1]} \leq k^{\frac{1}{q_1} - \frac{1}{q_2}} W_p^{[q_2]}$  if  $1 \leq q_1 \leq q_2 \leq +\infty$  for any  $p \in [1, +\infty]$ .

In this section, we establish two ways for deriving 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  $\Omega$ . Setting  $T_0 = T_{\max}(\text{supp}(\mu_0))$  (as given by Lemma 1.1), we consider the unique solution  $t \mapsto \mu(t) = \varphi_{\mu_0}(t)_* \mu_0$  in  $\mathcal{C}_{\text{comp}}^0([0, T_0], \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  of the Vlasov equation (45) such that  $\mu(0) = \mu_0$ , as given by Theorem 3.1. Recall that  $\mu_{t,x} = \varphi_{\mu_0}(t, x, \cdot)_* \mu_{0,x}$  for  $\nu$ -almost every  $x \in \Omega$ .

Hereafter, we propose two possible choices of  $\rho_0^N \in \mathcal{P}_c(\Omega^N \times \mathbb{R}^{dN})$ , generating by Proposition 4.1 the solution  $\rho^N(t) = \Phi^N(t)_* \rho_0^N$  of the Liouville equation (61) from which we recover at the larger  $N$  limit the solution  $\mu(t)$  of the Vlasov equation (45) by taking marginals.

In Theorem 4.1, we take  $\rho_0^N$  Dirac; in Theorem 4.2, we take  $\rho_0^N$  “semi-Dirac”. In both cases, we prove that  $\rho^N(t)_{N:k}^s$  converges to  $\mu(t)^{\otimes k}$  as  $N \rightarrow +\infty$  and we establish convergence estimates in Wasserstein distance  $W_p^{[q]}$ . The fact that the  $k^{\text{th}}$ -order marginal  $\rho^N(t)_{N:k}^s$  of the symmetrization of  $\rho^N(t)$ , which is absolutely not a tensor product at time  $t = 0$ , becomes however the tensor product  $\mu(t)^{\otimes k}$  at the limit  $N \rightarrow +\infty$ , is usually referred to as *propagation of chaos* (formalized in the pioneering articles [47, 53], see also [35, 56, 75, 76]).

### 4.2.1 First way, with $\rho_0^N$ Dirac

Given any fixed  $N \in \mathbb{N}^*$ , let  $(X^N, \Xi_0^N) \in \text{supp}(\mu_0)^N \subset \Omega^N \times \mathbb{R}^{dN}$  be arbitrary. Typically we may want that the empirical measure  $\mu_{(X^N, \Xi_0^N)}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_{0,i}^N}$  converges to  $\mu_0$  in Wasserstein distance as  $N \rightarrow +\infty$  (see Appendix A.3 for such conditions). Let  $t \mapsto \Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t))$  be the solution on  $[0, T_0)$  of the particle system (9) such that  $\Xi^N(0) = \Xi_0^N$ . If  $\mu_{(X^N, \Xi_0^N)}^e$  converges to  $\mu_0$  then, by Corollary 3.2, the empirical measure  $\mu_{(X^N, \Xi^N(t))}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N(t)}$  converges to  $\mu(t)$  in Wasserstein distance as  $N \rightarrow +\infty$ .

Defining  $\rho_0^N \in \mathcal{P}_c(\Omega^N \times \mathbb{R}^{dN})$  as the Dirac measure  $\rho_0^N = \delta_{X^N} \otimes \delta_{\Xi_0^N}$ , by Remark 4.1, the unique solution of the Liouville equation (61) such that  $\rho^N(0) = \rho_0^N$ , is given by the Dirac measure

$$\rho^N(t) = \Phi^N(t)_* \rho_0^N = \delta_{X^N} \otimes \delta_{\Xi^N(t)} \quad \forall t \in [0, T_0).$$

It is then easy to see that  $\rho^N(t)_{N:1}^s = \mu_{(X^N, \Xi^N(t))}^e$  (see the proof of the theorem below). Therefore, if  $\mu_{(X^N, \Xi_0^N)}^e$  converges weakly to  $\mu_0$  then  $\rho^N(t)_{N:1}^s$  converges weakly to  $\mu(t)$  as  $N \rightarrow +\infty$ . Actually, this first fact re-expresses results seen in Sections 3.1 and 3.2. The convergence is less obvious for the marginals of order  $k \geq 2$ .

Recall that  $G$  satisfies Assumption (G).

**Theorem 4.1.** *We have the following statements, for any  $p \in [1, +\infty)$  and  $q \in [1, +\infty]$ .*

- (A) *If  $\mu_{(X^N, \Xi_0^N)}^e$  converges weakly (equivalently, in Wasserstein distance  $W_p$ ) to  $\mu_0$  as  $N \rightarrow +\infty$ , then, for every  $k \in \mathbb{N}^*$ ,  $\rho^N(t)_{N:k}^s$  converges weakly (equivalently, in Wasserstein distance  $W_p^{[q]}$ ) to  $\mu(t)^{\otimes k}$  as  $N \rightarrow +\infty$ , uniformly with respect to  $t$  on any compact interval of  $[0, T_0)$ .*
- (B) *Assuming that  $G$  is locally Lipschitz with respect to  $(x, x', \xi, \xi')$  (uniformly with respect to  $t$  on any compact), setting*

$$S_\mu^N(\tau) = \text{supp}(\mu(\tau)) \cup \{(x_i^N, \xi_i^N(\tau)) \mid i \in \{1, \dots, N\}\} \quad (64)$$

and defining

$$C_\mu^N(t) = \exp\left(2 \int_0^t \text{Lip}(G(\tau, \cdot, \cdot, \cdot, \cdot))|_{S_\mu^N(\tau)^2} d\tau\right), \quad (65)$$

for every  $N \in \mathbb{N}^*$  and for every  $t \in [0, T_0)$  we have  $\rho^N(t)_{N:1}^s = \mu_{(X^N, \Xi^N(t))}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N(t)}$  and

$$W_p(\rho^N(t)_{N:1}^s, \mu(t)) = W_p(\bar{\mu}_{\Xi^N(t)}^e, \bar{\mu}(t)) \leq C_\mu^N(t) W_p(\mu_{(X^N, \Xi_0^N)}^e, \mu_0) \quad (66)$$

and, for every  $k \in \mathbb{N}^*$  such that  $k^2 \leq 2N \ln(1 + \frac{1}{2^p})$ ,

$$\begin{aligned} W_p^{[q]}(\rho^N(t)_{N:k}^s, \mu(t)^{\otimes k}) &\leq 2k^{1/q} \left(\frac{k^2}{N}\right)^{1/p} (\text{diam}_\Omega(\text{supp}(\nu)) + \text{diam}_{\mathbf{R}^d}(\Xi^N(t))) \\ &\quad + k^{1/q} C_\mu^N(t) W_p(\mu_{(X^N, \Xi_0^N)}^e, \mu_0) \end{aligned} \quad (67)$$

where  $\text{diam}_{\mathbf{R}^d}(\Xi^N(t)) = \max_{1 \leq i, j \leq N} \|\xi_i^N(t) - \xi_j^N(t)\|$ .

Theorem 4.1 is proved in Appendix B.4.

In (67),  $\text{diam}_\Omega(\text{supp}(\nu)) = \max_{x, x' \in \text{supp}(\nu)} d_\Omega(x, x')$ , and the Wasserstein distance  $W_p^{[q]}$  on  $\Omega^k \times (\mathbb{R}^d)^k$  is computed with respect to the distance  $d_{\Omega^k \times (\mathbb{R}^d)^k}^{[q]}$  defined by (63). Since  $W_p^{[q]} \leq k^{\frac{1}{q}} W_p^{[\infty]}$  (by (88)), the strongest inequality (67) is obtained when  $q = +\infty$ .

Lemmas A.15 and A.16 in Appendix A.3 show that there always exists a sequence of empirical measures  $\mu_{(X^N, \Xi_0^N)}^e$  converging weakly to  $\mu_0$ . As alluded above, to obtain an interesting convergence estimate from Item (B) of this theorem, we apply Lemma A.17 in Appendix A.3, which yields the estimate  $W_p(\mu_{(X^N, \Xi_0^N)}^e, \mu_0) \leq \frac{1}{N^{r/p}} C_{\Omega \times \mathbb{R}^d}^{1/p} \text{diam}_{\Omega \times \mathbb{R}^d}(\text{supp}(\mu_0))^{1-1/p}$  under the assumption of the existence of a family of tagged partitions. As noted in this appendix, there exist plenty of results quantifying the convergence of empirical measures to a given measure (see, e.g., [31]). Lemma A.17 is a rough result.

**Corollary 4.1.** *In the context of Item (B) of Theorem 4.1, we assume moreover that there exists a family of tagged partitions of  $\text{supp}(\mu_0)$  associated with  $\mu_0$  (see Section 1.5), i.e., for every  $N \in \mathbb{N}^*$  there exists a partition of  $\text{supp}(\mu_0) = \cup_{i=1}^N F_i^N$  such that all subsets  $F_i^N \subset \Omega \times \mathbb{R}^d$  are  $\mu_0$ -measurable, pairwise disjoint, satisfy  $\mu_0(F_i^N) = \frac{1}{N}$  and  $\text{diam}_{\Omega \times \mathbb{R}^d}(F_i^N) \leq C_{\Omega \times \mathbb{R}^d}/N^r$  for some  $C_{\Omega \times \mathbb{R}^d} > 0$  and  $r > 0$  not depending on  $N$ , and  $N$ -tuples  $X^N = (x_1^N, \dots, x_N^N) \in \Omega^N$  and  $\Xi_0^N = (\xi_{0,1}^N, \dots, \xi_{0,N}^N) \in (\mathbb{R}^d)^N$  such that  $(x_i^N, \xi_{0,i}^N) \in F_i^N$  for every  $i \in \{1, \dots, N\}$ . Then, for every  $t \in [0, T_0)$ ,*

$$W_p(\rho^N(t)_{N:1}^s, \mu(t)) = W_p(\bar{\mu}_{\Xi^N(t)}^e, \bar{\mu}(t)) \leq \frac{1}{N^{r/p}} C_{\Omega \times \mathbb{R}^d}^{1/p} \text{diam}_E(\text{supp}(\mu_0))^{1-1/p} C_\mu^N(t)$$

and, for every  $k \in \mathbb{N}^*$  such that  $k^2 \leq 2N \ln(1 + \frac{1}{2^p})$ ,

$$W_p^{[\infty]}(\rho^N(t)_{N:k}^s, \mu(t)^{\otimes k}) \leq 2 \left( \frac{k^2}{N} \right)^{1/p} (\text{diam}_\Omega(\text{supp}(\nu)) + \text{diam}_{\mathbb{R}^d}(\Xi^N(t))) + \frac{C_{\Omega \times \mathbb{R}^d}^{1/p}}{N^{r/p}} \text{diam}_{\Omega \times \mathbb{R}^d}(\text{supp}(\mu_0))^{1-1/p} C_\mu^N(t). \quad (68)$$

When  $\Omega$  is a  $n$ -dimensional manifold (thus  $\dim(\Omega \times \mathbb{R}^d) = n + d$ ), we have  $r = 1/(n + d) < 1$ .

According to the estimate (68),  $\rho^N(t)_{N:k}^s$  converges to  $\mu(t)^{\otimes k}$  in Wasserstein distance  $W_p^{[\infty]}$  as  $N \rightarrow +\infty$ , uniformly with respect to  $t$  on compact intervals of  $[0, T_0)$ , at rate  $1/N^{r/p}$  if  $k \ll N^{(1-r)/2}$  and at rate  $k^{2/p}/N^{1/p}$  if  $N^{(1-r)/2} \ll k \ll N^{1/2}$ . The rate of convergence can be improved if one uses better results for convergence of empirical measures.

Note that the assumption of a family of tagged partitions in Corollary 4.1 essentially entails that  $\mu_0$  be absolutely continuous with respect to a Lebesgue measure on  $\Omega \times \mathbb{R}^d$ .

**Particular case where  $G$  does not depend on  $(x, x')$ .** When  $G$  does not depend on  $(x, x')$ , particles are indistinguishable and the mean field is given by (55). We have the following corollary of Theorem 4.1.

**Corollary 4.2.** *Let  $\bar{\mu}_0 \in \mathcal{P}_c(\mathbb{R}^d)$  and let  $t \mapsto \bar{\mu}(t)$  be the unique solution on  $[0, T_0)$ , with  $T_0 = T_{\max}(\text{supp}(\bar{\mu}_0))$ , of the Vlasov equation (56) such that  $\bar{\mu}(0) = \bar{\mu}_0$  (see Corollary 3.1). Besides, let  $\bar{\rho}_0^N = \delta_{\Xi_0^N}$  and let  $t \mapsto \bar{\rho}^N(t) = \delta_{\Xi^N(t)}$  be the unique solution on  $[0, T_0)$  of the Liouville equation (61) (without dependence on  $X$ ) such that  $\bar{\rho}^N(0) = \bar{\rho}_0^N$ . Then, for every  $t \in [0, T_0)$ ,*

$$W_p(\bar{\rho}^N(t)_{N:1}^s, \bar{\mu}(t)) = W_p(\bar{\mu}_{\Xi^N(t)}^e, \bar{\mu}(t)) \leq C_{\bar{\mu}}^N(t) W_p(\bar{\mu}_{\Xi_0^N}^e, \bar{\mu}_0) \quad (69)$$

and we have  $\bar{\rho}^N(t)_{N:1}^s = \mu_{\Xi^N(t)}^e = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_i^N(t)}$  (empirical measure), where  $C_{\bar{\mu}}^N(t)$  is defined by (65) (without dependence on  $x, x'$ ), and, for every  $k \in \{2, \dots, N\}$ ,

$$W_p^{[\infty]}(\bar{\rho}^N(t)_{N:k}^s, \bar{\mu}(t)^{\otimes k}) \leq 2 \left( \frac{k^2}{N} \right)^{1/p} \text{diam}_{\mathbb{R}^d}(\Xi^N(t)) + C_{\bar{\mu}}^N(t) W_p(\bar{\mu}_{\Xi_0^N}^e, \bar{\mu}_0). \quad (70)$$

*Proof.* Following the proof of Corollary 3.1 and choosing  $\bar{\nu} = \delta_{\bar{x}}$  for some arbitrary  $\bar{x} \in \Omega$ , when  $G$  does not depend on  $(x, x')$ ,  $\bar{\mu}(\cdot)$  is solution of the Vlasov equation (56) (without dependence on  $x$ ) if and only if  $\mu(\cdot) = \delta_{\bar{x}} \otimes \bar{\mu}(\cdot)$  is solution of the Vlasov equation (45). We now define  $X^N = (\bar{x}, \dots, \bar{x}) \in \Omega^N$ , and we take  $\rho_0^N = \delta_{X^N} \otimes \delta_{\Xi_0^N}$  as initial condition for the Liouville equation in Theorem 4.1, so that  $\rho^N(t) = \delta_{X^N} \otimes \bar{\rho}^N(t)$  where  $\bar{\rho}^N(t) = \delta_{\Xi^N(t)}$ . With these choices, we obviously have  $\rho^N(t)_{N:k}^s = \delta_{\bar{x}}^{\otimes k} \otimes \bar{\rho}^N(t)_{N:k}^s$ , and then (69) and (70) straightforwardly follows from (66) and (67), by applying Remark A.2 in Appendix A.1.4.  $\square$

#### 4.2.2 Second way, with $\rho_0^N$ “semi-Dirac”

Given any fixed  $N \in \mathbb{N}^*$ , let  $X^N = (x_1^N, \dots, x_N^N) \in \Omega^N$  be arbitrary. We set  $\delta_{X^N} = \delta_{x_1^N} \otimes \dots \otimes \delta_{x_N^N}$  and  $\rho_{0, X^N}^N = \mu_{0, x_1^N} \otimes \dots \otimes \mu_{0, x_N^N}$ . Defining  $\rho_0^N \in \mathcal{P}_c(\Omega^N \times \mathbb{R}^{dN})$  as the “semi-Dirac” measure  $\rho_0^N = \delta_{X^N} \otimes \rho_{0, X^N}^N$ , we consider the unique solution on  $[0, T_0]$  of the Liouville equation (61) such that  $\rho^N(0) = \rho_0^N$ , given by the “semi-Dirac” measure

$$\rho^N(t) = \Phi^N(t) * \rho_0^N = \delta_{X^N} \otimes \Phi(t, X^N, \cdot) * \rho_{0, X^N}^N = \delta_{X^N} \otimes \rho_{t, X^N}^N.$$

Note indeed that the marginal  $\theta^N = (\pi^{\otimes N}) * \rho_t^N$  of  $\rho_t^N = \rho^N(t)$  on  $\Omega^N$  is  $\theta^N = \delta_{X^N}$ , and that  $\rho_{t, X^N}^N = \Phi^N(t, X^N, \cdot) * \rho_{0, X^N}^N$ .

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

$$(\rho_0^N)_{N:1}^s = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \mu_{0, x_i^N} = (\mu_0)_{X^N}^{se} \quad (71)$$

(semi-empirical measure), which converges weakly to  $\mu_0$  as  $N \rightarrow +\infty$  under slight assumptions on  $\mu_0$ , by Lemma A.19 in Appendix A.4.2. More generally,  $(\rho_0^N)_{N:k}^s$  converges weakly to  $\mu_0^{\otimes k}$  (in the proof of the theorem hereafter, we give an explicit expression for  $(\rho_0^N)_{N:k}^s$ , using (105) in Appendix A.2.3). In the theorem below, we establish that this convergence is propagated in time.

**Theorem 4.2.** *We assume that the norm  $\|\cdot\|$  on  $\mathbb{R}^d$  is induced by a scalar product on  $\mathbb{R}^d$ . Let  $p \in [1, 2]$  and  $q \in [1, +\infty]$  be arbitrary.*

- (A) *Assume that  $x \mapsto \mu_{0, x}$  is  $\nu$ -almost everywhere continuous for the Wasserstein distance  $W_p$ . Then, for every  $k \in \mathbb{N}^*$ ,  $\rho^N(t)_{N:k}^s$  converges weakly to  $\mu(t)^{\otimes k}$  (equivalently, in Wasserstein distance  $W_p^{[q]}$ ) as  $N \rightarrow +\infty$ , uniformly with respect to  $t$  on any compact interval of  $[0, T_0]$ .*
- (B) *Assuming that  $G$  is locally Lipschitz with respect to  $(x, x', \xi, \xi')$  (uniformly with respect to  $t$  on any compact), defining  $S_{\bar{\mu}}^N(\tau)$  by (64) and*

$$C_{\mu}(t) = 11 \left( 1 + 70 \max_{0 \leq \tau \leq t} \text{diam}_{\Omega \times \mathbb{R}^d}(\text{supp}(\mu(t))) \right)^{1/2} \exp \left( 2t \max_{0 \leq \tau \leq t} \|G(\tau, \cdot, \cdot, \cdot, \cdot)_{|S_{\bar{\mu}}^N(\tau)^2} \|_{\mathcal{C}^{0,1}} \right), \quad (72)$$

we have, for every  $N \in \mathbb{N}^*$ , for every  $k \in \{1, \dots, N\}$  such that  $k^2 \leq 2N \ln\left(\frac{3}{2}\right)$ ,

$$W_p^{[q]}(\rho^N(t)_{N:k}^s, \mu(t)^{\otimes k}) \leq k^{1/q} C_\mu(t) \max\left(\left(\frac{k^2}{N}\right)^{1/p}, \frac{1}{N^{\frac{1}{q}-\frac{1}{2}}}, N^{1-\frac{1}{q}} \sqrt{W_1((\mu_0)_{X^N}^{se}, \mu_0)}, W_p((\mu_0)_{X^N}^{se}, \mu_0)\right) \quad (73)$$

for every  $t \in [0, T_0)$  (for  $k = 1$ , without the first term in the above parenthesis).

Theorem 4.2 is proved in Appendix B.5. Note that the  $p$ -Wasserstein distance at the left-hand side of (73) is considered with  $p \leq 2$ , because in the proof we use in an instrumental way a variance-type estimate, measuring the  $L^2$  discrepancy between the mean field and the particle vector field (see Appendix A.6). Besides,  $q \in [1, +\infty]$  is arbitrary, but only the values  $q \in [1, 2)$  are meaningful. The strongest estimate inferred from (73) is when  $q = 1$ , i.e., when one takes the  $\ell^1$  distance on  $\Omega^k \times (\mathbb{R}^d)^k$ . This choice has no importance while  $k$  is small, but becomes important if one takes for instance  $k = N^{1/4}$ .

To obtain an interesting convergence result from this theorem, we apply the second item of Lemma A.19 of Appendix A.4.2, which yields  $W_1((\mu_0)_{X^N}^{se}, \mu_0) \leq \frac{(L+1)C_\Omega}{N^r}$  and  $W_p((\mu_0)_{X^N}^{se}, \mu_0) \leq \text{diam}_{\Omega \times \mathbb{R}^d}(\text{supp}(\mu_0))^{1-1/p} ((L+1)C_\Omega/N^r)^{1/p}$  under a regularity assumption on  $\mu_0$ .

**Corollary 4.3.** *In the context of Item (B) of Theorem 4.2, we assume moreover that, for every  $N \in \mathbb{N}^*$ , there exists a tagged partition  $(\mathcal{A}^N, X^N)$  of  $\Omega$  associated with  $\nu$  satisfying (27) (see Section 1.5), and that  $x \mapsto \mu_{0,x}$  is Lipschitz for the Wasserstein distance  $W_1$ , i.e., that there exists  $L > 0$  such that  $W_1(\mu_{0,x}, \mu_{0,y}) \leq L d_\Omega(x, y)$  for  $\nu$ -almost all  $x, y \in \Omega$ . Then*

$$W_p^{[q]}(\rho^N(t)_{N:k}^s, \mu(t)^{\otimes k}) \leq k^{1/q} (L+1) C_\Omega C_\mu(t) \max\left(\left(\frac{k^2}{N}\right)^{1/p}, \frac{1}{N^{\frac{1}{q}-\frac{1}{2}}}, \frac{1}{N^{\frac{r}{2}+\frac{1}{q}-1}}, \frac{1}{N^{r/p}}\right) \quad (74)$$

for every  $t \in [0, T_0)$ .

When  $\Omega$  is a  $n$ -dimensional manifold, we have  $r = 1/n$ , hence, if we take  $q = 1$  and  $p = 1$ , the rate of convergence provided by (74) is  $\frac{k}{N^{1/2n}}$ .

Note that the assumption of a family of tagged partitions in Corollary 4.3 essentially entails that  $\nu$  be absolutely continuous with respect to a Lebesgue measure on  $\Omega$ .

**Particular case where  $G$  does not depend on  $(x, x')$ .** When  $G$  does not depend on  $(x, x')$ , we have the following corollary of Theorem 4.2 (still assuming that the norm  $\|\cdot\|$  on  $\mathbb{R}^d$  is induced by a scalar product on  $\mathbb{R}^d$ , that  $p \in [1, 2]$  and that  $q \in [1, +\infty]$ ).

**Corollary 4.4.** *Let  $\bar{\mu}_0 \in \mathcal{P}_c(\mathbb{R}^d)$  and let  $t \mapsto \bar{\mu}(t)$  be the unique solution on  $[0, T_0)$ , with  $T_0 = T_{\max}(\text{supp}(\bar{\mu}_0))$ , of the Vlasov equation (56) such that  $\bar{\mu}(0) = \bar{\mu}_0$  (see Corollary 3.1). Besides, let  $\bar{\rho}_0^N = \bar{\mu}_0^{\otimes N}$  and let  $t \mapsto \bar{\rho}^N(t) = \Phi(t, \cdot)_* \bar{\rho}_0^N$  be the unique solution on  $[0, T_0)$  of the Liouville equation (61) (without dependence on  $X$ ) such that  $\bar{\rho}^N(0) = \bar{\rho}_0^N$ . Then, for every  $N \in \mathbb{N}^*$ , for every  $k \in \{1, \dots, N\}$ , we have*

$$W_p^{[q]}(\bar{\rho}^N(t)_{N:k}, \bar{\mu}(t)^{\otimes k}) \leq k^{1/q} C_{\bar{\mu}}(t) \max\left(\left(\frac{k^2}{N}\right)^{1/p}, \frac{1}{N^{\frac{1}{q}-\frac{1}{2}}}\right) \quad (75)$$

for every  $t \in [0, T_0)$ , where  $C_{\bar{\mu}}(t)$  is defined as in (72) (without dependence on  $x$ ).

*Proof.* The proof is the same as the one of Corollary 4.2: we take  $\bar{\nu} = \delta_{\bar{x}}$  for an arbitrary  $\bar{x} \in \Omega$ . Then  $(\mu_0)_{X^N}^{se} = \delta_{\bar{x}} \otimes \bar{\mu}_0$  and thus  $W_1((\mu_0)_{X^N}^{se}, \mu_0) = W_p((\mu_0)_{X^N}^{se}, \mu_0) = 0$ . We conclude the proof by noticing that, since the particle dynamics are invariant under permutations, we have  $\bar{\rho}^N(t)_{N:k} = \bar{\rho}^N(t)_{N:k}^s$ .  $\square$



**Remark 4.2.** It is interesting to observe that, in (75), we have not taken the symmetrization of the measure  $\bar{\rho}^N(t)$  (in contrast to Corollary 4.2).

**Remark 4.3.** Applying Corollary 4.4 to the kinetic plus potential Hamiltonian case where we have  $G(t, (q_i, p_i), (q_j, p_j)) = (p_i, \nabla V(q_i - q_j))$ , we recover [36, Theorem 3.1]. The corollary can also be applied to more general Hamiltonian systems, for example,  $G(t, (q_i, p_i), (q_j, p_j)) = (p_i, -\nabla(V(q_i - q_j) + (p_i - A(q_i))^2))$ , where a magnetic field associated to a vector potential  $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ; or to Cucker–Smale systems, for which  $G(t, (q_i, p_i), (q_j, p_j)) = (p_i, F(|q_i - q_j|)(p_i - p_j))$ , and generalizations introduced in [58].

## 5 From mesoscopic to macroscopic scale: hydrodynamic limit (from Vlasov to Euler)

### 5.1 Averaged dynamical quantities defined on $\Omega$

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  $\mu$  with respect to  $\xi$  (see, e.g., [75]), leading to define, for  $\nu$ -almost every  $x \in \Omega$ :

- the total mass  $\rho(x) \geq 0$  of  $\mu_x$  by

$$\rho(x) = \int_{\mathbb{R}^d} d\mu_x(\xi) = \mu_x(\mathbb{R}^d) = 1,$$

(moment of order 0) which is here assumed to be equal to 1 for  $\nu$ -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  $\mu_x$ ;

- and the “temperature”  $T(x) \geq 0$  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, or equivalently, if  $\|\cdot\|$  is the Euclidean norm, by

$$\frac{1}{2}\rho(x)\|y(x)\|^2 + \frac{d}{2}\rho(x)T(x) = \frac{1}{2}\int_{\mathbb{R}^d} \|\xi\|^2 d\mu_x(\xi).$$

Let  $t \mapsto \mu(t)$  be a fixed solution of the Vlasov equation (45) (recall that the mean field  $\mathcal{X}[\mu]$  is defined by (44)). According to Remark 3.1, its marginal  $\nu(t) = \nu$  on  $\Omega$  does not depend on  $t$ . Following the hydrodynamic limit procedure recalled above (see also, e.g. [17, 29, 58]), for every  $t \in \mathbb{R}$  and for  $\nu$ -almost every  $x \in \Omega$ , we define the three first moments  $\rho(t, x)$ ,  $y(t, x)$  and  $T(t, x)$  of  $\mu(t)$ . The moment  $\rho(t, x)$  of order 0 does not depend on  $t$  and is equal to 1 for  $\nu$ -almost every  $x \in \Omega$  and 0 otherwise. Let us study the moments of order one and two.

## 5.2 Moment of order 1, Euler equation

Given any solution  $t \mapsto \mu(t)$  of the Vlasov equation (45) on  $[0, T]$  (for some  $T > 0$ ), of marginal  $\nu$  on  $\Omega$ , using the disintegration of  $\mu$  with respect to  $\nu$  we define

$$y(t, x) = \int_{\mathbf{R}^d} \xi d\mu_{t,x}(\xi) \quad (76)$$

for  $\nu$ -almost every  $x \in \Omega$ , and  $y(t, x) = 0$  for every  $x \in \Omega \setminus \text{supp}(\nu)$ , for every  $t \in [0, T]$ . As a preliminary remark, using (45) (or, rather, (46)), we have

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

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

$$\partial_t y(t, x) = \int_{\mathbf{R}^d} \int_{\Omega \times \mathbf{R}^d} G(t, x, x', \xi, \xi') d\mu_t(x', \xi') d\mu_{t,x}(\xi). \quad (77)$$

It is remarkable that, for some classes of functions  $G$ , and for some classes of initial data, the right-hand side of (77) can be expressed in terms of  $y(t, x)$  only: we thus obtain a “closed” equation in  $y$ , as seen next.

### 5.2.1 Linear Euler equation

**Proposition 5.1.** *Assume that  $G$  is linear with respect to  $(\xi, \xi')$ , i.e.,*

$$G(t, x, x', \xi, \xi') = a_1(t, x, x')\xi + a_2(t, x, x')\xi' \quad \forall (t, x, x', \xi, \xi') \in \mathbf{R} \times \Omega \times \Omega \times \mathbf{R}^d \times \mathbf{R}^d.$$

*For any solution  $t \mapsto \mu(t)$  of the Vlasov equation (45), the mapping  $t \mapsto y(t, \cdot)$ , where  $y(t, x)$  is defined by (76), is solution of the continuum / graph limit equation (31), in which the operator  $A$  is linear, given by*

$$(A(t)y)(x) = \int_{\Omega} a_1(t, x, x') d\nu(x') y(x) + \int_{\Omega} a_2(t, x, x') y(x') d\nu(x') \quad \forall y \in L_{\nu}^{\infty}(\Omega, \mathbf{R}^d),$$

where  $\nu$  is the marginal of  $\mu(t)$  on  $\Omega$  (not depending on  $t$ ).

*Proof.* Using the disintegration of the measure, we infer from (77) and from the specific expression of  $G$  that

$$\begin{aligned} \partial_t y(t, x) &= \underbrace{\int_{\mathbf{R}^d} \xi d\mu_{t,x}(\xi)}_{=y(t,x)} \int_{\Omega} a_1(t, x, x') \underbrace{\int_{\mathbf{R}^d} d\mu_{t,x'}(\xi') d\nu(x')}_{=1} \\ &\quad + \underbrace{\int_{\mathbf{R}^d} d\mu_{t,x}(\xi)}_{=1} \int_{\Omega} a_2(t, x, x') \underbrace{\int_{\mathbf{R}^d} \xi' d\mu_{t,x'}(\xi') d\nu(x')}_{=y(t,x')} \end{aligned}$$

and the result follows.  $\square$

**Remark 5.1.** If  $\mu(0) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N(0)} = \mu_{(X^N, \Xi_0^N)}^e$  as in Proposition 3.1, then  $\mu(t) = \mu_{(X^N, \Xi^N(t))}^e$  whose marginal on  $\Omega$  is  $\nu = \nu_{X^N}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N}$  and whose disintegration with respect to  $\nu$  is  $\mu_{t,x} = \delta_{\xi_i^N(t)}$  if  $x = x_i^N$  for  $i \in \{1, \dots, N\}$  and 0 otherwise. In this case, in the context of Proposition 5.1, we have then  $y(t, x) = \xi_i^N(t)$  if  $x = x_i^N$  for  $i \in \{1, \dots, N\}$  and 0 otherwise, and the differential equation (41) exactly coincides with the particle system (13).

Proposition 5.1 applies, for example, to the Hegselmann–Krause system: under Assumption (G), if  $t \mapsto \mu(t)$  is a solution of the Vlasov equation associated to the mean field (60) then  $t \mapsto y(t, \cdot)$ , with  $y(t, x)$  defined by (76), is a solution of the continuum / graph limit equation (41).

The result of Proposition 5.1 fails however if  $G$  is not linear with respect to  $(\xi, \xi')$ , since then we do not obtain a “closed” equation in  $y$ .

### 5.2.2 Open issue: how to obtain a closed equation?

An open question is to characterize the mappings  $G$  such that, for any solution  $\mu$  of (45), the function  $y$  defined by (76) satisfies the nonlinear Euler equation (31),  $\partial_t y(t, \cdot) = A(t, y(t, \cdot))$ . We face here with the classical problem in kinetic theory of considering the three first moments of a solution  $\mu$  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 [17] for interesting comments, see also Section 5.4 further). This is why it is usual to consider a monokinetic ansatz for  $\mu$ , as explained in the following section.

### 5.2.3 The $\nu$ -monokinetic case

In this section, we assume that  $\Omega$  is compact (for the non-compact case, see Remark 2.1). Let us consider specific solutions  $\mu$  of the Vlasov equation (45), that are  $\nu$ -monokinetic, meaning that  $\mu$  is delta-valued in the  $\xi$  variable and has the marginal  $\nu$  on  $\Omega$ . Given any  $\nu \in \mathcal{P}(\Omega)$  and any measurable function  $y : \Omega \rightarrow \mathbb{R}^d$ , we define the  $\nu$ -monokinetic measure  $\mu_y^\nu$  on  $\Omega \times \mathbb{R}^d$  by

$$\mu_y^\nu = \nu \otimes \delta_{y(\cdot)}. \quad (78)$$

We have  $y(x) = \int_\Omega \xi d(\mu_y^\nu)_x(\xi)$  for  $\nu$ -almost every  $x \in \Omega$  (as in (76)), where the disintegration of  $\mu_y^\nu$  with respect to its marginal  $\nu$  on  $\Omega$  is given by the family of conditional measures defined by  $(\mu_y^\nu)_x = \delta_{y(x)}$ .

**Proposition 5.2.** *Let  $\nu \in \mathcal{P}(\Omega)$ . Let  $T > 0$  and let  $t \mapsto y(t, \cdot) \in L^\infty_\nu(\Omega, \mathbb{R}^d)$  be a locally Lipschitz mapping on  $[0, T]$ .*

*The mapping  $t \mapsto \mu(t) = \mu_{y(t, \cdot)}^\nu \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ , of marginal  $\nu$  on  $\Omega$ , is a ( $\nu$ -monokinetic) solution on  $[0, T]$  of the Vlasov equation (45) with the general mean field (44) if and only if the mapping  $t \mapsto y(t, \cdot) \in L^\infty_\nu(\Omega, \mathbb{R}^d)$  is a solution on  $[0, T]$  of the (nonlinear) Euler equation (31).*

*Proof.* When  $\mu_t = \mu_{y(t, \cdot)}^\nu$ , (44) gives  $\mathcal{X}[\mu_t](t, x, \xi) = \int_\Omega G(t, x, x', \xi, y(t, x')) d\nu(x')$ . The proof is then straightforward, and we can note that  $A(t, y)(x) = \mathcal{X}[\mu_y^\nu](t, x, y(x))$  (where the nonlinear operator  $A$  is defined by (30)).  $\square$

**Remark 5.2.** Proposition 5.2 implies Theorem 2.1 (existence and uniqueness for the Euler equation (31)). Indeed, assume that  $\Omega$  is compact, let  $\nu \in \mathcal{P}(\Omega)$ , let  $y^0 \in L^\infty_\nu(\Omega, \mathbb{R}^d)$ , set  $K' = \text{ess.im}(y^0)$  its essential range (compact subset of  $\mathbb{R}^d$ ) and  $K = \Omega \times K'$  (compact). Since the unique solution of the Vlasov equation (45) such that  $\mu(0) = \mu_{y^0}^\nu = \nu \otimes \delta_{y^0(\cdot)}$  is well defined on  $[0, T_{\max}(K))$  (by Theorem 3.1) and is given by  $\mu(t) = \mu_{y(t, \cdot)}^\nu$  (by Proposition 5.2), it follows that the nonlinear Euler equation (31) has a unique solution on  $[0, T_{\max}(K))$  such that  $y(0, \cdot) = y^0(\cdot)$ .

When  $\mu_t$  is not of the form  $\mu_{y(t, \cdot)}^\nu$ ,  $t \mapsto y(t, \cdot)$  fails in general to satisfy a “closed” equation (i.e.,  $\partial_t y(t, \cdot)$  may not be expressible only in function of the first moment  $y(t, \cdot)$ ). Instead, there may be a full hierarchy of equations coupling all moments of  $\mu_{t,x}$  (see Section 5.4). Anyway, when convergence to consensus holds, we may expect that any solution  $\mu$  of (45) is asymptotically of the form  $\mu_{y(t, \cdot)}^\nu$ .

**Remark 5.3.** In Remark 2.3 in Section 2.1, we have seen how to embed the solutions of the particle system to solutions of the Euler equation by considering an empirical measure  $\nu$ . This embedding works because, when  $\nu = \nu_{X^N}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N}$  and  $y(t, x_i^N) = \xi_i^N(t)$ , the  $\nu$ -monokinetic measure  $\mu_{y(t, \cdot)}^\nu$  coincides with the empirical measure  $\mu_{(X^N, \Xi^N(t))}^e$ . Indeed,

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

**Remark 5.4.** In Appendix A.5, we provide estimates on the discrepancy between empirical measures and  $\nu$ -monokinetic measures. Lemma A.20 of that appendix, combined with Theorem 2.2 and with the proof of that theorem, yields estimates on the discrepancy of the empirical measure  $\mu_{(X^N, \Xi^N(t))}^e$  with respect to the  $\nu$ -monokinetic measures  $\mu_{y(t, \cdot)}^\nu$  or  $\mu_{y^N(t, \cdot)}^\nu$ .

**Remark 5.5.** The proofs of Theorems 2.2 and 2.3 that we provide in Appendices B.1 and B.2 are direct, but actually one can also prove these propositions by applying Corollary 4.1 with  $\mu(t) = \mu_{y(t, \cdot)}^\nu = \nu \otimes \delta_{y(t, \cdot)}$  (the  $\nu$ -monokinetic measure) and use Lemma A.20 of Appendix A.5.

### 5.3 Moment of order 2

In this section, we assume that the norm  $\|\cdot\|$  on  $\mathbb{R}^d$  is induced by a scalar product  $\langle \cdot, \cdot \rangle_{\mathbb{R}^d}$  on  $\mathbb{R}^d$ . We define

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

for  $\nu$ -almost every  $x \in \Omega$ . Note that  $T(t, x) = 0$  for  $\nu$ -almost every  $x \in \Omega \setminus \text{supp}(\nu)$ . Using (45) (or, rather, (46)) and noting that  $\int_{\mathbb{R}^d} \langle \xi - y(t, x), \partial_t y(t, x) \rangle_{\mathbb{R}^d} d\mu_{t, x}(\xi) = 0$ , we compute

$$\partial_t T(t, x) = \frac{2}{d} \int_{\mathbb{R}^d} \langle \xi - y(t, x), \mathcal{X}[\mu_t](t, x, \xi) \rangle_{\mathbb{R}^d} d\mu_{t, x}(\xi). \quad (79)$$

**Proposition 5.3.** *In the Hegselmann–Krause model (13), we have  $G(t, x, x', \xi, \xi') = \sigma(x, x')(\xi' - \xi)$  and*

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

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

*Proof.* We have  $\mathcal{X}[\mu_t](t, x, \xi) = \int_{\Omega} \int_{\mathbb{R}^d} \sigma(x, x')(\xi' - \xi) d\mu_{t, x'}(\xi') d\nu(x')$  in the Hegselmann–Krause model (see (60)), and thus

$$\mathcal{X}[\mu_t](t, x, \xi) = -S(x) \int_{\mathbb{R}^d} (\xi - y(t, x)) d\mu_{t, x'}(\xi') + \int_{\Omega} \int_{\mathbb{R}^d} \sigma(x, x')(\xi' - y(t, x)) d\mu_{t, x'}(\xi') d\nu(x').$$

Since the second term does not depend on  $\xi$ , using again the fact that  $\int_{\mathbb{R}^d} (\xi - y(t, x)) d\mu_{t, x}(\xi) = 0$  by definition, and using (79), the result follows.  $\square$

**Remark 5.6.** We will see in Remark 5.7 in Section 5.4 that, in the Hegselmann–Krause model, all moments of order  $\geq 2$  satisfy the same differential equation, and thus, decrease exponentially to 0 as  $t \rightarrow +\infty$  as soon as  $S(x) > 0$  for  $\nu$ -almost every  $x \in \Omega$ . This shows that, under the latter assumption, the solution  $t \mapsto \mu(t)$  of the Vlasov equation (45) is such that  $\mu_{t, x}$  is exponentially close (in Wasserstein distance) to the Dirac measure  $\delta_{y(t, x)}$  as  $t \rightarrow +\infty$ .

In [14], convergence to consensus is proved for the Euler equation under the assumptions that  $d\nu(x) = dx$ , that  $S(x) \geq \delta > 0$  for almost every  $x \in \Omega$  and that the (infinite-dimensional) graph associated with  $\sigma$  be strongly connected. This remark shows that the result of [14] can be generalized by relaxing the assumption on  $S$  to:  $S(x) > 0$  for  $\nu$ -almost every  $x \in \Omega$ .

For general mappings  $G$ , the question of whether or not  $T$  is the solution of some ‘‘closed’’ equation is open.

In the  $\nu$ -monokinetic case, i.e., assuming that  $\mu$  is of the form (78) and is a solution of (45), we have  $T(t, x) = 0$ . This is expected since  $T(t, x)$  is the variance and thus measures the distance to the average  $y(t, x)$ .

## 5.4 Generalization: coupled equations of moments

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

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

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

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

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

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

and thus, using (30),

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

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

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

(actually since  $y_1 = 0$  the above sum can be taken over all pairs  $(i, j)$  such that  $i + j \geq 2$ ) and, for

every  $k \in \mathbb{N} \setminus \{0, 1\}$ ,

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

(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.

Closing the hierarchy of equations satisfied by all the moments  $y_i(t, x)$ , for  $i \in \mathbb{N}^*$ , might be done by adding a small parameter  $\varepsilon$ . This is an open question.

**Remark 5.7.** In the Hegselmann–Krause model (13), we have  $G(t, x, x', \xi, \xi') = \sigma(x, x')(\xi' - \xi)$  and thus  $g_{ij} = 0$  if  $i + j \geq 2$  and  $g_{01} = -g_{10} = \sigma$ . We recover the facts that the equation in  $y$  is closed and that  $\partial_t y_2(t, x) = -2S(x)y_2(t, x)$ . Moreover, a straightforward computation shows that

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

thus generalizing the case  $k = 2$  studied in Proposition 5.3. Therefore,  $y_k(t, x) = y_k(0, x)e^{-tS(x)}$ .

## 5.5 Directly from Liouville to Euler

In this section we answer the question : *which “Liouville” quantity the first moment of the solution to Vlasov (i.e., the solution of Euler in the good cases) is the large  $N$  limit of the symmetrized first marginal of?*

Ler us motivate this question by reminding that, in accordance with Section 4.2, the true “vector” for passing from Liouville to Vlasov is the process of taking marginals: Vlasov appears as a subsystem of the large  $N$  limit of Liouville by keeping only the information encapsulated in the marginals of the solutions of Liouville. The task of the present section is to do the same for Euler.

In Section 2.2 we considered the direct passage from the particle system (8) to the Euler (continuum / graph limit , graph limit) equation (31) through the system of ODEs defining the particle dynamics whereas in the present section we also reached the same Euler equation but via the Vlasov equation in the pure mean field paradigm, using the results of Scions 3 and 4.

The Liouville equation (61) being the transport equation lifting the particle system (8), 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 (61), converging to the solution of the Euler equation as  $N \rightarrow +\infty$ . 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 (61) 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  $\Xi = (\xi_1, \dots, \xi_N) \in \mathbb{R}^{dN}$ ,  $\rho(X, \Xi)$  is the joint probability that, for every  $i \in \{1, \dots, N\}$ , the  $i^{\text{th}}$  particle has position and momentum  $(x_i, \xi_i)$ . In Section 4.2 we have

shown that, for appropriate initial conditions  $\rho(0)$ , we recover the mean field limit by taking the limit  $N \rightarrow +\infty$  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 (or at least to the velocity field of the solution of the Vlasov equation) and answer the aforementioned question: which  $N$  particles system’s quantity is the solution to the graph limit equation the large  $N$  limit of the first marginal of?

Given any  $X^N = (x_1^N, \dots, x_N^N) \in \Omega^N$ , we set

$$\rho(X, \Xi) = \prod_{i=1}^n \delta_{x_i^N}(x_i) \mu(x_i, \xi_i)$$

and, for any  $i \in \{1, \dots, N\}$ ,

$$\mathcal{M}_1^i[\rho] = \int_{\mathbf{R}^d} \xi_i \rho(X, \Xi) d\xi_i \quad (80)$$

defined as a marginal against test functions on  $\Omega^N \times \mathbf{R}^{d(N-1)}$ , i.e.,

$$\langle \mathcal{M}_1^i[\rho], \varphi \rangle = \int_{\Omega^N \times \mathbf{R}^{d(N-1)}} \varphi(X, \Xi_{-i}) \xi_i \rho(X, \Xi) dX d\xi$$

where  $\Xi_{-i} = (\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_N)$ . Given any  $\mu \in \mathcal{P}(\Omega \times \mathbf{R}^d)$ , we set

$$m_1[\mu](x) = \int_{\mathbf{R}^d} \xi d\mu_x(x, \xi).$$

**Proposition 5.4.** *Under the framework of Corollary 4.3 and uniformly with respect to  $i \in \{1, \dots, N\}$ , we have*

$$W_1 \left( (\mathcal{M}_1^i[\rho^t])_{N:1}^s, m_1[\mu^t] \right) = o(N^{-1})$$

where  $o(N^{-1})$  is given by the right-hand side of (74) for  $k = p = q = 1$ . Here  $\rho^t$  is the pushforward of  $\rho^{t=0} = \rho$  under the particle flow, and  $\mu^t$  solves Vlasov with initial condition  $\mu^{t=0} = \mu$ .

**Corollary 5.1** (monokinetic case). *When  $\mu(x, \xi) = \delta(\xi - y(x))$ ,*

$$W_1 \left( (\mathcal{M}_1^i[\rho^t])_{N:1}^s, y^t \right) = o(N^{-1})$$

(same as in Proposition 5.4), where  $y^t$  solves Euler with initial condition  $y^{t=0} = y$ .

It is also true for general  $\rho$  for the Hegselmann Krause model, with  $y(x) = \int \xi d\mu_x(x, \xi)$ .

In other words, the answer to the question asked in this section is any “first moment” of the Liouville solution as defined by (80).

*Proof.* The fact that the proposition and the corollary are valid uniformly with respect to  $i \in \{1, \dots, N\}$  comes from the fact that symmetrization commutes with the action of taking first-order marginals (only first-order). Therefore  $(\mathcal{M}_1^i[\rho^t])_{N:1}^s = m_1 [([\rho^t]_{N:1}^s)]$  and the proposition easily follows from Lemma A.11. The proof of the corollary is immediate.  $\square$

## 5.6 Hydrodynamic limit in the second-order case

For second-order particle systems (16) where  $q$  is interpreted as a position and  $p$  as a speed (or a momentum), in the classical kinetic literature where particles are assumed to be indistinguishable, the hydrodynamic quantities that are most often considered are the three first moments of  $\mu$  integrated with respect to  $p$  and kept as functions of  $q$ .

Let us recall that the “hydrodynamics” replaces a scalar equation on “phase space” (the one of  $q$  and  $p$ ), the Vlasov one, by a system of equations on the “configuration space” (the one of  $q$  only). The interest of this approach is twofold: firstly at the conceptual level, as it puts the dynamics on a physical, directly observable space (a feature particularly important for situations not naturally embedded in the classical physics cultural paradigm (for example biology, economy, social sciences), and secondly at the numerical point of view since the increasing number of variables for PDEs is very costly (see [60] for a comparison between numerics associated to Vlasov and Euler).

More precisely, in the second-order case, any solution  $t \mapsto \mu(t)$  of the Vlasov equation (not depending on  $x$ ) is such that  $\mu(t) \in \mathcal{P}_c(\mathbb{R}^r \times \mathbb{R}^r)$ , where  $d = 2r$  and  $\xi = (q, p)$ , and, assuming that  $\frac{d\mu_t(q,p)}{dq dp} = f(t, q, p)$ , the three first (marginal) moments of  $\mu(t)$  under consideration are:

- the mass  $m(q) = \int_{\mathbb{R}^r} f(t, q, p) dp$ ,
- the momentum  $m(q)v(q) = \int_{\mathbb{R}^r} p f(t, q, p) dp$ ,
- the energy (or temperature)  $m(q)E(q) = \int_{\mathbb{R}^r} \|p - v(q)\|^2 f(t, q, p) dp$ ,

as defined in [75]. This is different from what we did in Section 5.1.

The above quantities are introduced for example in [38] in the Cucker–Smale model. In general, they provide alternative objects of investigation, in the specific case of second-order models. But let us notice that they do not satisfy a system of closed equations; the equations for the mass and the momentum close themselves under a monokinetic condition on the initial condition of the Vlasov equation.

Let us assume that  $d = 2r$ ,  $\xi = (q, p) \in \mathbb{R}^r \times \mathbb{R}^r$  and that the system of particles is an indistinguishable second-order system of the form

$$\dot{q}_i(t) = p_i(t), \quad \dot{p}_i(t) = \frac{1}{N} \sum_{j=1}^N K(t, q_i(t), p_i(t), q_j(t), p_j(t))$$

for some continuous mapping  $K$  of class  $C^1$  with respect to its four last variables. Here,  $q_i(t)$  is the position and  $p_i(t)$  is the speed of the agent number  $i$ . Assumption (G) in Section 1.2 is satisfied with  $G(t, x, x', \xi, \xi') = (p, K(t, \xi, \xi'))$ . Since  $G$  does not depend on  $(x, x')$ , particles are indistinguishable. When considering the limit equations (Euler or Vlasov), we can however choose to distinguish them, as already discussed.

**Euler equation.** Choosing a set  $\Omega$  of labels and a probability measure  $\nu$  on  $\Omega$ , the Euler equation (31) is

$$\begin{aligned} \partial_t y_1(t, x) &= y_2(t, x) \\ \partial_t y_2(t, x) &= \int_{\Omega} K(t, y_1(t, x), y_2(t, x), y_1(t, x'), y_2(t, x')) d\nu(x') \end{aligned} \tag{81}$$

where it is understood that  $y_1(t, x_i) \simeq q_i(t)$  and  $y_2(t, x_i) \simeq p_i(t)$  (see Section 2.2).



**Vlasov equation.** As we have seen before in (55) and (56), since in the present case the mean field  $\mathcal{X}[\mu]$  does not depend on the variable  $x$ , we have  $\mathcal{X}[\mu](t, x, \xi) = \bar{\mathcal{X}}[\bar{\mu}](t, \xi)$  where the mean field  $\bar{\mathcal{X}}[\bar{\mu}]$  is given by

$$\bar{\mathcal{X}}[\bar{\mu}](t, \xi) = \left( \int_{\mathbb{R}^r \times \mathbb{R}^r} K(t, \xi, \xi') d\bar{\mu}(\xi') \right)^p$$

and the Vlasov equation is  $\partial_t \bar{\mu} + \operatorname{div}(\bar{\mathcal{X}}[\bar{\mu}]\bar{\mu}) = 0$ .

**Usual hydrodynamic limit considered in the literature: the *standard* Euler equation.**

We are going to consider the hydrodynamic variables in the sense that is usually considered in the literature: moments of the measure  $\bar{\mu}$  where the integration is performed with respect to  $p$ , but keeping  $q$  as a parameter.

(1) If we proceed by disintegration, we do not obtain something much interesting. Let us explain. Given any probability measure  $\bar{\mu}$  on  $\mathbb{R}^r \times \mathbb{R}^r$  (where the variables are  $(q, p)$ ), let  $\theta$  be the marginal of  $\bar{\mu}$  on the first copy of  $\mathbb{R}^r$  (where the variable is  $q$ ). The disintegration of  $\bar{\mu}$  with respect to  $\theta$  is  $\bar{\mu} = \int_{\mathbb{R}^r} \mu_q d\theta(q)$  where the measures  $\mu_q$  are probability measures. Then, if we take the first moments of  $\mu$ : the moment of order 0 is  $\rho(q) = \int_{\mathbb{R}^r} d\mu_q(p) = 1$  and thus is uninteresting; we could then consider the moments of order 1 and 2. But let us proceed in another way, following what is usually done in the literature:

(2) In the usual way, we assume that, for every  $t$ ,  $\bar{\mu}(t)$  (solution of Vlasov) is absolutely continuous, i.e.,  $d\bar{\mu}_t(q, p) = f(t, q, p) dq dp$ , and we define

- the moment of order 0:  $\rho(t, q) = \int_{\mathbb{R}^r} f(t, q, p) dp$ ;
- the moment of order 1:  $\rho(t, q)u(t, q) = \int_{\mathbb{R}^r} p f(t, q, p) dp$ .

This is different from what was done in (1) because now  $\rho(t, q)$  is not constant. Actually, comparing with the disintegration of  $\bar{\mu}_t$ , we have  $\frac{d\bar{\mu}_{t,q}}{dp}(p) = f(t, q, p) / \int_{\mathbb{R}^r} f(t, q, p') dp'$ . So, here, the way we consider the hydrodynamic variables is different.

Let us recall how to compute the equations satisfied by  $\rho$  and  $u$  (this is well known in the existing literature). The Vlasov equation for  $f$  is

$$\partial_t f + \langle p, \nabla_q f \rangle + \operatorname{div}_p(\mathcal{X}_K[f]f) = 0$$

where  $\mathcal{X}_K[f](t, q, p) = \int_{\mathbb{R}^r} \int_{\mathbb{R}^r} K(t, q, p, q', p') f(t, q', p') dq' dp'$ . Multiplying by  $\varphi_1(q)\varphi_2(p)$  and integrating, we get

$$\begin{aligned} & \int_{\mathbb{R}^r} \int_{\mathbb{R}^r} \varphi_1(q)\varphi_2(p)\partial_t f(t, q, p) dq dp \\ &= \int_{\mathbb{R}^r} \int_{\mathbb{R}^r} \left( \varphi_2(p) d\varphi_1(q) \cdot p + \varphi_1(q) d\varphi_2(p) \cdot \mathcal{X}_K[f](t, q, p) \right) f(t, q, p) dq dp \end{aligned}$$

First, taking  $\varphi_2(p) = 1$ , replacing above and integrating by parts, and noting that the equation is valid for any function  $\varphi_1$  (thus, one can localize in  $q$ ), we get

$$\boxed{\partial_t \rho + \operatorname{div}_q(\rho u) = 0}$$

which is the classical continuity equation. It is obtained without any specific assumption, in contrast to the next one.

Taking  $\varphi_2 = p_i$  for any  $i \in \{1, \dots, r\}$  (where  $p = (p_1, \dots, p_r)$ ), replacing above and integrating by parts does not suffice in producing a “closed” equation. As done usually, we assume that the

velocity distribution is monokinetic:  $f(t, q, p) = \rho(t, q)\delta(p - u(t, q))$ . In this way we obtain the equation

$$\partial_t(\rho u) + \operatorname{div}_q(\rho u \otimes u) = \int_{\mathbf{R}^r} b(t, q, u(t, q), q', u(t, q')) \rho(t, q) \rho(t, q') dq'$$

which is the pressureless Euler equation, as obtained by [21, 29] and [59] for more general systems including chemotaxis. Although we have used the same name, this Euler equation (and the way it has been obtained) is of course completely different from the one given in (81).

## 6 Summary: relationships between various scales

(Euler and continuum / graph limit equations are the same)

In the previous sections, we have investigated the following three scales (recall that  $G$  satisfies Assumption **(G)**):

- The microscopic model, which is the particle system

$$\dot{\xi}_i^N(t) = \frac{1}{N} \sum_{j=1}^N G(t, x_i^N, x_j^N, \xi_i^N(t), \xi_j^N(t)), \quad i \in \{1, \dots, N\}. \quad (82)$$

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

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

$$\partial_t \mu + \operatorname{div}_\xi(\mathcal{X}[\mu]\mu) = 0 \quad (83)$$

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

- the macroscopic model, which is the Euler equation

$$\partial_t y(t, x) = (A(t, y(t)))(x) = \int_{\Omega} G(t, x, x', y(t, x), y(t, x')) d\nu(x') \quad (84)$$

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

Additionally, we have also considered the Liouville equation,

$$\partial_t \rho^N + \operatorname{div}_\Xi(Y^N \rho^N) = 0 \quad (85)$$

where  $Y^N$  is the vector field in  $\mathbf{R}^{dN}$  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  $\Xi^N(\cdot)$  of the particle system (82) can be embedded as a Dirac measure  $\rho^N(\cdot) = \delta_{X^N} \otimes \delta_{\Xi^N(\cdot)}$  that is a solution of the Liouville equation (85).

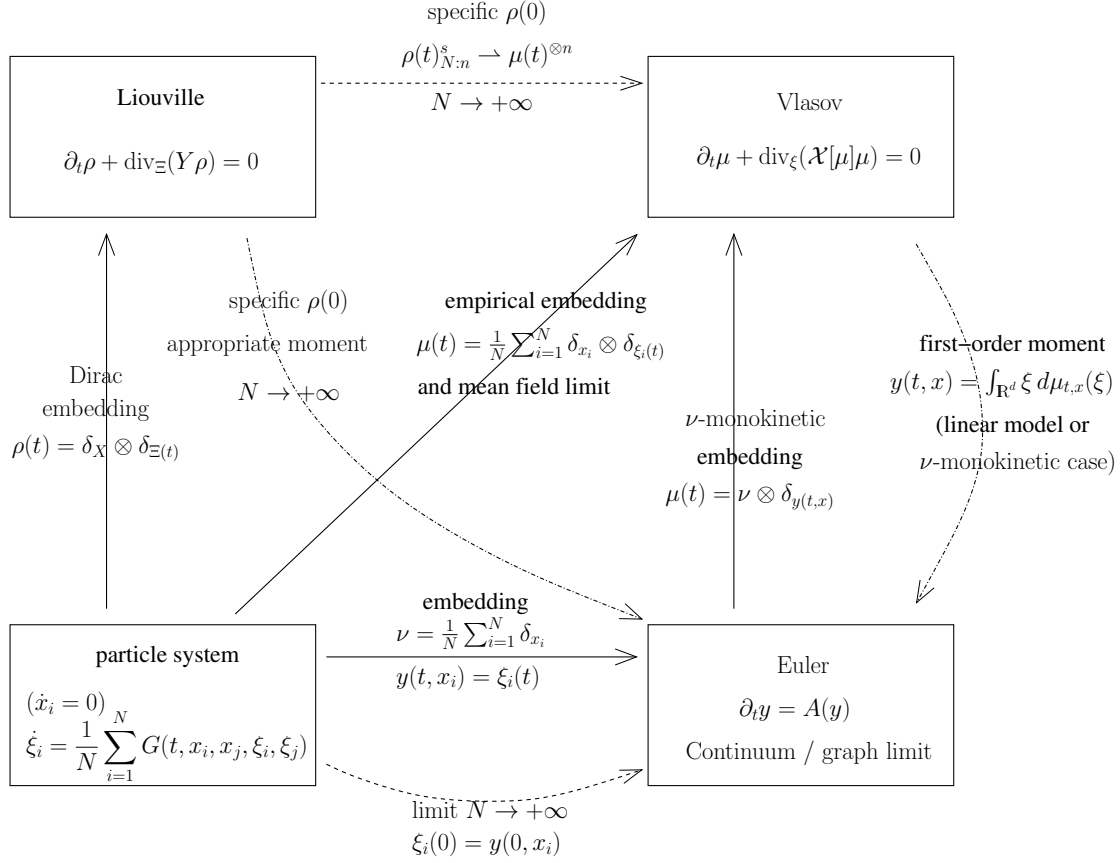


Figure 1: Relationships between particle (microscopic) system, Liouville (probabilistic) equation, Vlasov (mesoscopic, mean field) equation, Euler (macroscopic, graph limit) equation. We do not write the upperscript  $N$  in the various formulas to keep a better readability.

**Particle to Vlasov.** By Proposition 3.1, any solution  $\Xi^N(\cdot)$  of the particle system (82) can be embedded to an empirical measure  $\mu(\cdot) = \mu_{(X^N, \Xi^N(\cdot))}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N(\cdot)}$  that is a solution of the Vlasov equation (83). Conversely if an empirical measure  $\mu(\cdot) = \mu_{(X^N, \Xi^N(\cdot))}^e$  (with distinct points) is a solution of the Vlasov equation (83) then  $\Xi^N(\cdot)$  must be a solution of (82).

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

**Particle to Euler.** Any solution  $\Xi^N(\cdot)$  of the particle system (82) can be embedded to a solution of the general nonlinear Euler equation (84) by using an empirical measure  $\nu$  (see Remark 2.3).

Alternatively and much more interestingly, to pass from the microscopic to the macroscopic scale, by Theorems 2.2 and 2.3, one can take the graph limit of the particle system (Riemann sum theorem) and thus obtain the Euler equation, with estimates of convergence as  $N \rightarrow +\infty$ .

**Liouville to Vlasov.** By Theorems 4.1 or 4.2, one can recover the solutions of the Vlasov equation (83) from those of the Liouville equation (85), for some appropriate initial conditions  $\rho(0)$ , by taking marginals and taking the limit  $N \rightarrow +\infty$ .

**Euler to Vlasov.** By Proposition 5.2 in Section 5.2.3, given any  $\nu \in \mathcal{P}(\Omega)$  and any solution  $t \mapsto y(t, \cdot)$  of the Euler equation (84), the  $\nu$ -monokinetic measure mapping  $t \mapsto \mu(t) = \mu_{y(t, \cdot)}^\nu = \nu \otimes \delta_{y(t, \cdot)}$  defined by (78) is a solution of the Vlasov equation (83). This embedding from the macroscopic to the mesoscopic scale is general and is valid for the mean field  $\mathcal{X}[\mu]$  defined by (44) and for the nonlinear operator  $A$  defined by (30).

**Vlasov equation to Euler equation.** Here, and only here, we assume, first, that  $G$  is linear with respect to  $(\xi, \xi')$  (as it is the case for the Hegselmann–Krause model). Proposition 5.1 says that, given any solution  $t \mapsto \mu(t)$  of the Vlasov equation (83), defining  $\nu = \pi_* \mu(t)$  (marginal of  $\mu(t)$ , which does not depend on  $t$ ), the moment mapping  $t \mapsto y(t, \cdot)$  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 (84) (which is linear in this case).

As discussed in Section 5.2.3, there is a second way, still not general, to go from Vlasov to Euler, by assuming that the solution  $\mu(\cdot)$  of the Vlasov equation is  $\nu$ -monokinetic. In this case, its moment  $y$  of order 1 is solution of the nonlinear Euler equation (31).

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

**Liouville equation to Euler equation.** Proposition 5.4 and its Corollary in Section 5.5 show how to derive Euler from Liouville, for specific initial conditions  $\rho^N(0)$ , by taking an adequate moment of  $\rho^N(t)$  and then passing to the limit  $N \rightarrow +\infty$ .

Finally, all above 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 if  $G$  is linear with respect to  $(\xi, \xi')$  but fails in general. The graph limit procedure is of a different nature and relies on the Riemann sum theorem (see Section 2).

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.

In conclusion let us emphasize the following important novelty of our article:

**a Vlasov-mesoscopic scale equation for agent systems**  
and the fact that  
**the (continuum) graph limit equation is an Euler equation**

## 7 Further comments and perspectives

In this paper we have surveyed and generalized various ways to pass to the limit in finite systems of particles, and described precise relationships between the various limit equations: Euler, Vlasov and Liouville equations. This has been done under the standing assumption **(G)** on the interaction mapping  $G$  modeling the particle dynamics.

As already said, we have restricted our study to regular mappings  $G$ , because our objective was to highlight in the simplest possible way the basic relationships between the microscopic, mesoscopic and macroscopic scales, with the most possible general viewpoint. The study of singular kernels is much more challenging and requires the development of other techniques.

Related to this issue, we give in the section hereafter a surprising consequence of our results.

### 7.1 Approximation of PDEs by finite particle systems

As stated in the longer preprint version [62] of the present article, and as already presented at a number of conferences and talks since 2022, there is a surprising consequence of our analysis,

namely, that sufficiently regular solutions of any quasilinear PDE can be approximated by solutions of systems of  $N$  particles, to within  $1/\ln \ln(N)$ .

This intriguing result is stated in [62] but since this preprint was becoming too long we decided to split it and to present the PDE application in a separate paper.

In this section, we give a flavor of this result. To simplify the setting, we restrict our exposition to the linear case. Let us consider a linear evolution equation

$$\dot{y}(t) = Ay(t) \tag{86}$$

where  $A : D(A) \rightarrow L^2(\Omega, \mathbb{R}^d)$  is a linear operator generating a  $C_0$  semigroup and  $\Omega$  is an open bounded subset of  $\mathbb{R}^n$ . Let  $[A]$  be the (distributional) Schwartz kernel of  $A$ , so that  $(Af)(x) = \langle [A](x, \cdot), f \rangle$  in the distributional sense, for every  $f \in \mathcal{C}_c^\infty(\Omega, \mathbb{R}^d)$ . For instance if  $[A](x, x') = \delta'_x$ , the distributional derivative of the Dirac measure  $\delta_x$  at  $x$ , then  $A = -\partial_x$ .

Writing, with a slight abuse of notation,  $(Af)(x) = \int_\Omega [A](x, x')f(x')$ , we see a strong similarity with the bounded operator given in (41) (in Section 2.3), of kernel  $\sigma$ . The idea then consists in approximating the distribution  $[A]$  with a family of smooth kernels  $\sigma_\varepsilon$ , depending on a new parameter  $\varepsilon > 0$ , such that  $\sigma_\varepsilon$  converges in the distributional sense to  $[A]$  as  $\varepsilon \rightarrow 0$ .

There are a number of ways to perform such an approximation. Explicit constructions are provided in [62] by using convolutions. In all cases, we now have a family of bounded operators  $A_\varepsilon$  on  $L^2(\Omega)$ , such that  $(A_\varepsilon f)(x) = \int_\Omega \sigma_\varepsilon(x, x')f(x') dx'$  for every  $f \in \mathcal{C}_c^\infty(\Omega, \mathbb{R}^d)$ . Following Section 2.1, for every  $\varepsilon > 0$  fixed, the operator  $A_\varepsilon$  is of the form (30) with the mapping  $G_\varepsilon$  given by  $G_\varepsilon(t, x, x', \xi, \xi') = \sigma_\varepsilon(x, x')\xi'$ , and the corresponding Euler equation (31), which is here the linear evolution equation  $\dot{y}_\varepsilon(t) = A_\varepsilon y_\varepsilon(t)$ , is viewed as an approximation of the linear evolution equation (86). Actually, it is not difficult to stand general assumptions under which  $\|y_\varepsilon(t) - y(t)\|_{L^2} \leq C e^{\beta t}$  for some  $C > 0$  and  $\beta \in \mathbb{R}$  not depending on the solutions.

Now, in a second step, following the path drawn in this paper and passing from the macroscopic to the microscopic scale, the family of finite particle systems, indexed by  $N$ , naturally associated to the  $\varepsilon$ -Euler equation  $\dot{y}_\varepsilon(t) = A_\varepsilon y_\varepsilon(t)$  is

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

and the discrepancy between their respective solutions can be estimated thanks to Theorem 2.2 in Section 2.2 (in particular, (37)). Under appropriate assumptions, or by an explicit construction of the kernel  $\sigma_\varepsilon$  done in [62], it happens that the Lipschitz constants of  $G_\varepsilon$  and of  $\sigma_\varepsilon$  are estimated by integer powers of  $\frac{1}{\varepsilon}$  so that, finally, by the triangular inequality, we obtain an estimate of the form

$$\left\| y(t, \cdot) - \sum_{i=1}^N \xi_{\varepsilon,i}^N(t) \mathbb{1}_{\Omega_i^N}(\cdot) \right\|_{L^2(\Omega, \mathbb{R}^d)} \leq C \left( \varepsilon + \frac{1}{N} \exp \left( \frac{C}{\varepsilon} \exp \left( \frac{C}{\varepsilon} \right) \right) \right).$$

Actually, there are powers of  $N$  and of  $\varepsilon$  but we drop them to simplify. Of course, in the above estimate, if  $N$  is fixed and  $\varepsilon \rightarrow 0$  then the estimate blows up, which is expected because of the unboundedness of  $A$ . It is thus required to let  $N$  tend to infinity and  $\varepsilon$  to zero in an appropriate way. Optimizing the above estimate leads to choose  $\varepsilon \sim \frac{1}{\ln \ln N}$ , and then

$$\left\| y(t, \cdot) - \sum_{i=1}^N \xi_{\varepsilon,i}^N(t) \mathbb{1}_{\Omega_i^N}(\cdot) \right\|_{L^2(\Omega, \mathbb{R}^d)} \leq \frac{C}{\ln \ln N}$$

(actually, a positive power of it). We have thus shown that we can approximate any linear (and actually, quasilinear) PDE by an explicit family of finite particle systems, to within  $1/\ln \ln(N)$ .

It is intriguing that this result is universal and requires very few assumptions (see [62] for more details and complete statements). However, coming back to the basics of statistical physics, one should not forget that a given domain  $\Omega$ , say of volume 1, contains  $N \simeq 6.10^{23}$  particles and that  $\ln \ln(N) \simeq 4$ , thus making irrelevant the general obtained estimate. This shows that, in some sense, the estimate  $1/\ln \ln(N)$  is a kind of physical barrier. But we refer the reader to the last part of [62], soon becoming a separate article, for more results and comments on this issue.

We conclude the present paper with a comment. The above discovery has been made possible only by adding the variable  $x \in \Omega$  to the particle system, to the Vlasov equation, yielding the Euler equation obtained by graph limit. This additional variable  $x$ , which stands for the label of a given particle (and is used to distinguish particles one from each other), becomes naturally here, in the context of PDEs, the spatial variable. We refer to [63] for further developments.

## 7.2 Some open questions

We provide hereafter several further comments and open issues.

**Closing the hierarchy of moments.** In Section 5, we have defined the moments of the measure  $\mu$  solution of the Vlasov equation. We have seen that the hierarchy of moments is closed if  $G$  is linear with respect to  $(\xi, \xi')$  but is not closed in general otherwise. Although Proposition 5.1 remains an obvious observation, it is still mysterious to us. The question is open to characterize all mappings  $G$  so that the hierarchy is closed at a certain level.

In cases where the hierarchy is not closed (like for fluid equations), we wonder whether it is possible to add a small parameter  $\varepsilon$  which would be used to close the hierarchy by taking adequate limits (see [35] for similar comments).

**Convergence to consensus.** We have shown in Section 5.3 that, surprisingly, the “temperature” always decreases exponentially, pointwisely, as soon as  $S(x) > 0$ , for the Hegselmann–Krause model. This fact allows one to easily recover (and improve) some known results on convergence to consensus. We do not know to what extent this observation may be generalized but we think that it can be used to derive consensus results under weaker assumptions.

In more general, for nonlinear systems enjoying consensus properties (like nonlinear Hegselmann–Krause models) or synchronization properties (like the Kuramoto model), following Section 5.2.3, we expect that, under appropriate assumptions, any solution  $t \mapsto \mu(t)$  of the Vlasov equation (45) is asymptotically of the form  $\mu_{y(t, \cdot)}^\nu$  where  $t \mapsto y(t, \cdot)$  is a solution of the Euler equation (31).

In any case, establishing a consensus result at the level of the Vlasov equation is interesting because it should a priori imply consensus at the level of the (macroscopic) Euler equation and for the (microscopic) particle system.

**Improving error estimates.** In our results, we establish error estimates between solutions of the particle system and a limit equation (Euler or Vlasov) on compact intervals of time, and the error grows exponentially in time, essentially due to a Gronwall argument. Such errors can certainly be much improved for some classes of mappings  $G$ , maybe under consensus convergence properties, in order to obtain uniform in time estimates.

**Use of the measure  $\nu$ .** In the existing literature, the measure  $\nu$  used in the definition (30) of the operator  $A$  of the Euler equation (31) (graph limit) is always the Lebesgue measure. We have shown in this paper the interest of considering other measures, in particular empirical measures, to obtain (trivial) relationships with the particle system. But more generally, it is certainly of interest to use other measures  $\nu$ , depending on the context. For example, in social sciences, each

agent could have a probability of decision, but several agents could have the same probability of opinion.

**Weakly regular (but not singular) mapping  $G$ .** It is likely that, in our results and in particular in Theorem 3.1, we can weaken the continuity assumption on  $G$ , provided we consider the solutions of the Vlasov equation in a weaker sense. This is what is done in [44] for some classes of opinion propagation models: the authors do not assume that there exists a limit mapping  $G$  (as we do in (G)) but to take the mean field limit, in a weaker sense, they make another assumption of uniform boundedness on their dynamics. However, at the limit they lose the distinguishability of the particles.

It can be noted that when  $G$  is weakly regular (for instance  $L^\infty$ ), we do not have, a priori, an existence and uniqueness result for the particle system. But, following [45], we can study the Liouville equation (61), for which we can have existence (but not uniqueness) for rough vector fields, and then derive the Vlasov equation by taking marginals.

**Multiplewise interactions.** As alluded at the end of Section 1.3, we have considered particle systems having pairwise interactions. We could consider dynamics with “triplewise” interactions (or more):

$$\dot{\xi}_i(t) = \frac{1}{N^2} \sum_{j,k=1}^N G(t, x_i, x_j, x_k, \xi_i(t), \xi_j(t), \xi_k(t)).$$

For such dynamics, the mean field is then formally obtained as

$$\mathcal{X}[\mu](t, x, \xi) = \int_{\Omega \times \mathbf{R}^d} \int_{\Omega \times \mathbf{R}^d} G(t, x, x', x'', \xi, \xi', \xi'') d\mu(x', \xi') d\mu(x'', \xi'')$$

and the Vlasov equation (45) remains the same. Obtaining a theorem of existence and uniqueness of a solution  $t \mapsto \mu(t)$  to the Vlasov equation, like in Theorem 3.1, is an open problem. The above mean field  $\mathcal{X}[\mu]$  is indeed now quadratic with respect to  $\mu$  and this may complicate significantly the analysis. We do not know if such triplewise (or more) interaction particle systems have been studied in the literature.

**Similar models.** There exist some interesting models in the literature, not covered by our analysis but are not far. A first example is the more general Hegselmann–Krause model

$$\dot{\xi}_i^N(t) = \frac{1}{N} \sum_{j=1}^N \frac{\phi(\|\xi_i^N(t) - \xi_j^N(t)\|)}{\frac{1}{N} \sum_{k=1}^N \phi(\|\xi_i^N(t) - \xi_k^N(t)\|)} (\xi_j^N(t) - \xi_i^N(t)), \quad i \in \{1, \dots, N\},$$

studied in (see [57]). Its graph limit is the Euler equation

$$\partial_t y(t, x) = \int_{\mathbf{R}^d} \frac{\phi(\|y(t, x) - y(t, x')\|)}{\int_{\mathbf{R}^d} \phi(\|y(t, x) - y(t, x'')\|) dx''} (y(t, x') - y(t, x)) dx'$$

and the mean field is

$$\mathcal{X}[\mu](\xi) = \int_{\mathbf{R}^d} \frac{\phi(\|\xi - \xi'\|)}{\int_{\mathbf{R}^d} \phi(\|\xi - \xi''\|) d\mu(\xi'')} (\xi' - \xi) d\mu(\xi').$$

The corresponding Vlasov equation is studied in [57].

A second example is the Transformers model studied in [34]

$$\dot{\xi}_i^N(t) = \frac{1}{N} \sum_{j=1}^N \frac{\exp\langle Q\xi_i^N(t), K\xi_j^N(t) \rangle}{\frac{1}{N} \sum_{k=1}^N \exp\langle Q\xi_i^N(t), K\xi_k^N(t) \rangle} V\xi_j^N(t), \quad i \in \{1, \dots, N\},$$

where  $Q$ ,  $K$  and  $V$  are matrices. This interacting particle model seems to be particularly relevant in artificial intelligence. Although the above dynamics cannot be written in the form of the particle system (8), it is quite evident that all the theory developed in this paper extends to such cases and that its graph limit is the Euler equation

$$\partial_t y(t, x) = \int_{\mathbf{R}^d} \frac{\exp\langle Qy(t, x), Ky(t, x') \rangle}{\int_{\mathbf{R}^d} \exp\langle Qy(t, x), Ky(t, x'') \rangle dx''} Vy(t, x') dx'$$

and that the Vlasov equation (studied [34, Section 6.3]) is (45) with the mean field

$$\mathcal{X}[\mu](\xi) = \int_{\mathbf{R}^d} \frac{\exp\langle Q\xi, K\xi' \rangle}{\int_{\mathbf{R}^d} \exp\langle Q\xi, K\xi'' \rangle d\mu(\xi'')} V\xi' d\mu(\xi').$$

There exist also other variants of particle systems, involving some delays, or some coupling with other equations (like in the Keller-Segel model). We think that many of them can be covered by slight extensions of the analysis done in this paper.

**Stochastic particle systems.** Throughout this article, we have focused on *deterministic* finite systems of particles. Since many stochastic systems of interacting particles, involving noise, can be relevant in modeling collective behavior, it is of interest to extend the results of this paper to the stochastic context. For example, using Ito calculus, it is proved in [12] that taking the stochastic mean field limit in a kinetic McKean-Vlasov type finite particle system, involving some Brownian motion, leads to a kinetic Fokker-Planck equation. The general picture drawn on Figure 1 remains to be investigated in the stochastic setting.

**Control at the various scales.** From the control theory viewpoint, it is natural to add a control term in the particle system (8) and, accordingly, in the limit Euler equation (31) and in the Vlasov equation (45). This was done in [1, 15, 30, 64, 68] (just to cite a few). The main objective is then to obtain “commutative diagrams” in the following sense: if  $u$  is a control for a limit equation then one wants that there is an explicit sequence of controls  $u^N$  for the particle system, converging to  $u$  (this is the easy part); conversely, and much more difficultly, one wants to design controls  $u^N$  for the family of particle systems, indexed by  $N$ , converging to a control  $u$  for the limit equation. This question can be settled in various contexts: exact control, optimal control, stabilization. This is a major challenge.

**Numerical consequences.** All results and error estimates derived in our paper show that the solutions of particle systems provide good approximations of the Euler or of the Vlasov equation. In numerical analysis, particle methods, or particle-in-cell methods, have been much used in particular to approximation solutions of fluid equations (see, e.g., [23, 70]). Of course, these equations involve unbounded operators. But the results announced in Section 7.1 open a new perspective regarding numerical issues, to be explored.



## A Appendix

Let  $E$  be a Polish space, endowed with a distance  $d_E$ .

### A.1 Some general facts on the Wasserstein distance

**Choice of a distance on  $E^k$ .** Let  $q \in [1, +\infty]$  be arbitrarily fixed. Given any  $k \in \mathbb{N}^*$ , we endow  $E^k$  with the  $\ell^q$  distance based on  $d_E$ , defined by

$$d_{E^k}^{[q]}(y, y') = \|(d_E(y_1, y'_1), \dots, d_E(y_k, y'_k))\|_{\ell^q} = \begin{cases} \left( \sum_{i=1}^k d_E(y_i, y'_i)^q \right)^{1/q} & \text{if } 1 \leq q < +\infty \\ \max_{1 \leq i \leq k} d_E(y_i, y'_i) & \text{if } q = +\infty \end{cases} \quad (87)$$

for all  $y = (y_1, \dots, y_k)$  and  $y' = (y'_1, \dots, y'_k)$  in  $E^k$ .

Fixing such a choice has an impact on the computation of the Wasserstein distance  $W_p$  between two probability measures on  $E^k$ . Indeed, this means that the distance (87) is used in the definition (21) of  $W_p$ , and that, in the definition (24) of  $W_1$ , the Lipschitz constants must be computed with the distance (87). The lemma below is thus important to compute Lipschitz constants.

**Lemma A.1.** *Let  $f \in \text{Lip}(E^k)$ . Then, for any  $y_2, \dots, y_k \in E$ , the mapping  $y_1 \mapsto f(y_1, y_2, \dots, y_k)$  is Lipschitz, of Lipschitz constant less than  $\text{Lip}(f)$ . We set  $\text{Lip}_{y_1}(f) = \max\{\text{Lip}(f(\cdot, y_2, \dots, y_k)) \mid y_2, \dots, y_k \in E\}$ . All other  $\text{Lip}_{y_i}(f)$  are defined similarly, for  $i = 2, \dots, k$ . We have*

$$\text{Lip}(f) = \|(\text{Lip}_{y_1}(f), \dots, \text{Lip}_{y_k}(f))\|_{\ell^{q'}} = \begin{cases} \left( \sum_{i=1}^k \text{Lip}_{y_i}(f)^{q'} \right)^{1/q'} & \text{if } q' < +\infty \\ \max_{1 \leq i \leq k} \text{Lip}_{y_i}(f) & \text{if } q' = +\infty \end{cases}$$

where  $q' \in [1, +\infty]$  is defined by  $\frac{1}{q} + \frac{1}{q'} = 1$ .

*Proof.* It suffices to write

$$\begin{aligned} & |f(y_1, \dots, y_k) - f(y'_1, \dots, y'_k)| \\ & \leq |f(y_1, y_2, \dots, y_k) - f(y'_1, y_2, \dots, y_k)| + \dots + |f(y'_1, \dots, y'_{k-1}, y_k) - f(y'_1, \dots, y'_{k-1}, y'_k)| \\ & \leq \sum_{i=1}^k \text{Lip}_{y_i}(f) d_E(y_i, y'_i) \end{aligned}$$

and to use the Hölder inequality. □

**Remark A.1.** The choice of a distance  $d_{E^k}^{[q]}$  on the tensor product  $E^k$  (i.e., the choice of  $q \in [1, +\infty]$ ) is far from being insignificant because, although all norms are equivalent in  $E^k$ , comparing them gives constants depending on  $k$ . The choice thus becomes particularly meaningful when  $k$  is large.

Another remark is that the definition (87) is based on the usual  $\ell^q$  norm, for  $q \in [1, +\infty]$ . Other choices are possible, but in order to keep many of the statements further the convexity of the norm is important.

**Notation**  $W_p^{[q]}$ . For all  $p, q \in [1, +\infty]$ , following Remark A.1, hereafter we denote by  $W_p^{[q]}$  the Wasserstein distance  $W_p$  on  $\mathcal{P}(E^k)$  (defined by (21)) with respect to the distance  $d_{E^k}^{[q]}$  on  $E^k$ .

It follows from the usual inequalities for  $\ell^q$  norms in  $\mathbb{R}^k$  that  $q \mapsto d_{E^k}^{[q]}$  is decreasing and

$$1 \leq q_1 \leq q_2 \leq +\infty \Rightarrow d_{E^k}^{[q_2]} \leq d_{E^k}^{[q_1]} \leq k^{\frac{1}{q_1} - \frac{1}{q_2}} d_{E^k}^{[q_2]}$$

and thus

$$1 \leq q_1 \leq q_2 \leq +\infty \Rightarrow W_p^{[q_2]} \leq W_p^{[q_1]} \leq k^{\frac{1}{q_1} - \frac{1}{q_2}} W_p^{[q_2]} \quad (88)$$

for any  $p \in [1, +\infty]$ . These inequalities complement (23). For  $p$  fixed, in the family of distances  $W_p^{[q]}$ , for  $q \in [1, +\infty]$ , the  $\ell^1$  distance  $W_p^{[1]}$  is the weakest one. This is an important point because, in the existing literature, very often the  $\ell^2$  distance  $W_p^{[2]}$  is used, but in this work the use of  $q = 1$  is crucial for some parts.

In all subsections hereafter, we fix an arbitrary  $p \in [1, +\infty)$ . The case  $p = +\infty$  is obtained by taking the limit when it makes sense. We also fix an arbitrary  $q \in [1, +\infty]$ .

### A.1.1 Convexity

**Lemma A.2** ( $(W_p)^p$  is convex). *Given any  $\mu_1, \mu_2, \mu'_1, \mu'_2 \in \mathcal{P}(E)$  and any  $\lambda \in [0, 1]$ , we have*

$$W_p(\lambda\mu_1 + (1-\lambda)\mu_2, \lambda\mu'_1 + (1-\lambda)\mu'_2)^p \leq \lambda W_p(\mu_1, \mu'_1)^p + (1-\lambda) W_p(\mu_2, \mu'_2)^p.$$

*Proof.* This result is a particular case of [78, Part I, Chapter 4, Theorem 4.8]. Let  $\Pi_i$  be an optimal coupling between  $\mu_i$  and  $\mu'_i$ , for  $i = 1, 2$ . Then  $\Pi = \lambda\Pi_1 + (1-\lambda)\Pi_2$  couples  $\lambda\mu_1 + (1-\lambda)\mu_2$  and  $\lambda\mu'_1 + (1-\lambda)\mu'_2$  (maybe not optimally). Hence

$$\begin{aligned} & W_p(\lambda\mu_1 + (1-\lambda)\mu_2, \lambda\mu'_1 + (1-\lambda)\mu'_2)^p \leq \int_E d_E(x, x')^p d\Pi(x, x') \\ & = \lambda \int_E d_E(x, x')^p d\Pi_1(x, x') + (1-\lambda) \int_E d_E(x, x')^p d\Pi_2(x, x') = \lambda W_p(\mu_1, \mu'_1)^p + (1-\lambda) W_p(\mu_2, \mu'_2)^p \end{aligned}$$

and the lemma follows.  $\square$

**Lemma A.3.** *Let  $\mu_1, \mu_2, \beta \in \mathcal{P}(E)$  and let  $\varepsilon \in (0, 1]$  be such that  $\mu_1 = (1+\varepsilon)\mu_2 - \varepsilon\beta$ . Then*

$$W_p(\mu_1, \mu_2) \leq \varepsilon^{1/p} W_p(\mu_1, \beta)$$

and, assuming that  $\varepsilon < 1$ ,

$$W_p(\mu_1, \mu_2) \leq \frac{\varepsilon^{1/p}}{1 - \varepsilon^{1/p}} W_p(\mu_2, \beta).$$

In the particular case  $p = 1$ , we have  $W_1(\mu_1, \mu_2) = \varepsilon W_1(\mu_2, \beta)$ .

*Proof.* We have  $\mu_2 = \frac{1}{1+\varepsilon}\mu_1 + \frac{\varepsilon}{1+\varepsilon}\beta$  (convex combination), and applying Lemma A.2 we get  $W_p(\mu_1, \mu_2)^p \leq \frac{\varepsilon}{1+\varepsilon} W_p(\mu_1, \beta)^p \leq \varepsilon W_p(\mu_1, \beta)^p$ , and the first inequality follows. The second inequality is obtained by using the triangular inequality  $W_p(\mu_1, \beta) \leq W_p(\mu_1, \mu_2) + W_p(\mu_2, \beta)$ . When  $p = 1$ , given any  $f \in \mathcal{C}_c^0(E)$ , we have  $\int_E f d(\mu_1 - \mu_2) = \varepsilon \int_E f d(\mu_2 - \beta)$ , and taking (in two steps) the supremum over all  $f$  such that  $\text{Lip}(f) \leq 1$ , we get  $W_1(\mu_1, \mu_2) = \varepsilon W_1(\mu_2, \beta)$ .  $\square$

### A.1.2 Symmetrization

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

$$\mu^s = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \sigma_* \mu \quad (89)$$

where the measure  $\sigma_* \mu$  is defined by  $\langle \sigma_* \mu, f \rangle = \langle \mu, \sigma^* f \rangle$  and  $(\sigma^* f)(y) = f(\sigma \cdot y)$ , with  $\sigma \cdot y = (y_{\sigma(1)}, \dots, y_{\sigma(N)})$  for every  $y \in E^N$  and for every  $\sigma \in \mathfrak{S}_N$ , where  $\mathfrak{S}_N$  is the group of permutations of  $N$  elements. Here,  $\langle \cdot, \cdot \rangle$  is the duality bracket. Equivalently,

$$\int_{E^N} f(y) d\mu^s(y) = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \int_{E^N} f(\sigma \cdot y) d\mu(y) \quad \forall f \in \mathcal{C}_c^0(E^N).$$

**Lemma A.4.** *Given any  $\mu_1, \mu_2 \in \mathcal{P}(E^N)$ , we have*

$$W_p^{[q]}(\mu_1^s, \mu_2^s) \leq W_p^{[q]}(\mu_1, \mu_2).$$

In this lemma, the Wasserstein distance  $W_p$  is computed with respect to the  $\ell^q$  distance  $d_{E^N}^{[q]}$ .

*Proof.* This follows from Lemma A.2, since  $\mu^s$  is written as the convex combination (89), noting that  $W_p^{[q]}(\sigma_* \mu_1, \sigma_* \mu_2) = W_p^{[q]}(\mu_1, \mu_2)$  for any  $\sigma \in \mathfrak{S}_N$  because the distance  $d_{E^N}^{[q]}$  defined by (87) is itself symmetric and because, for any  $\Pi$  coupling  $\mu_1$  and  $\mu_2$  and for any  $\sigma \in \mathfrak{S}_N$ ,  $(\sigma \otimes \sigma)_* \Pi$  couples  $\sigma_* \mu_1$  and  $\sigma_* \mu_2$ .  $\square$

### A.1.3 Marginals

Let  $N \in \mathbb{N}^*$  be arbitrary. Given any  $\mu \in \mathcal{P}(E^N)$  and any  $k \in \{1, \dots, N\}$ , the  $k^{\text{th}}$ -order marginal  $\mu_{N:k} \in \mathcal{P}(E^k)$  of  $\mu$  is the image of  $\mu$  under the canonical projection  $\pi_k : E^N = E^k \times E^{N-k} \rightarrow E^k$ .

**Lemma A.5.** *Given any  $\mu_1, \mu_2 \in \mathcal{P}(E^N)$  and any  $k \in \{1, \dots, N\}$ , we have*

$$W_p^{[q]}((\mu_1)_{N:k}, (\mu_2)_{N:k}) \leq W_p^{[q]}(\mu_1, \mu_2). \quad (90)$$

The Wasserstein distance at the left-hand (resp., right-hand) side of (90) is computed with respect to the  $\ell^q$  distance  $d_{E^k}^{[q]}$  (resp.,  $d_{E^N}^{[q]}$ ). We will establish in Lemma A.13 in Appendix A.2.2 a stronger estimate when  $\mu_1$  and  $\mu_2$  are symmetric.

*Proof.* Let  $\Pi$  be an optimal coupling between  $\mu_1$  and  $\mu_2$ . Then, obviously,  $(\pi_k)_* \Pi$  couples (maybe not optimally)  $(\pi_k)_* \mu_1 = (\mu_1)_{N:k}$  and  $(\pi_k)_* \mu_2 = (\mu_2)_{N:k}$ . Therefore

$$\begin{aligned} W_p^{[q]}((\mu_1)_{N:k}, (\mu_2)_{N:k})^p &\leq \int_{E^k} d_{E^k}^{[q]}((y_1, \dots, y_k), (y'_1, \dots, y'_k))^p d(\pi_k)_* \Pi((y_1, \dots, y_k), (y'_1, \dots, y'_k)) \\ &\leq \int_{E^N} d_{E^k}^{[q]}(\pi_k(y), \pi_k(y'))^p d\Pi(y, y') \\ &\leq \int_{E^N} d_{E^N}^{[q]}(y, y')^p d\Pi(y, y') = W_p^{[q]}(\mu_1, \mu_2)^p \end{aligned}$$

where we have used that  $d_{E^k}^{[q]}(\pi_k(y), \pi_k(y')) \leq d_{E^N}^{[q]}(y, y')$ .  $\square$

### A.1.4 Tensor product

Let  $k \in \mathbb{N}^*$ . For every  $i \in \{1, \dots, k\}$ , let  $E_i$  be a Polish space, endowed with a distance  $d_{E_i}$ . We endow the product space  $E_1 \times \dots \times E_k$  with the distance

$$d_{E_1 \times \dots \times E_k}^{[q]}(y, y') = \|(d_{E_1}(y_1, y'_1), \dots, d_{E_k}(y_k, y'_k))\|_{\ell^q} = \begin{cases} \left( \sum_{i=1}^k d_{E_i}(y_i, y'_i)^q \right)^{1/q} & \text{if } 1 \leq q < +\infty \\ \max_{1 \leq i \leq k} d_{E_i}(y_i, y'_i) & \text{if } q = +\infty \end{cases} \quad (91)$$

for all  $y = (y_1, \dots, y_k)$ ,  $y' = (y'_1, \dots, y'_k) \in E_1 \times \dots \times E_k$ .

**Lemma A.6.** *Given any  $\mu_1, \mu'_1 \in \mathcal{P}(E_1)$ ,  $\dots$ ,  $\mu_k, \mu'_k \in \mathcal{P}(E_k)$ , we have, for every  $j \in \{1, \dots, k\}$ ,*

$$W_p(\mu_j, \mu'_j) \leq W_p^{[q]} \left( \bigotimes_{i=1}^k \mu_i, \bigotimes_{i=1}^k \mu'_i \right) \leq \max \left( k^{\frac{1}{q} - \frac{1}{p}}, 1 \right) \left( \sum_{i=1}^k W_p(\mu_i, \mu'_i)^p \right)^{1/p} \quad (92)$$

and the right-hand side inequality in (92) is an equality if  $p = q$ .

Taking  $E_i = E$ ,  $d_{E_i} = d_E$ ,  $\mu_i = \mu$  and  $\mu'_i = \mu'$  for every  $i \in \{1, \dots, n\}$ , we have the slightly stronger inequality

$$W_p^{[q]}(\mu^{\otimes k}, (\mu')^{\otimes k}) \leq k^{1/q} W_p(\mu, \mu') \quad (93)$$

and the inequality is an equality if  $p = q$ .

Lemma A.6 can be found in [51].

The Wasserstein distance  $W_p$  at the left-hand side of (92) is computed with respect to the distance  $d_{E_j}$ . The Wasserstein distance  $W_p^{[q]}$  in the middle of (92) is computed with respect to the distance  $d_{E_1 \times \dots \times E_k}^{[q]}$  defined by (91).

The Wasserstein distance  $W_p^{[q]}$  at the left-hand side of (93) is computed with respect to the distance  $d_{E^k}^{[q]}$  defined by (87). Recall that  $q \in [1, +\infty]$  has been chosen arbitrarily to define this distance. At the right-hand side of (93), if  $q = +\infty$  then  $k^{1/q} = 1$ .

**Remark A.2.** As a particular case of (92), taking  $k = 2$  and  $\mu_2 = \mu'_2 = \mu$ , we have

$$W_p(\mu_1, \mu'_1) \leq W_p^{[q]}(\mu_1 \otimes \mu, \mu'_1 \otimes \mu) = W_p^{[q]}(\mu \otimes \mu_1, \mu \otimes \mu'_1) \leq \max \left( 2^{\frac{1}{q} - \frac{1}{p}}, 1 \right) W_p(\mu_1, \mu'_1).$$

In particular, if  $p \leq q$  then  $W_p(\mu_1, \mu'_1) = W_p^{[q]}(\mu_1 \otimes \mu, \mu'_1 \otimes \mu) = W_p^{[q]}(\mu \otimes \mu_1, \mu \otimes \mu'_1)$ .

*Proof.* We have  $W_p(\mu_j, \mu'_j) \leq W_p^{[q]} \left( \bigotimes_{i=1}^k \mu_i, \bigotimes_{i=1}^k \mu'_i \right)$  for every  $i \in \{1, \dots, k\}$ : this is proved like in Lemma A.5 because  $\mu_j$  is the marginal on  $E_j$  of the measure  $\bigotimes_{i=1}^k \mu_i$  on  $E$ , and similarly for  $\mu'_j$ . Therefore the left-hand side inequality in (92) follows.

Let us now establish the right-hand side inequality in (92), for  $q < +\infty$ . For every  $i \in \{1, \dots, k\}$ , let  $\Pi_i$  be an optimal coupling between  $\mu_i$  and  $\mu'_i$ . Then, obviously,  $\Pi = \bigotimes_{i=1}^k \Pi_i$  couples (maybe not optimally)  $\bigotimes_{i=1}^k \mu_i$  and  $\bigotimes_{i=1}^k \mu'_i$ . Therefore

$$W_p^{[q]} \left( \bigotimes_{i=1}^k \mu_i, \bigotimes_{i=1}^k \mu'_i \right)^p \leq \int_{E_1 \times E_1} \dots \int_{E_k \times E_k} \left( \sum_{i=1}^k d_{E_i}(y_i, y'_i)^q \right)^{p/q} d\Pi_k(y_k, y'_k) \dots d\Pi_1(y_1, y'_1).$$

If  $p \geq q$ , using the convexity inequality  $(|a_1| + \dots + |a_k|)^r \leq k^{r-1}(|a_1|^r + \dots + |a_k|^r)$  for  $r \geq 1$  (with equality for  $r = 1$ ), we obtain

$$W_p^{[q]} \left( \bigotimes_{i=1}^k \mu_i, \bigotimes_{i=1}^k \mu'_i \right)^p \leq k^{\frac{p}{q}-1} \sum_{i=1}^k W_p(\mu_i, \mu'_i)^p$$

and the inequality is an equality if  $p = q$  because in this case  $\Pi$  is an optimal coupling. If  $p \leq q$ , using the inequality  $(|a_1| + \dots + |a_k|)^{1/r} \leq |a_1|^{1/r} + \dots + |a_k|^{1/r}$  for  $r \geq 1$ , we obtain

$$W_p^{[q]} \left( \bigotimes_{i=1}^k \mu_i, \bigotimes_{i=1}^k \mu'_i \right)^p \leq \sum_{i=1}^k W_p(\mu_i, \mu'_i)^p.$$

All in all, we have established (92).

To prove (93), using the definition (22) of  $W_p$ , we note that

$$W_p^{[q]}(\mu^{\otimes k}, (\mu')^{\otimes k})^p \leq \mathbb{E} \left( \sum_{i=1}^k d_E(Y, Y')^q \right)^{p/q} = k^{p/q} \mathbb{E} d_E(Y, Y')^p = k^{p/q} W_p(\mu, \mu')^p$$

where  $Y$  and  $Y'$  are random variables (with values in  $E$ ) of laws  $\mu$  and  $\mu'$ , such that  $W_p(\mu, \mu')^p = \mathbb{E} d_E(Y, Y')^p$ .  $\square$

### A.1.5 Diameter of the support

**Lemma A.7.** *Given any  $\mu_1, \mu_2 \in \mathcal{P}_c(E)$ , we have*

$$W_p(\mu_1, \mu_2) \leq \text{diam}_E(\text{supp}(\mu_1) \cup \text{supp}(\mu_2)) = \max\{d_E(y, y') \mid y, y' \in \text{supp}(\mu_1) \cup \text{supp}(\mu_2)\}.$$

*Proof.* By (21), since  $W_p(\mu_1, \mu_2)^p$  is the infimum of  $\int_{E^2} d_E(y, y')^p d\Pi(y, y')$  over all probability measures  $\Pi$  on  $E^2$  coupling  $\mu_1$  and  $\mu_2$ , we have  $W_p(\mu_1, \mu_2) \leq \max\{d_E(y_1, y_2) \mid y_1 \in \text{supp}(\mu_1), y_2 \in \text{supp}(\mu_2)\}$ , and the result follows.  $\square$

### A.1.6 Propagation

In this section, we assume that  $E$  is a Banach space, endowed with a norm  $\|\cdot\|_E$ . Let also  $\Lambda$  (space of parameters) be a Polish space, endowed with a distance  $d_\Lambda$ . The space  $\Lambda \times E$  is endowed with the distance  $d_{\Lambda \times E} = d_\Lambda + d_E$ , where  $d_E$  is the distance on  $E$  induced by the norm  $\|\cdot\|_E$ .

**Lemma A.8.** *For  $i = 1, 2$ , let  $Y^i(t, \lambda, \cdot)$  be a continuous time-varying vector field on  $E$ , depending on the parameter  $\lambda \in \Lambda$ , locally Lipschitz with respect to  $(\lambda, y) \in \Lambda \times E$  uniformly with respect to  $t$  on any compact interval, generating a flow  $(\Phi^i(t, t_0, \lambda, \cdot))_{t \in \mathbb{R}}$  (assumed to be well defined for every  $t \in \mathbb{R}$ ) for any  $t_0 \in \mathbb{R}$ , that is,*

$$\begin{aligned} \partial_t \Phi^i(t, t_0, \lambda, y) &= Y^i(t, \lambda, \Phi^i(t, t_0, \lambda, y)) \\ \Phi^i(t_0, t_0, \lambda, y) &= y \end{aligned}$$

for all  $t, t_0 \in \mathbb{R}$ ,  $y \in E$  and  $\lambda \in \Lambda$ . Given any  $t_0 \in \mathbb{R}$  and any  $\mu^1(t_0), \mu^2(t_0) \in \mathcal{P}_c(\Lambda \times E)$ , we set  $\mu_t^i = \mu^i(t) = \Phi^i(t, t_0)_* \mu^i(t_0)$  for every  $t \geq t_0$ , for  $i = 1, 2$ ; this notation means, denoting by  $\nu^i$  the (constant in time) marginal of  $\mu^i(t)$  on  $\Lambda$  and disintegrating  $\mu_t^i = \int_\Lambda \mu_{t, \lambda}^i d\nu^i(\lambda)$ , that  $\mu_{t, \lambda}^i = \Phi^i(t, t_0, \lambda, \cdot)_* \mu^i(t_0)$  for  $\nu^i$ -almost every  $\lambda \in \Lambda$ . For every  $p \in [1, +\infty)$ , we have

$$W_p(\mu^1(t), \mu^2(t)) \leq e^{(t-t_0)L([t_0, t])} W_p(\mu^1(t_0), \mu^2(t_0)) + M([t_0, t]) \frac{e^{(t-t_0)L([t_0, t])} - 1}{L([t_0, t])} \quad (94)$$

for every  $t \geq t_0$ , where<sup>6</sup>

$$L([t_0, t]) = \max_{t_0 \leq \tau \leq t} \text{Lip}(Y^1(\tau, \cdot, \cdot)|_{S(\tau)}), \quad (95)$$

$$S(t) = (\text{supp}(\nu^1) \cup \text{supp}(\nu^2)) \times \Phi^1(t, t_0, \text{supp}(\mu^1(t_0)) \cup \text{supp}(\mu^2(t_0))) \cup \text{supp}(\mu^2(t)),$$

$$M([t_0, t]) = \max\{\|Y^1(\tau, \lambda, y) - Y^2(\tau, \lambda, y)\|_E \mid t_0 \leq \tau \leq t, (\lambda, y) \in \text{supp}(\mu^2(\tau))\}. \quad (96)$$

Alternatively, the second term at the right-hand side of (94) can be replaced by

$$M_p([t_0, t])(t - t_0)^{1/p} \left( \frac{e^{(t-t_0)p'L([t_0, t])} - 1}{p'L([t_0, t])} \right)^{1/p'} \quad (97)$$

where  $\frac{1}{p} + \frac{1}{p'} = 1$  and

$$M_p([t_0, t]) = \max_{t_0 \leq \tau \leq t} \left( \int_{\Lambda \times E} \|Y^1(\tau, \lambda, y) - Y^2(\tau, \lambda, y)\|_E^p d\mu_\tau^2(\lambda, y) \right)^{1/p}. \quad (98)$$

Some remarks are in order:

- In (94) (and in (97)), it is understood that if  $L([t_0, t]) = 0$  then  $\frac{e^{(t-t_0)L([t_0, t])} - 1}{L([t_0, t])}$  is replaced by  $t - t_0$ . Lemma A.8 extends [67, Proposition 4] to the case with parameters and to the local Lipschitz case; also, the alternative (not usual) estimate with (97) is useful to derive some results of this paper.
- If  $Y^1 = Y^2$  then  $M(\cdot) = 0$ .
- When  $t_0 = 0$ , we denote  $\Phi^i(t, \lambda, y) = \Phi^i(t, 0, \lambda, y)$ ,  $L(t) = L([0, t])$  and  $M(t) = M([0, t])$ .
- Finally, it is interesting to observe that, in Lemma A.8, actually only the first vector field  $Y^1$  is required to be locally Lipschitz. Concerning the second, it is only required that  $Y^2$  is regular enough so that (98) is well defined, and also that the flow  $\Phi^2$  is well defined.

*Proof.* Given any  $(\lambda_1, y_1) \in \text{supp}(\mu^1(t_0))$  and  $(\lambda_2, y_2) \in \text{supp}(\mu^2(t_0))$ , using (95) we have

$$\begin{aligned} & \partial_t \|\Phi^1(t, t_0, \lambda_1, y_1) - \Phi^1(t, t_0, \lambda_2, y_2)\|_E \\ & \leq \|Y^1(t, \lambda_1, \Phi^1(t, t_0, \lambda_1, y_1)) - Y^1(t, \lambda_2, \Phi^1(t, t_0, \lambda_2, y_2))\|_E \\ & \leq L([t_0, t]) (d_\Lambda(\lambda_1, \lambda_2) + \|\Phi^1(t, t_0, \lambda_1, y_1) - \Phi^1(t, t_0, \lambda_2, y_2)\|_E) \end{aligned}$$

because  $\Phi^1(t, t_0, \lambda_1, y_1) \in \Phi^1(t, t_0, \text{supp}(\mu^1(t_0)))$  and  $\Phi^1(t, t_0, \lambda_2, y_2) \in \Phi^1(t, t_0, \text{supp}(\mu^2(t_0)))$  (this motivates the definition of  $S(t)$ ), and by integration we get that

$$d_\Lambda(\lambda_1, \lambda_2) + \|\Phi^1(t, t_0, \lambda_1, y_1) - \Phi^1(t, t_0, \lambda_2, y_2)\|_E \leq e^{(t-t_0)L([t_0, t])} (d_\Lambda(\lambda_1, \lambda_2) + \|y_1 - y_2\|_E) \quad (99)$$

for every  $t \geq t_0$ , for  $i = 1, 2$  (we have used the fact that  $t \mapsto L([t_0, t])$  is nondecreasing).

Taking an optimal coupling  $\Pi_{t_0} \in \mathcal{P}((\Lambda \times E)^2)$  between  $\mu^1(t_0)$  and  $\mu^2(t_0)$ , the probability measure  $\Pi_t = (\Phi^1(t, t_0) \otimes \Phi^2(t, t_0))_* \Pi_{t_0}$  couples (maybe not optimally)  $\mu^1(t)$  with  $\mu^2(t)$ .<sup>7</sup> Therefore,

<sup>6</sup>Note that  $S(t)$  is compact and that  $\Phi^i(t, t_0, \text{supp}(\mu^i(t_0))) = \text{supp}(\mu^i(t))$ .

<sup>7</sup>Indeed, denoting by  $\pi_i$  the projection of  $(\Lambda \times E)^2$  onto the  $i^{\text{th}}$ -copy of  $\Lambda \times E$ , we have  $\pi_i \circ (\Phi^1 \otimes \Phi^2) = \Phi^i \circ \pi_i$ .

using the definition (21) of  $W_p$ ,

$$\begin{aligned}
& W_p(\mu^1(t), \mu^2(t))^p \\
& \leq \int_{(\Lambda \times E)^2} (\mathrm{d}_\Lambda(\lambda_1, \lambda_2) + \|y_1 - y_2\|_E)^p d\Pi_t(\lambda_1, y_1, \lambda_2, y_2) \\
& = \int_{(\Lambda \times E)^2} (\mathrm{d}_\Lambda(\lambda_1, \lambda_2) + \|\Phi^1(t, t_0, \lambda_1, y_1) - \Phi^2(t, t_0, \lambda_2, y_2)\|_E)^p d\Pi_{t_0}(\lambda_1, y_1, \lambda_2, y_2) \\
& \leq \int_{(\Lambda \times E)^2} \left( \mathrm{d}_\Lambda(\lambda_1, \lambda_2) + \|\Phi^1(t, t_0, \lambda_1, y_1) - \Phi^1(t, t_0, \lambda_2, y_2)\|_E \right. \\
& \quad \left. + \|\Phi^1(t, t_0, \lambda_2, y_2) - \Phi^2(t, t_0, \lambda_2, y_2)\|_E \right)^p d\Pi_{t_0}(\lambda_1, y_1, \lambda_2, y_2)
\end{aligned}$$

and thus, and using the triangular inequality in  $L^p$ , we get

$$\begin{aligned}
& W_p(\mu^1(t), \mu^2(t)) \\
& \leq \left( \int_{(\Lambda \times E)^2} (\mathrm{d}_\Lambda(\lambda_1, \lambda_2) + \|\Phi^1(t, t_0, \lambda_1, y_1) - \Phi^1(t, t_0, \lambda_2, y_2)\|_E)^p d\Pi_{t_0}(\lambda_1, y_1, \lambda_2, y_2) \right)^{1/p} \\
& \quad + \left( \int_{(\Lambda \times E)^2} \|\Phi^1(t, t_0, \lambda_2, y_2) - \Phi^2(t, t_0, \lambda_2, y_2)\|_E^p d\Pi_{t_0}(\lambda_1, y_1, \lambda_2, y_2) \right)^{1/p}
\end{aligned} \tag{100}$$

Using (99), the first term of the sum at the right-hand side of (100) is less than or equal to

$$\begin{aligned}
& e^{(t-t_0)L([t_0, t])} \left( \int_{(\Lambda \times E)^2} (\mathrm{d}_\Lambda(\lambda_1, \lambda_2) + \|y_1 - y_2\|_E)^p d\Pi_{t_0}(\lambda_1, y_1, \lambda_2, y_2) \right)^{1/p} \\
& = e^{(t-t_0)L([t_0, t])} W_p(\mu^1(t_0), \mu^2(t_0)),
\end{aligned}$$

the latter equality being because  $\Pi_{t_0}$  is an optimal coupling between  $\mu^1(t_0)$  and  $\mu^2(t_0)$ .

To treat the second term, we first observe that, for  $(\lambda, y) \in \text{supp}(\mu^2(t_0))$ ,

$$\begin{aligned}
\partial_t \|\Phi^1(t, t_0, \lambda, y) - \Phi^2(t, t_0, \lambda, y)\|_E & \leq \|Y^1(t, \lambda, \Phi^1(t, t_0, \lambda, y)) - Y^1(t, \lambda, \Phi^2(t, t_0, \lambda, y))\|_E \\
& \quad + \|Y^1(t, \lambda, \Phi^2(t, t_0, \lambda, y)) - Y^2(t, \lambda, \Phi^2(t, t_0, \lambda, y))\|_E \\
& \leq L([t_0, t]) \|\Phi^1(t, t_0, \lambda, y) - \Phi^2(t, t_0, \lambda, y)\|_E \\
& \quad + \|Y^1(t, \lambda, \Phi^2(t, t_0, \lambda, y)) - Y^2(t, \lambda, \Phi^2(t, t_0, \lambda, y))\|_E
\end{aligned}$$

where we have used (95), noting that  $(\lambda, \Phi^1(t, t_0, \lambda, y)) \in S(t)$  and  $(\lambda, \Phi^2(t, t_0, \lambda, y)) \in S(t)$ , and thus, using the Gronwall lemma and the fact that  $\tau \mapsto L([t_0, \tau])$  is nondecreasing,

$$\begin{aligned}
& \|\Phi^1(t, t_0, \lambda, y) - \Phi^2(t, t_0, \lambda, y)\|_E \\
& \leq \int_{t_0}^t e^{(t-\tau)L([t_0, t])} \|Y^1(\tau, \lambda, \Phi^2(\tau, t_0, \lambda, y)) - Y^2(\tau, \lambda, \Phi^2(\tau, t_0, \lambda, y))\|_E d\tau. \tag{101}
\end{aligned}$$

Using the definition (96) of  $M([t_0, t])$  and the fact that  $\Phi^2(\tau, t_0, \text{supp}(\mu^2(t_0))) = \text{supp}(\mu^2(\tau))$ , we get

$$\|\Phi^1(t, t_0, \lambda, y) - \Phi^2(t, t_0, \lambda, y)\|_E \leq M([t_0, t]) \frac{e^{(t-t_0)L([t_0, t])} - 1}{L([t_0, t])}.$$

Therefore, the second term of the sum at the right-hand side of (100) is estimated by

$$\begin{aligned} & \left( \int_{(\Lambda \times E)^2} \|\Phi^1(t, t_0, \lambda_2, y_2) - \Phi^2(t, t_0, \lambda_2, y_2)\|_E^p d\Pi_{t_0}(\lambda_1, y_1, \lambda_2, y_2) \right)^{1/p} \\ &= \left( \int_{\Lambda \times E} \|\Phi^1(t, t_0, \lambda, y) - \Phi^2(t, t_0, \lambda, y)\|_E^p d\mu_{t_0}^2(\lambda, y) \right)^{1/p} \leq M([t_0, t]) \frac{e^{(t-t_0)L([t_0, t])} - 1}{L([t_0, t])} \end{aligned}$$

where we have used that the second marginal of  $\Pi_{t_0}$  is  $\mu_{t_0}^2 = \mu^2(t_0)$ . The estimate (94) follows.

To obtain the alternative estimate with the term (97), we apply the Hölder inequality to the right-hand side of (101), obtaining

$$\begin{aligned} & \|\Phi^1(t, t_0, \lambda, y) - \Phi^2(t, t_0, \lambda, y)\|_E \\ & \leq \left( \frac{e^{p'(t-t_0)L([t_0, t])} - 1}{p'L([t_0, t])} \right)^{1/p'} \left( \int_{t_0}^t \|Y^1(\tau, \lambda, \Phi^2(\tau, t_0, \lambda, y)) - Y^2(\tau, \lambda, \Phi^2(\tau, t_0, \lambda, y))\|_E^p d\tau \right)^{1/p}. \end{aligned}$$

Therefore, the second term of the sum at the right-hand side of (100) is estimated by

$$\begin{aligned} & \left( \int_{(\Lambda \times E)^2} \|\Phi^1(t, t_0, \lambda_2, y_2) - \Phi^2(t, t_0, \lambda_2, y_2)\|_E^p d\Pi_{t_0}(\lambda_1, y_1, \lambda_2, y_2) \right)^{1/p} \\ &= \left( \int_{\Lambda \times E} \|\Phi^1(t, t_0, \lambda, y) - \Phi^2(t, t_0, \lambda, y)\|_E^p d\mu_{t_0}^2(\lambda, y) \right)^{1/p} \\ & \leq \left( \frac{e^{p'(t-t_0)L([t_0, t])} - 1}{p'L([t_0, t])} \right)^{1/p'} \left( \int_{t_0}^t \int_{\Lambda \times E} \|Y^1(\tau, \lambda, \Phi^2(\tau, t_0, \lambda, y)) \right. \\ & \qquad \qquad \qquad \left. - Y^2(\tau, \lambda, \Phi^2(\tau, t_0, \lambda, y))\|_E^p d\mu_{t_0}^2(\lambda, y) d\tau \right)^{1/p} \\ & \leq \left( \frac{e^{p'(t-t_0)L([t_0, t])} - 1}{p'L([t_0, t])} \right)^{1/p'} \left( \int_{t_0}^t \int_{\Lambda \times E} \|Y^1(\tau, \lambda, y) - Y^2(\tau, \lambda, y)\|_E^p d\mu_{\tau}^2(\lambda, y) d\tau \right)^{1/p} \\ & \leq \left( \frac{e^{p'(t-t_0)L([t_0, t])} - 1}{p'L([t_0, t])} \right)^{1/p'} (t - t_0)^{1/p} M_p([t_0, t]) \end{aligned}$$

The lemma is proved.  $\square$

**Lemma A.9.** *Let  $Y(t, \lambda, \cdot)$  be a continuous time-varying vector field on  $E$ , depending on the parameter  $\lambda \in \Lambda$ , locally Lipschitz with respect to  $y \in E$  uniformly with respect to  $(t, \lambda)$  on any compact, generating a flow  $(\Phi(t, t_0, \lambda, \cdot))_{t \in \mathbb{R}}$  (assumed to be well defined for every  $t \in \mathbb{R}$ ) for any  $t_0 \in \mathbb{R}$  (as in Lemma A.8). Given any  $t_0 \in \mathbb{R}$  and any  $\mu_{t_0} \in \mathcal{P}_c(\Lambda \times E)$ , we set  $\mu(t) = \Phi(t, t_0)_* \mu_{t_0}$  for every  $t \geq t_0$ . For every  $p \in [1, +\infty)$ , we have*

$$W_p(\mu(t), \mu(t_0)) \leq M([t_0, t]) |t - t_0| \quad \forall t \geq t_0$$

where  $M([t_0, t]) = \max \{\|Y(\tau, \lambda, y)\| \mid t_0 \leq \tau \leq t, (\lambda, y) \in \text{supp}(\mu(\tau))\}$ .

*Proof.* It would suffice to apply Lemma A.8 with  $Y^1 = 0$  and  $Y^2 = Y$ , if  $Y$  were also Lipschitz with respect to  $\lambda$  (or, to adapt this lemma to vector fields that depend only continuously on  $\lambda$ ). Without this assumption, let us give a quick proof. We first establish the following general result.



**Lemma A.10.** *Let  $F$  be a Polish space, endowed with a distance  $d_F$ , let  $\mu \in \mathcal{P}_c(F)$  and let  $\phi : F \rightarrow F$  be a measurable mapping. For every  $p \in [1, +\infty)$ , we have*

$$W_p(\phi_*\mu, \mu) \leq \left( \int_F d_F(y, \phi(y))^p d\mu(y) \right)^{1/p}$$

*Proof of Lemma A.10.* With a slight abuse of notation, we define  $\Pi \in \mathcal{P}(F \times F)$  by  $\Pi(y, y') = \mu(y) \delta_{y'=\Phi(y)}$ . Since we also have  $\Pi = (\Phi_*\mu)(y') \delta_{y'=\Phi(y)}$ , it follows that  $\Pi$  couples  $\mu$  and  $\phi_*\mu$ . Therefore  $W_p(\phi_*\mu, \mu)^p \leq \int_{F^2} d_F(y, y')^p d\Pi(y, y') = \int_F d_F(y, \phi(y))^p d\mu(y)$ . Lemma A.10 is proved.  $\square$

Applying Lemma A.10 with  $F = \Lambda \times E$ ,  $\mu = \mu_{t_0}$  and  $\phi = \Phi(t, t_0)$ , we have

$$W_p(\mu(t), \mu(t_0))^p \leq \int_{\Lambda \times E} d_{\Lambda \times E}((\lambda, y), (\lambda, \Phi(t, t_0, \lambda, y)))^p d\mu_{t_0}(\lambda, y)$$

and we note that  $d_{\Lambda \times E}((\lambda, y), (\lambda, \Phi(t, t_0, \lambda, y))) = \|\Phi(t, t_0, \lambda, y) - y\|$ . Now, since  $\Phi(t, t_0, \lambda, y) = y + \int_{t_0}^t Y(\tau, \lambda, \Phi(\tau, t_0, \lambda, y)) d\tau$  and  $\text{supp}(\mu(t)) = \Phi(t, t_0, \text{supp}(\mu_{t_0}))$ , Lemma A.9 easily follows.  $\square$

### A.1.7 Moment of order one

Let  $\Omega$  be a Polish space and let  $d \in \mathbb{N}^*$ .

**Lemma A.11.** *For  $i = 1, 2$ , let  $\mu_i \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ , disintegrated as  $\mu_i = \int_{\Omega} (\mu_i)_x d\nu_i(x)$  with respect to its marginal  $\nu_i$  on  $\Omega$ , and let  $y_i$  be the moment of order one of  $\mu_i$ , defined by  $y_i(x) = \int_{\mathbb{R}^d} \xi d(\mu_i)_x(\xi)$  for  $\nu_i$ -almost every  $x \in \Omega$ . Then*

$$W_1(y_1\nu_1, y_2\nu_2) \leq W_1(\mu_1, \mu_2).$$

*Proof.* By the definition (24) of  $W_1$ , we have  $W_1(y_1\nu_1, y_2\nu_2) = \int_{\Omega} y_1 g d\nu_1 - \int_{\Omega} y_2 g d\nu_2$  for some  $g \in \text{Lip}(\Omega)$  such that  $\text{Lip}(g) \leq 1$ . Hence  $W_1(y_1\nu_1, y_2\nu_2) = \int_{\Omega \times \mathbb{R}^d} f d(\mu_1 - \mu_2)$  where  $f(x, \xi) = g(x)\xi$ . Since  $f \in \text{Lip}(\Omega \times \mathbb{R}^d)$  and  $\text{Lip}(f) \leq 1$ , the result follows.  $\square$

## A.2 More precise facts on the marginals of a symmetrization

Let  $N \in \mathbb{N}^*$  be arbitrary. Recall that the symmetrization of a measure is defined by (89) (see Appendix A.1.2).

### A.2.1 First marginal of the symmetrization

For every  $i \in \{1, \dots, N\}$ , we denote by  $p^i$  the projection of  $E^N$  onto the  $i^{\text{th}}$  copy of  $E$ , i.e., in coordinates,  $p^i(y) = y_i$ .

**Lemma A.12.** *Let  $\mu \in \mathcal{P}(E^N)$  be arbitrary.*

- *The first marginal  $\mu_{N:1}^s = p_*^1 \mu^s$  of the symmetrization  $\mu^s$  of  $\mu$  is given by*

$$\mu_{N:1}^s = \frac{1}{N} \sum_{i=1}^N p_*^i \mu$$

*where  $p_*^i \mu$  is the image of  $\mu$  under the projection  $p^i$ . In other words,  $\mu_{N:1}^s$  is the average of the marginals of  $\mu$  on the copies of  $E$ .*

- We have  $p_*^i \mu^s = \frac{1}{N} \sum_{j=1}^N p_*^j \mu^s$  for every  $i \in \{1, \dots, N\}$  and thus  $p_*^i \mu^s$  does not depend on  $i$ .

In other words, the marginals of a symmetric measure on the copies of  $E$  are all equal; the same is true for the marginals of higher order.

*Proof.* Given any  $f \in \mathcal{C}_c^0(E)$ , we have

$$\begin{aligned} \langle \mu_{N:1}^s, f \rangle &= \langle p_*^1 \mu^s, f \rangle = \langle \mu^s, (p^1)^* f \rangle = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \langle \sigma_* \mu, (p^1)^* f \rangle = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \langle \mu, \sigma^* (p^1)^* f \rangle \\ &= \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \int_{E^N} f \circ p^1(\sigma \cdot y) d\mu(y) = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \int_{E^N} f(y_{\sigma(1)}) d\mu(y) \end{aligned}$$

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  $\text{card}(\mathfrak{S}_{N-1}) = (N - 1)!$ , we get that

$$\langle \mu_{N:1}^s, f \rangle = \frac{1}{N} \sum_{i=1}^N \int_{E^N} f(y_i) d\mu(y) = \frac{1}{N} \sum_{i=1}^N \int_{E^N} f \circ p^i(y) d\mu(y) = \frac{1}{N} \sum_{i=1}^N \langle p_*^i \mu, f \rangle$$

whence the first item.

The second item is proved in the same way, replacing  $\mu$  by  $\mu^s$ .  $\square$

## A.2.2 Marginals of symmetric measures

We have seen in Lemma A.5 (Appendix A.1.3) that  $W_p^{[q]}((\mu_1)_{N:k}, (\mu_2)_{N:k}) \leq W_p^{[q]}(\mu_1, \mu_2)$ , for every  $k \in \{1, \dots, N\}$ , for any  $\mu_1, \mu_2 \in \mathcal{P}(E^N)$ . We have a stronger estimate when  $\mu_1$  and  $\mu_2$  are symmetric, i.e., when  $\mu_1 = \mu_1^s$  and  $\mu_2 = \mu_2^s$ .

**Lemma A.13.** *Let  $\mu_1, \mu_2 \in \mathcal{P}(E^N)$  be symmetric measures. Then, for any  $k \in \{1, \dots, N\}$ ,*

$$W_p^{[q]}((\mu_1)_{N:k}, (\mu_2)_{N:k}) \leq \left( \frac{k}{N} \right)^{1/q} W_p^{[q]}(\mu_1, \mu_2). \quad (102)$$

In (102), the  $W_p^{[q]}$  distances are computed with respect to the  $\ell^q$  distances  $d_{E^k}^{[q]}$  and  $d_{E^N}^{[q]}$  defined by (87).

*Proof.* Assume that  $q < +\infty$  (for  $q = +\infty$ , it suffices to take limits). Let  $\Pi \in \mathcal{P}((E^N)^2)$  be an optimal coupling between  $\mu_1$  and  $\mu_2$  for the  $W_p^{[q]}$  distance, i.e., using the definitions (21) and (22) of  $W_p^{[q]}$ ,

$$W_p^{[q]}(\mu_1, \mu_2)^p = \int_{(E^N)^2} \left( \sum_{i=1}^N d_E(y_i^1, y_i^2) \right)^{p/q} d\Pi(y^1, y^2) = \mathbb{E} \left( \sum_{i=1}^N d_E(Y_i^1, Y_i^2) \right)^{p/q}$$

where  $y^j = (y_1^j, \dots, y_N^j)$  for  $j = 1, 2$ , and where the  $Y_i^j$  are random variables of laws the respective marginals of  $\Pi$ . Using that the cost is symmetric, without loss of generality we assume that  $\Pi$  is symmetric, i.e.,  $\Pi = (\sigma \otimes \sigma)_* \Pi$  for every  $\sigma \in \mathfrak{S}_N$  (the symmetrization is performed in each copy  $E^N$  of  $(E^N)^2$ ). By an obvious adaptation of Lemma A.12 in Appendix A.2.1, the marginals of  $\Pi$  (which are, accordingly, probability measures on  $E^2$ , by considering the product of the  $i^{\text{th}}$  copy

of  $E$  with the  $i^{\text{th}}$  copy of  $E$ ) are all equal. The same is true for the marginals of higher order. It follows that  $Y_i^1$  and  $Y_i^2$  do not depend on  $i$  and thus

$$\begin{aligned} W_p^{[q]}(\mu_1, \mu_2)^p &= N^{p/q} \mathbb{E} d_E(Y_1^1, Y_1^2)^p = N^{p/q} \int_{E^2} d_E(y_1^1, y_1^2)^p d\Pi_{N:1}(y_1^1, y_1^2) \\ &= \left(\frac{N}{k}\right)^{p/q} \mathbb{E} \left( \sum_{i=1}^k d_E(Y_i^1, Y_i^2)^q \right)^{p/q} = \left(\frac{N}{k}\right)^{p/q} \int_{(E^k)^2} d_{E^k}^{[q]}(y^1, y^2)^p d\Pi_{N:k}(y^1, y^2) \end{aligned}$$

for every  $k \in \{1, \dots, N\}$ . The latter quantity is greater than or equal to  $W_p^{[q]}((\mu_1)_{N:k}, (\mu_2)_{N:k})^p$  because  $\Pi_{N:k}$  couples  $(\mu_1)_{N:k}$  and  $(\mu_2)_{N:k}$ . The lemma follows.  $\square$

### A.2.3 A combinatorial lemma towards propagation of chaos

**Lemma A.14.** *Let  $\mu_1, \dots, \mu_N \in \mathcal{P}(E)$ , and let  $\rho \in \mathcal{P}(E^N)$  be defined by*

$$\rho = \mu_1 \otimes \dots \otimes \mu_N.$$

*The symmetrization of  $\rho$  is given by*

$$\rho^s = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \mu_{\sigma(1)} \otimes \dots \otimes \mu_{\sigma(N)}. \quad (103)$$

*The first marginal  $\rho_{N:1}^s \in \mathcal{P}(E)$  of  $\rho^s$  is*

$$\rho_{N:1}^s = \frac{1}{N} \sum_{i=1}^N \mu_i \quad (104)$$

*and, for every  $k \in \{2, \dots, N\}$ , its  $k^{\text{th}}$ -order marginal  $\rho_{N:k}^s \in \mathcal{P}(E)$  is*

$$\rho_{N:k}^s = (1 + \varepsilon_k) (\rho_{N:1}^s)^{\otimes k} - \varepsilon_k \beta_k \quad (105)$$

*where*

$$\varepsilon_k = \frac{N^k (N-k)!}{N!} - 1 \in \left[0, e^{\frac{k^2}{2N}} - 1\right] \quad (106)$$

*and*

$$\beta_k = \frac{1}{\varepsilon_k} \frac{(N-k)!}{N!} \sum \mu_{i_1} \otimes \dots \otimes \mu_{i_k} \in \mathcal{P}(E^k) \quad (107)$$

*where the sum in (107) is taken over all  $k$ -tuples  $(i_1, \dots, i_k) \in \{1, \dots, N\}^k$  for which at least two elements are equal. For every  $p \in [1, +\infty)$ , for every  $k \in \mathbb{N}^*$  such that  $k^2 \leq 2N \ln(1 + \frac{1}{2p})$ , we have*

$$W_p^{[q]}(\rho_{N:k}^s, (\rho_{N:1}^s)^{\otimes k}) \leq 2 \left(\frac{k^2}{N}\right)^{1/p} W_p^{[q]}((\rho_{N:1}^s)^{\otimes k}, \beta_k) \quad (108)$$

*and therefore, assuming moreover that  $\mu_1, \dots, \mu_N \in \mathcal{P}_c(E)$ ,*

$$W_p^{[q]}(\rho_{N:k}^s, (\rho_{N:1}^s)^{\otimes k}) \leq 2k^{1/q} \left(\frac{k^2}{N}\right)^{1/p} \text{diam}_E \left( \bigcup_{i=1}^N \text{supp}(\mu_i) \right). \quad (109)$$

In (108) and (109), the Wasserstein distance  $W_p^{[q]}$  is computed with respect to the  $\ell^q$  distance  $d_E^{[q]}$ . The estimate (109) is used several times in the proofs of our main results, in an instrumental way to express that, for  $N$  large, the  $k^{\text{th}}$ -order marginal  $\rho_{N:k}^s$  of the symmetric measure  $\rho^s$  is close to the tensor power  $(\rho_{N:1}^s)^{\otimes k}$ , with an error that is precisely estimated.

The first part of the lemma, in particular the formulas (104) and (105), are certainly known by experts and can be found, e.g., in [69, Section 3], for Dirac measures.

*Proof.* The formula (103) straightforwardly follows from (89), and the formula (104) follows from Lemma A.12 in Appendix A.2.1 because  $p_*^i \rho = \mu_i$ .

Let us now compute the  $k^{\text{th}}$ -order marginal  $\rho_{N:k}^s$  of  $\rho^s$ , for every  $k \in \{2, \dots, N\}$ . Let  $I_k^N$  be the set of all  $k$ -tuples  $(i_1, \dots, i_k)$  consisting of distinct integers chosen in  $\{1, \dots, N\}$ . We have  $\text{card}(I_k^N) = \frac{N!}{(N-k)!}$ . Denoting by  $\mathfrak{S}_N^{i_1, \dots, i_k}$  the set of all  $\sigma \in \mathfrak{S}_N$  such that  $(\sigma(1), \dots, \sigma(k)) = (i_1, \dots, i_k)$ , we have  $\text{card}(\mathfrak{S}_N^{i_1, \dots, i_k}) = (N-k)!$ . Now, since

$$\sum_{\sigma \in \mathfrak{S}_N} \mu_{\sigma(1)} \otimes \dots \otimes \mu_{\sigma(N)} = \sum_{(i_1, \dots, i_N) \in I_k^N} \mu_{i_1} \otimes \dots \otimes \mu_{i_k} \otimes \sum_{\sigma \in \mathfrak{S}_N^{i_1, \dots, i_k}} \mu_{\sigma(n+1)} \otimes \dots \otimes \mu_{\sigma(N)}$$

we infer that

$$\rho_{N:k}^s = \frac{(N-k)!}{N!} \sum_{(i_1, \dots, i_k) \in I_k^N} \mu_{i_1} \otimes \dots \otimes \mu_{i_k}. \quad (110)$$

Now, writing  $I_k^N = \{1, \dots, N\}^k \setminus (\{1, \dots, N\}^k \setminus I_k^N)$ , we write the sum in (110) as a sum over  $\{1, \dots, N\}^k$  minus a sum over  $\{1, \dots, N\}^k \setminus I_k^N$  (where at least two of the indices are equal). For the first sum, we have

$$\sum_{(i_1, \dots, i_k) \in \{1, \dots, N\}^k} \mu_{i_1} \otimes \dots \otimes \mu_{i_k} = \left( \sum_{i=1}^N \mu_i \right)^{\otimes k} = N^k (\rho_{N:1}^s)^{\otimes k}. \quad (111)$$

We infer from (110) and (111) that

$$\rho_{N:k}^s = \frac{N^k (N-k)!}{N!} (\rho_{N:1}^s)^{\otimes k} - \frac{(N-k)!}{N!} \beta$$

where

$$\beta = \sum_{(i_1, \dots, i_k) \in \{1, \dots, N\}^k \setminus I_k^N} \mu_{i_1} \otimes \dots \otimes \mu_{i_k}$$

is a nonnegative Radon measure of total mass  $|\beta| = \text{card}(\{1, \dots, N\}^k \setminus I_k^N) = N^k - \frac{(N-k)!}{N!}$ . Besides, we have

$$1 \leq \frac{N^k (N-k)!}{N!} = \frac{N^k}{N(N-1) \dots (N-k+1)} = \frac{1}{\prod_{i=1}^{k-1} \left(1 - \frac{i}{N}\right)} \leq e^{\frac{k^2}{2N}}$$

where we have used the inequality

$$\ln \prod_{i=1}^{k-1} \left(1 - \frac{i}{N}\right) = - \sum_{i=1}^{k-1} \ln \left(1 - \frac{i}{N}\right) \geq - \frac{1}{N} \sum_{i=1}^{k-1} i = - \frac{(k-1)k}{2N} \geq - \frac{k^2}{2N}.$$

Therefore, defining  $\varepsilon_k$  by (106) and

$$\beta_k = \frac{1}{\varepsilon_k} \frac{(N-k)!}{N!} \beta \in \mathcal{P}(\mathbb{R}^{dk}),$$

we obtain  $\rho_{N:k}^s = (1 + \varepsilon_k) (\rho_{N:1}^s)^{\otimes k} - \varepsilon_k \beta_k$ , which is (105). Then, applying Lemma A.3 in Appendix A.1.1, using that  $\varepsilon_k < 1$  if  $e^{k^2/2N} - 1 < 1$ , or equivalently,  $k^2 < 2N \ln(2)$ , we obtain that

$$W_p^{[q]}((\rho_{N:1}^s)^{\otimes k}, \rho_{N:k}^s) \leq \frac{\varepsilon_k^{1/p}}{1 - \varepsilon_k^{1/p}} W_p^{[q]}((\rho_{N:1}^s)^{\otimes k}, \beta_k) \quad \text{if } k^2 < 2 \ln(2)N. \quad (112)$$

The estimate (108) is now inferred from (112) as follows: if  $k^2 \leq 2N \ln(1 + \frac{1}{2^p})$  then  $e^{k^2/2N} - 1 \leq \frac{1}{2^p}$ , hence  $\varepsilon_k^{1/p} \leq \frac{1}{2}$  (using (106)) and thus  $\frac{\varepsilon_k^{1/p}}{1 - \varepsilon_k^{1/p}} \leq 2\varepsilon_k^{1/p}$ , and it follows from (112) that

$$W_p^{[q]}((\rho_{N:1}^s)^{\otimes k}, \rho_{N:k}^s) \leq 2(e^{\frac{k^2}{2N}} - 1)^{1/p} W_p^{[q]}((\rho_{N:1}^s)^{\otimes k}, \beta_k).$$

Using the inequality  $\frac{e^x - 1}{x} \leq 1/2^p \ln(1 + \frac{1}{2^p}) \leq 2$  for every  $x \in (0, \ln(1 + \frac{1}{2^p})]$ , we obtain (108).

Let us finally establish (109). Using (111) and (107), which express  $(\rho_{N:1}^s)^{\otimes k}$  and  $\beta_k$  as linear combinations, applying two times Lemma A.2 in Appendix A.1.1 and then Lemma A.7 in Appendix A.1.5, we infer that

$$W_p^{[q]}((\rho_{N:1}^s)^{\otimes k}, \beta_k) \leq \max \left( \sum_{i=1}^k d_E(y_i, y'_i)^q \right)^{1/q} \leq k^{1/q} \text{diam}_E \left( \bigcup_{i=1}^N \text{supp}(\mu_i) \right)$$

where, above, the maximum has been taken over all possible  $y_i, y'_i \in \text{supp}(\mu_i)$ , for  $i \in \{1, \dots, k\}$ . Then, (109) follows from (108) combined with the above inequality.  $\square$

### A.3 Density of empirical measures in the set of probability measures

Let  $E$  be a Polish space, endowed with a distance  $d_E$ . For every  $N \in \mathbb{N}^*$ , let  $Y^N = (y_1^N, \dots, y_N^N) \in E^N$ , and define the *empirical measure*  $\mu_{Y^N}^e \in \mathcal{P}(E)$  by

$$\mu_{Y^N}^e = \frac{1}{N} \sum_{i=1}^N \delta_{y_i^N}.$$

The points  $y_i^N$  are not required to be distinct, so that the empirical measure  $\mu_{Y^N}^e$  can equivalently be defined as a *convex combination with rational coefficients of Dirac masses*. Note that

$$\int_E f d\mu_{Y^N}^e = \frac{1}{N} \sum_{i=1}^N f(y_i^N) \quad \forall f \in \mathcal{C}^0(E).$$

A sequence  $(\mu_j)_{j \in \mathbb{N}^*}$  of  $\mathcal{P}(E)$  converges weakly to  $\mu \in \mathcal{P}(E)$  if  $\int_E f d\mu_j \rightarrow \int_E f d\mu$  as  $j \rightarrow +\infty$  for any  $f \in \mathcal{C}_b(E)$  (narrow convergence), where  $\mathcal{C}_b(E)$  is the Banach space of bounded functions on  $E$ .

**Lemma A.15.** *When  $E$  is compact, the set  $\{\mu_{Y^N}^e \mid N \in \mathbb{N}^*, Y^N \in E^N\}$  is weakly dense in  $\mathcal{P}(E)$ . In other words, any probability measure on  $E$  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., [52, Lemma 7]). Let us anyway recall a proof. The set  $\mathcal{P}(E)$  is convex and weak star compact, and its extreme points are Dirac masses. The Krein-Milman theorem implies that any  $\mu \in \mathcal{P}(E)$  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.  $\square$

Recall that the Wasserstein distance  $W_p$  metrizes the weak convergence in  $\mathcal{P}_p(E)$  (which also entails the convergence of first moments), for any  $p \in [1, +\infty)$ . We have then the following variant of the above lemma (see [78, Theorem 6.18]).

**Lemma A.16.** *The set  $\{\mu_{Y^N}^e \mid N \in \mathbb{N}^*, Y^N \in E^N\}$  is dense in  $\mathcal{P}_p(E)$  for the Wasserstein distance  $W_p$ . In other words, any  $\mu \in \mathcal{P}_p(E)$  is the limit of a sequence of empirical measures for the Wasserstein distance  $W_p$ .*

*Proof.* It suffices to consider  $R > 0$  sufficiently large such that  $\int_{E \setminus B(y_0, R)} d_E(y_0, y)^p d\mu(y) < \varepsilon$ , for  $\varepsilon > 0$  small enough, so that the argument can be performed in the compact set  $\overline{B}(y_0, R)$ , and the statement readily follows (see also [73, Chap. 5]).  $\square$

There exist a number of results in the literature quantifying the convergence of empirical measures  $\mu_{Y^N}^e$  towards  $\mu \in \mathcal{P}(E)$  and providing rates of convergence, most in a probabilistic context, like [31] where  $Y$  consists of  $N$  random variables having the same distribution as  $\mu$ . In the result hereafter,  $Y$  is deterministic and the rate of convergence is the one obtained by Riemann integration.

**Lemma A.17.** *Let  $\mu \in \mathcal{P}_c(E)$  and let  $N \in \mathbb{N}^*$ . We assume that there exists a family of tagged partitions of  $\text{supp}(\mu)$  associated with  $\mu$  (see (27)), i.e., for every  $N \in \mathbb{N}^*$  there exists a partition of  $\text{supp}(\mu) = \cup_{i=1}^N F_i^N$  such that all subsets  $F_i^N$  are  $\mu$ -measurable, pairwise disjoint,  $\mu(F_i^N) = \frac{1}{N}$ , and satisfy  $\text{diam}_E(F_i^N) \leq \frac{C_E}{N^r}$  for some  $C_E > 0$  not depending on  $N$ , and a  $N$ -tuple  $Y^N = (y_1^N, \dots, y_N^N) \in E^N$  such that  $y_i^N \in F_i^N$  for every  $i \in \{1, \dots, N\}$ . Then*

$$W_1(\mu_{Y^N}^e, \mu) \leq \frac{C_E}{N^r}$$

and thus also, using (23),  $W_p(\mu_{Y^N}^e, \mu) \leq \text{diam}_E(\text{supp}(\mu))^{1-1/p} \frac{C_E^{1/p}}{N^{r/p}}$ , for any  $p \in [1, +\infty)$ .

Note that, when  $E$  is a finite-dimensional manifold,  $r = 1/\dim(E)$ .

When one wants that the assumption on the tagged partition be satisfied for any  $N \in \mathbb{N}^*$ , this requires that the mass of  $\mu$  be quite well uniformly distributed; for instance it is satisfied if  $\mu$  is absolutely continuous with respect to a Lebesgue measure with a density that is bounded above and below on  $\text{supp}(\mu)$ . This result is quite obvious and has nothing to see with much deeper and general results like those of [31].

*Proof.* For every  $i \in \{1, \dots, N\}$ , we have  $\int_{F_i^N} f(y_i^N) d\mu(y) = f(y_i^N) \mu(F_i^N) = \frac{1}{N} f(y_i^N)$  because  $\mu(F_i^N) = \frac{1}{N}$  and thus, for every  $f \in \text{Lip}(E)$  such that  $\text{Lip}(f) \leq 1$ ,

$$\begin{aligned} \left| \int_E f d(\mu - \mu_{Y^N}^e) \right| &= \left| \sum_{i=1}^N \int_{F_i^N} f(y) d\mu(y) - \frac{1}{N} \sum_{i=1}^N f(y_i^N) \right| = \left| \sum_{i=1}^N \int_{F_i^N} (f(y) - f(y_i^N)) d\mu(y) \right| \\ &\leq \sum_{i=1}^N \int_{F_i^N} |f(y) - f(y_i^N)| d\mu(y) \leq \sum_{i=1}^N \int_{F_i^N} d_E(y, y_i^N) d\mu(y) \leq \sum_{i=1}^N \mu(F_i^N) \text{diam}_E(F_i^N) \leq \frac{C_E}{N^r} \end{aligned}$$

and the conclusion follows by taking the supremum over all  $f$ .  $\square$

## A.4 Convergence of empirical and semi-empirical measures

Let  $(\Omega, d_\Omega)$  be a complete metric space and let  $\nu \in \mathcal{P}_c(\Omega)$ . We assume that there exists a family of tagged partitions  $(\mathcal{A}^N, X^N)$  of  $\text{supp}(\nu)$  associated with  $\nu$  satisfying (27) (see Section 1.5), with

$\mathcal{A}^N = (\Omega_1^N, \dots, \Omega_N^N)$  and  $X^N = (x_1^N, \dots, x_N^N)$ . We define the *empirical measure*  $\nu_{X^N}^e \in \mathcal{P}(\Omega)$  by

$$\nu_{X^N}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N}.$$

Note that, when  $\Omega$  is a  $n$ -dimensional manifold, one has  $r = 1/n$  in (27).

#### A.4.1 Convergence of empirical measures on $\Omega$

**Lemma A.18.** • *Let  $f$  be a bounded and  $\nu$ -almost everywhere continuous (i.e.,  $\nu$ -Riemann integrable) function on  $\Omega$ , of compact support. Then*

$$\int_{\Omega} f d(\nu - \nu_{X^N}^e) = \int_{\Omega} f d\nu - \frac{1}{N} \sum_{i=1}^N f(x_i^N) = o(1) \quad (113)$$

as  $N \rightarrow +\infty$ . As a consequence,  $\nu_{X^N}^e$  converges weakly to  $\nu$  as  $N \rightarrow +\infty$ ; equivalently,  $W_p(\nu_{X^N}^e, \nu) = o(1)$  as  $N \rightarrow +\infty$ .

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

$$\left| \int_{\Omega} f d\nu - \frac{1}{N} \sum_{i=1}^N f(x_i^N) \right| \leq \frac{C_{\Omega}^{\alpha}}{N^{r\alpha}} \text{Hol}_{\alpha}(f) \quad (114)$$

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

$$W_1(\nu_{X^N}^e, \nu) \leq \frac{C_{\Omega}}{N^r} \quad (115)$$

and thus also, using (23),  $W_p(\nu_{X^N}^e, \nu) \leq \text{diam}_{\Omega}(\text{supp}(\nu))^{1-1/p} \frac{C_{\Omega}^{1/p}}{N^{r/p}}$ , for any  $p \in [1, +\infty)$ .

*Proof.* In the first item, (113) follows from the theorem of convergence of Riemann sums, as already recalled in (29). Interpreted in terms of the empirical measure  $\nu_{X^N}^e$ , this means that  $\nu_{X^N}^e$  converges weakly to  $\nu$  as  $N \rightarrow +\infty$ . In accordance with the Portmanteau theorem (see, e.g., [9, Chapter 1, Section 2, Theorem 2.1]), since  $W_p$  metrizes the weak convergence, we have  $W_p(\nu_{X^N}^e, \nu) = o(1)$  as  $N \rightarrow +\infty$  since  $\text{supp}(\nu)$  is compact.

Writing  $\int_{\Omega} f d\nu = \sum_{i=1}^N \int_{\Omega_i^N} f d\nu$  and using that  $\nu(\Omega_i^N) = \frac{1}{N}$  (thus  $\frac{1}{N} f(x_i^N) = \int_{\Omega_i^N} f(x_i^N) d\nu(x)$ ) and that  $\text{diam}_{\Omega}(\Omega_i^N) \leq \frac{C_{\Omega}}{N^r}$  (see (27)), we have

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

which gives (114). Taking  $\alpha = 1$ , (115) follows by the definition (24) of  $W_1$ .  $\square$

### A.4.2 Convergence of semi-empirical measures

Let  $d \in \mathbb{N}^*$ . Let  $\mu \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ , disintegrated as  $\mu = \int_{\Omega} \mu_x d\nu(x)$  with respect to its marginal  $\nu = \pi_* \mu$  on  $\Omega$ . We define the *semi-empirical measure*  $\mu_{X^N}^{se} \in \mathcal{P}(\Omega \times \mathbb{R}^d)$  by

$$\mu_{X^N}^{se} = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \mu_{x_i^N} = \int_{\Omega} \mu_x d\nu_{X^N}^e(x).$$

Its marginal on  $\Omega$  is the empirical measure  $\nu_{X^N}^e$ . In other words, the disintegration of  $\mu_{X^N}^{se}$  with respect to  $\nu_{X^N}^e$  is the family of probability measures given by  $\mu_{x_i^N}$  when  $x = x_i^N$  for some  $i \in \{1, \dots, N\}$  and 0 otherwise.

**Lemma A.19.**

- We assume that  $x \mapsto \mu_x$  is  $\nu$ -almost everywhere continuous for the Wasserstein distance  $W_1$  (equivalently,  $W_p$ ). Let  $f$  be a bounded and  $\mu$ -almost everywhere continuous (i.e.,  $\mu$ -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

$$\int_{\Omega \times \mathbb{R}^d} f d(\mu - \mu_{X^N}^{se}) = o(1) \quad (116)$$

as  $N \rightarrow +\infty$ . As a consequence,  $\mu_{X^N}^{se}$  converges weakly to  $\mu$ ; equivalently,  $W_p(\mu_{X^N}^{se}, \mu) = o(1)$  as  $N \rightarrow +\infty$ .

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

$$\left| \int_{\Omega \times \mathbb{R}^d} f d(\mu - \mu_{X^N}^{se}) \right| \leq \frac{(L+1)C_{\Omega}}{N^r} \text{Lip}(f) \quad (117)$$

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

$$W_1(\mu_{X^N}^{se}, \mu) \leq \frac{(L+1)C_{\Omega}}{N^r}, \quad (118)$$

and thus also, using (23),  $W_p(\mu_{X^N}^{se}, \mu) \leq \text{diam}_{\Omega \times \mathbb{R}^d}(\text{supp}(\mu))^{1-1/p} \frac{((L+1)C_{\Omega})^{1/p}}{N^{r/p}}$ , for any  $p \in [1, +\infty)$ .

*Proof.* Let  $f : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}$  be a bounded and  $\mu$ -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  $\Omega$ , and

$$\begin{aligned} |F(x) - F(x')| &\leq \int_{\mathbb{R}^d} |f(x, \xi) - f(x', \xi)| d\mu_x(\xi) + \left| \int_{\mathbb{R}^d} f(x', \xi) d(\mu_x - \mu_{x'})(\xi) \right| \\ &\leq \int_{\mathbb{R}^d} |f(x, \xi) - f(x', \xi)| d\mu_x(\xi) + W_1(\mu_x, \mu_{x'}) \text{Lip}(f(x', \cdot)) \end{aligned} \quad (119)$$

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

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

$$\int_{\Omega \times \mathbb{R}^d} f d(\mu - \mu_{X^N}^{se}) = \int_{\Omega} F d(\nu - \nu_{X^N}^e) = \int_{\Omega} F d\nu - \frac{1}{N} \sum_{i=1}^N F(x_i^N) = o(1)$$

as  $N \rightarrow +\infty$  by convergence of Riemann sums ( $f$  and thus  $F$  being fixed), which gives (116).



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

$$|F(x) - F(x')| \leq \text{Lip}(f) d_\Omega(x, x') + W_1(\mu_x, \mu_{x'}) \text{Lip}(f) \leq \text{Lip}(f)(1 + L) d_\Omega(x, x')$$

and thus, using Lemma A.18, that  $\int_\Omega F d(\nu - \nu_{X^N}^e) \leq \frac{C_\Omega}{N^r} \text{Lip}(F)$ , whence (117) and (118).  $\square$

**Remark A.3.** In the first item of Lemma A.19, 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  $|f(x, \xi)| \leq |f(x, 0)| + \text{Lip}(f(x, \cdot))|\xi|$ , we infer that  $F$  is bounded. The rest of the proof is the same.

## A.5 Discrepancy between empirical and $\nu$ -monokinetic measures

Recall that:

- given any  $X^N = (x_1^N, \dots, x_N^N) \in \Omega^N$  and any  $\Xi^N = (\xi_1^N, \dots, \xi_N^N) \in \mathbb{R}^{dN}$ , the empirical measure  $\mu_{(X^N, \Xi^N)}^e$  on  $\Omega \times \mathbb{R}^d$  is defined by (58);
- given any  $\nu \in \mathcal{P}(\Omega)$  and any measurable function  $y : \Omega \rightarrow \mathbb{R}^d$ , the  $\nu$ -monokinetic measure  $\mu_y^\nu$  on  $\Omega \times \mathbb{R}^d$  is defined by (78).

**Lemma A.20.** Let  $\nu \in \mathcal{P}(\Omega)$  and let  $(\mathcal{A}^N, X^N)_{N \in \mathbb{N}^*}$  be a family of tagged partitions associated with  $\nu$  (see (27)), with  $\mathcal{A}^N = (\Omega_1^N, \dots, \Omega_N^N)$  and  $X^N = (x_1^N, \dots, x_N^N)$ .

- (i) Let  $y \in \text{Lip}(\Omega, \mathbb{R}^d)$ . For every  $N \in \mathbb{N}^*$ , taking  $\Xi^N = (\xi_1^N, \dots, \xi_N^N)$  with  $\xi_i^N = y(x_i^N)$  for every  $i \in \{1, \dots, N\}$ , we have

$$\left| \left\langle \mu_y^\nu - \mu_{(X^N, \Xi^N)}^e, f \right\rangle \right| \leq \frac{C_\Omega}{N^r} \text{Lip}(x \mapsto f(x, y(x))) \quad \forall f \in \text{Lip}_c(\Omega \times \mathbb{R}^d).$$

- (ii) For every  $N \in \mathbb{N}^*$ , let  $\Xi^N = (\xi_1^N, \dots, \xi_N^N) \in \mathbb{R}^{dN}$ . Defining the piecewise continuous function

$$y^N(x) = \sum_{i=1}^N \xi_i^N \mathbf{1}_{\Omega_i^N}(x) \quad \forall x \in \Omega,$$

so that  $y^N(x_i^N) = \xi_i^N$  for every  $i \in \{1, \dots, N\}$ , we have

$$\left| \left\langle \mu_{y^N}^\nu - \mu_{(X^N, \Xi^N)}^e, f \right\rangle \right| \leq \frac{C_\Omega}{N^r} \max_{1 \leq i \leq N} \text{Lip}(f(\cdot, \xi_i^N)) \quad \forall f \in \text{Lip}_c(\Omega \times \mathbb{R}^d).$$

*Proof.* Let us prove (i). We have  $\langle \mu_y^\nu, f \rangle = \int_\Omega f(x, y(x)) d\nu(x) = \sum_{i=1}^N \int_{\Omega_i^N} f(x, y(x)) d\nu(x)$  and (using that  $\nu(\Omega_i^N) = \frac{1}{N}$ )

$$\left\langle \mu_{(X^N, \Xi^N)}^e, f \right\rangle = \frac{1}{N} \sum_{i=1}^N f(x_i^N, y(x_i^N)) = \sum_{i=1}^N \int_{\Omega_i^N} f(x_i^N, y(x_i^N)) d\nu(x)$$

hence

$$\begin{aligned} \left| \left\langle \mu_y^\nu - \mu_{(X^N, \Xi^N)}^e, f \right\rangle \right| &\leq \sum_{i=1}^N \int_{\Omega_i^N} |f(x, y(x)) - f(x_i^N, y(x_i^N))| d\nu(x) \\ &\leq \text{Lip}(x \mapsto f(x, y(x))) \sum_{i=1}^N \int_{\Omega_i^N} d_\Omega(x, x_i^N) d\nu(x) \end{aligned}$$

and (i) follows because  $\int_{\Omega_i^N} d_{\Omega}(x, x_i^N) d\nu(x) \leq \nu(\Omega_i^N) \text{diam}_{\Omega}(\Omega_i^N) \leq \frac{C_{\Omega}}{N^{1+r}}$  (using (27)).

The estimate of (ii) is proved similarly: we have  $\langle \mu_{y^N}^{\nu}, f \rangle = \sum_{i=1}^N \int_{\Omega_i^N} f(x, \xi_i^N) d\nu(x)$  and thus

$$\begin{aligned} \left| \langle \mu_{y^N}^{\nu} - \mu_{(X^N, \Xi^N)}^e, f \rangle \right| &\leq \sum_{i=1}^N \int_{\Omega_i^N} |f(x, \xi_i^N) - f(x_i^N, \xi_i^N)| d\nu(x) \\ &\leq \sum_{i=1}^N \text{Lip}(f(\cdot, \xi_i^N)) \int_{\Omega_i^N} d_{\Omega}(x, x_i^N) d\nu(x) \end{aligned}$$

and (ii) follows.  $\square$

**Remark A.4.** 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  $\Omega_i^N$ . In particular, they may be discontinuous at the boundary of  $\Omega_i^N$ .

With that remark, we recover (ii) as a consequence of (i).

## A.6 Mean field and variance

Let  $\mu \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$  be arbitrary. Recall that the mean field  $\mathcal{X}[\mu](t, x, \xi)$  is defined by (44), which is the expectation of  $G(t, x, \cdot, \xi, \cdot)$  for the measure  $\mu$ , performed with respect to  $(x', \xi') \in \Omega \times \mathbb{R}^d$ :

$$\mathcal{X}[\mu](t, x, \xi) = \int_{\Omega \times \mathbb{R}^d} G(t, x, x', \xi, \xi') d\mu(x', \xi') = \mathbb{E}^{\mu} G(t, x, \cdot, \xi, \cdot)$$

Given any  $t \in \mathbb{R}$ , any  $x, x' \in \Omega$  and any  $\xi, \xi' \in \mathbb{R}^d$ , we set

$$e_t[\mu](x, x', \xi, \xi') = G(t, x, x', \xi, \xi') - \mathcal{X}[\mu](t, x, \xi).$$

Of course, we have  $\mathbb{E}^{\mu} e_t[\mu](x, \cdot, \xi, \cdot) = 0$  and thus also

$$\mathbb{E}^{\mu \otimes \mu} e_t[\mu] = 0.$$

This naturally leads to consider the variance of  $e_t[\mu]$  with respect to  $\mu \otimes \mu$ :

$$\text{Var}(e_t[\mu]) = \mathbb{E}^{\mu \otimes \mu} \|e_t\|^2 = \int_{\Omega^2 \times \mathbb{R}^{2d}} \|G(t, x, x', \xi, \xi') - \mathcal{X}[\mu](t, x, \xi)\|^2 d\mu(x', \xi') d\mu(x, \xi).$$

Note that

$$\text{Var}(e_t[\mu]) \leq 4 \|G(t, \cdot, \cdot, \cdot, \cdot)_{|\text{supp}(\mu)^2}\|_{\mathcal{C}^0}^2. \quad (120)$$

Let  $N \in \mathbb{N}^*$  be fixed. Recall that the particle (time-dependent) vector field  $Y^N = (Y_1^N, \dots, Y_N^N)$  is defined by (10) with  $Y_i^N$  defined by (11), i.e.,  $Y_i^N(t, X, \Xi) = \frac{1}{N} \sum_{j=1}^N G(t, x_i, x_j, \xi_i, \xi_j)$ , where we use the notations  $X = (x_1, \dots, x_N) \in \Omega^N$  and  $\Xi = (\xi_1, \dots, \xi_N) \in (\mathbb{R}^d)^N$ .

**Lemma A.21.** *We assume that the norm  $\|\cdot\|$  on  $\mathbb{R}^d$  is induced by a scalar product  $\langle \cdot, \cdot \rangle$  on  $\mathbb{R}^d$ . For every  $i \in \{1, \dots, N\}$  we have*

$$\int_{\Omega^N \times \mathbb{R}^{dN}} \|Y_i^N(t, X, \Xi) - \mathcal{X}[\mu](t, x_i, \xi_i)\|^2 d\mu^{\otimes N}(X, \Xi) = \frac{1}{N} \text{Var}(e_t[\mu]) \leq \frac{4}{N} \|G(t, \cdot, \cdot, \cdot, \cdot)_{|\text{supp}(\mu)^2}\|_{\mathcal{C}^0}^2.$$

*Proof of Lemma A.21.* By definition of  $e_t$ , we have, for every  $i \in \{1, \dots, N\}$ ,

$$Y_i^N(\tau, X, \Xi) - \mathcal{X}[\mu](t, x_i, \xi_i) = \frac{1}{N} \sum_{j=1}^N e_t[\mu](x_i, x_j, \xi_i, \xi_j).$$

Therefore,

$$\begin{aligned} & \int_{\Omega^N \times \mathbf{R}^{dN}} \|Y_i^N(\tau, X, \Xi) - \mathcal{X}[\mu](t, x_i, \xi_i)\|^2 d\mu^{\otimes N}(X, \Xi) \\ &= \frac{1}{N^2} \int_{\Omega^N \times \mathbf{R}^{dN}} \sum_{j=1}^N \|e_t[\mu](x_i, x_j, \xi_i, \xi_j)\|^2 d\mu^{\otimes N}(X, \Xi) \\ & \quad + \frac{1}{N^2} \int_{\Omega^N \times \mathbf{R}^{dN}} \sum_{\substack{j,k=1 \\ j \neq k}}^N \langle e_t[\mu](x_i, x_j, \xi_i, \xi_j), e_t[\mu](x_i, x_k, \xi_i, \xi_k) \rangle d\mu^{\otimes N}(X, \Xi). \end{aligned} \quad (121)$$

The first term at the right-hand side of (121) is equal to

$$\frac{1}{N} \int_{\Omega^2 \times \mathbf{R}^{2d}} \|e_t[\mu](x, x', \xi, \xi')\|^2 d\mu(x', \xi') d\mu(x, \xi) = \frac{1}{N} \text{Var}(e_t[\mu]). \quad (122)$$

The second term at the right-hand side of (121) is equal to

$$\begin{aligned} & \frac{N^2 - N}{N^2} \int_{\Omega^3 \times \mathbf{R}^{3d}} \langle e_t[\mu](x, x', \xi, \xi'), e_t[\mu](x, x'', \xi, \xi'') \rangle d\mu(x, \xi) d\mu(x', \xi') d\mu(x'', \xi'') \\ &= \frac{N^2 - N}{N^2} \int_{\Omega \times \mathbf{R}^d} \left\| \int_{\Omega \times \mathbf{R}^d} e_t[\mu](x, x', \xi, \xi') d\mu(x', \xi') \right\|^2 d\mu(x, \xi) = 0 \end{aligned} \quad (123)$$

because the expectation of  $e_t[\mu](x_i, \cdot, \xi_i, \cdot)$  is equal to 0. The lemma is proved, using (120).  $\square$

Although Lemma A.21 is not used as such in this article, we believe that it has its own interest. Actually, in the proof of Theorem 4.2 (in Appendix B.5), we will need the following result, in the spirit of Lemma A.21 but more technical.

**Lemma A.22.** *As in Lemma A.21, we assume that the norm  $\|\cdot\|$  on  $\mathbf{R}^d$  is induced by a scalar product  $\langle \cdot, \cdot \rangle$  on  $\mathbf{R}^d$ . Let  $\bar{X} = (\bar{x}_1, \dots, \bar{x}_N) \in \Omega^N$  be arbitrary, and let*

$$\rho = \delta_{\bar{x}_1} \otimes \dots \otimes \delta_{\bar{x}_N} \otimes \mu_{\bar{x}_1} \otimes \dots \otimes \mu_{\bar{x}_N}.$$

Then

$$\begin{aligned} M(t) &= \left( \int_{\Omega^N \times \mathbf{R}^{dN}} \left( \sum_{i=1}^N \|Y_i^N(t, X, \Xi) - \mathcal{X}[\mu](t, x_i, \xi_i)\| \right)^2 d\rho(X, \Xi) \right)^{1/2} \\ &\leq 2 \|G(t, \cdot, \cdot, \cdot, \cdot)_{|\text{supp}(\mu)^2}\|_{\mathcal{C}^{0,1}} \left( \sqrt{N} \sqrt{1 + 70 \text{diam}_{\Omega \times \mathbf{R}^d}(\text{supp}(\mu))} + N \sqrt{5 W_1(\mu, \mu_{\bar{X}}^{se})} \right) \end{aligned} \quad (124)$$

where  $\mu_{\bar{X}}^{se} = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{x}_i} \otimes \mu_{\bar{x}_i} = \rho_{N;1}^s$  (semi-empirical measure).

*Proof.* As a first remark, we note that, since the function inside the integral at the left-hand side of the inequality (124) is symmetric, we can replace  $\rho$  by the symmetrization  $\rho^s$  in the integral (indeed, when  $F : \Omega^N \times \mathbb{R}^{dN} \rightarrow \mathbb{R}$  is symmetric, we have  $\int F d\rho = \int F d\rho^s$ ). As a second remark, since  $M(t)$  is defined as the  $L^2$  norm of a sum, we infer from the triangular inequality that

$$\begin{aligned} M(t) &\leq \sum_{i=1}^N \left( \int_{\Omega^N \times \mathbb{R}^{dN}} \|Y_i^N(t, X, \Xi) - \mathcal{X}[\mu](t, x_i, \xi_i)\|^2 d\rho^s(X, \Xi) \right)^{1/2} \\ &\leq N \max_{1 \leq i \leq N} \sqrt{I_i(t)} \quad \text{with} \quad I_i(t) = \int_{\Omega^N \times \mathbb{R}^{dN}} \|Y_i^N(t, X, \Xi) - \mathcal{X}[\mu](t, x_i, \xi_i)\|^2 d\rho^s(X, \Xi) \end{aligned}$$

Note that it was important to symmetrize the measure  $\rho$  *before* applying the triangular inequality. Let us now estimate  $I_i(t)$ , for any fixed  $i \in \{1, \dots, N\}$ . We cannot apply directly Lemma A.21 because in the integral  $I_i(t)$  the integration is performed with respect to  $\rho^s$ , and not with respect to  $\mu^{\otimes N}$ . However, following the proof of Lemma A.21, we expand  $I_i(t)$  similarly as in (121); replacing  $\mu^{\otimes N}$  by  $\rho^s$  and thus the second-order marginal  $\rho_{N:2}^s$  and third-order marginal  $\rho_{N:3}^s$  appear. Note that, by Lemma A.12 in Appendix A.2, since  $\rho^s$  is symmetric all its second-order (resp., third-order) marginals on the various copies of  $(\Omega \times \mathbb{R}^d)^2$  (resp., of  $(\Omega \times \mathbb{R}^d)^2$ ) are equal. We obtain

$$I_i(t) = \frac{1}{N} \int_{\Omega^2 \times \mathbb{R}^{2d}} \|e_t[\mu]\|^2 d\rho_{N:2}^s + \frac{N^2 - N}{N^2} \int_{\Omega^3 \times \mathbb{R}^{3d}} F_t[\mu] d\rho_{N:3}^s \quad (125)$$

with

$$F_t[\mu](x, x', x'', \xi, \xi', \xi'') = \langle e_t[\mu](x, x', \xi, \xi'), e_t[\mu](x, x'', \xi, \xi'') \rangle$$

To estimate the first term at the right-hand side of (125), we observe that (using the definition (24) of the Wasserstein distance  $W_1$ )

$$\begin{aligned} \int_{\Omega^2 \times \mathbb{R}^{2d}} \|e_t[\mu]\|^2 d(\rho_{N:2}^s - \mu^{\otimes 2}) &\leq \text{Lip}(\|e_t[\mu]\|_{\text{supp}(\mu)^2}) W_1^{[1]}(\rho_{N:2}^s, \mu^{\otimes 2}) \\ &\leq 4 \|G(t, \cdot, \cdot, \cdot, \cdot)_{|\text{supp}(\mu)^2}\|_{\mathcal{C}^{0,1}}^2 W_1^{[1]}(\rho_{N:2}^s, \mu^{\otimes 2}) \end{aligned}$$

(the choice of  $q = 1$ , above, has little importance; other choices would change the constant 4, see Lemma A.1) and that, using (120) and (122),

$$\int_{\Omega^2 \times \mathbb{R}^{2d}} \|e_t[\mu]\|^2 d\mu^{\otimes 2} = \text{Var}(e_t[\mu]) \leq 4 \|G(t, \cdot, \cdot, \cdot, \cdot)_{|\text{supp}(\mu)^2}\|_{\mathcal{C}^0}^2 \leq 4 \|G(t, \cdot, \cdot, \cdot, \cdot)_{|\text{supp}(\mu)^2}\|_{\mathcal{C}^{0,1}}^2.$$

To estimate the second term at the right-hand side of (125), similarly, we observe that

$$\begin{aligned} \int_{\Omega^3 \times \mathbb{R}^{3d}} F_t[\mu] d(\rho_{N:3}^s - \mu^{\otimes 3}) &\leq \text{Lip}(F_t[\mu]_{|\text{supp}(\mu)^2}) W_1^{[1]}(\rho_{N:3}^s, \mu^{\otimes 3}) \\ &\leq 4 \|G(t, \cdot, \cdot, \cdot, \cdot)_{|\text{supp}(\mu)^2}\|_{\mathcal{C}^{0,1}}^2 W_1^{[1]}(\rho_{N:3}^s, \mu^{\otimes 3}) \end{aligned}$$

and that, as in (123),

$$\int_{\Omega^3 \times \mathbb{R}^{3d}} F_t[\mu] d\mu^{\otimes 3} = 0.$$

It follows that

$$I_i(t) \leq 4 \|G(t, \cdot, \cdot, \cdot, \cdot)_{|\text{supp}(\mu)^2}\|_{\mathcal{C}^{0,1}}^2 \left( \frac{1}{N} + W_1^{[1]}(\rho_{N:2}^s, \mu^{\otimes 2}) + W_1^{[1]}(\rho_{N:3}^s, \mu^{\otimes 3}) \right).$$

Now, applying Lemma A.14 (Appendix A.2.3), we infer from (104) that  $\rho_{N:1}^s = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{x}_i} \otimes \mu_{\bar{x}_i} = \mu_{\bar{X}}^{se}$  (semi-empirical measure) and from (109) (taking  $k = 2, 3$ ) that

$$W_1^{[1]}(\rho_{N:2}^s, (\mu_{\bar{X}}^{se})^{\otimes 2}) + W_1^{[1]}(\rho_{N:3}^s, (\mu_{\bar{X}}^{se})^{\otimes 3}) \leq \frac{70}{N} \text{diam}_{\Omega \times \mathbf{R}^d}(\text{supp}(\mu))$$

and thus, using the triangular inequality,

$$\sum_{k=2,3} W_1^{[1]}(\rho_{N:k}^s, \mu^{\otimes k}) \leq \frac{70}{N} \text{diam}_{\Omega \times \mathbf{R}^d}(\text{supp}(\mu)) + \sum_{k=2,3} W_1^{[1]}(\mu^{\otimes k}, (\mu_{\bar{X}}^{se})^{\otimes k}).$$

Applying Lemma A.6 in Appendix A.1.4, we have

$$W_1^{[1]}(\mu^{\otimes k}, (\mu_{\bar{X}}^{se})^{\otimes k}) \leq k W_1(\mu, \mu_{\bar{X}}^{se}).$$

Finally,

$$I_i(t) \leq 4 \|G(t, \cdot, \cdot, \cdot, \cdot)_{|\text{supp}(\mu)^2}\|_{\mathcal{C}^{0,1}}^2 \left( \frac{1}{N} + \frac{70}{N} \text{diam}_{\Omega \times \mathbf{R}^d}(\text{supp}(\mu)) + 5 W_1(\mu, \mu_{\bar{X}}^{se}) \right)$$

and the estimate (124) follows.  $\square$

The results of that section can be compared with some considerations done in [36, 58, 59].

## B Proofs

### B.1 Proof of Theorem 2.2

We start by proving the second item of Theorem 2.2. Hence, we assume that  $G$  is locally  $\alpha$ -Hölder continuous with respect to  $(x, x', \xi, \xi')$  (uniformly with respect to  $t$  on any compact).

**Lemma B.1.** *Let  $x, x' \in \Omega$  be arbitrary. We have*

$$\|y(t, x) - y(t, x')\| \leq e^{tL_y(t)} (\|y^0(x) - y^0(x')\| + d_\Omega(x, x')^\alpha) \quad \forall t \geq 0. \quad (126)$$

*Proof of Lemma B.1.* 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

$$\begin{aligned} \partial_t y(t, x) - \partial_t y(t, x') &= \int_\Omega G(t, x, x'', y(t, x), y(t, x'')) d\nu(x'') - \int_\Omega G(t, x', x'', y(t, x), y(t, x'')) d\nu(x'') \\ &\quad + \int_\Omega G(t, x', x'', y(t, x), y(t, x'')) d\nu(x'') - \int_\Omega G(t, x', x'', y(t, x'), y(t, x'')) d\nu(x'') \end{aligned} \quad (127)$$

and using the definition of  $L_y(t)$  we obtain

$$\|\partial_t(y(t, x) - y(t, x'))\| \leq L_y(t) (d_\Omega(x, x')^\alpha + \|y(t, x) - y(t, x')\|)$$

and (126) follows by integration (noting that  $\tau \mapsto L_y(\tau)$  is nondecreasing).  $\square$

**Remark B.1.** If  $G$  does not depend on  $(x, x')$  then one can remove the term  $d_\Omega(x, x')^\alpha$  in (126).

By assumption,  $\|y^0(x) - y^0(x')\| \leq \text{Hol}_\alpha(y^0)d_\Omega(x, x')^\alpha$  for all  $x, x' \in \Omega$ , hence, using (126) in Lemma B.1 we infer that  $y(t, \cdot)$  is  $\alpha$ -Hölder continuous and (35) follows.

Let us establish (36). We set  $r_i^N(t) = y(t, x_i^N) - \xi_i^N(t)$ , for  $i = 1, \dots, N$ . By definition, we have

$$\dot{r}_i^N(t) = \frac{1}{N} \sum_{j=1}^N (G(t, x_i^N, x_j^N, y(t, x_i^N), y(t, x_j^N)) - G(t, x_i^N, x_j^N, \xi_i^N(t), \xi_j^N(t))) + \epsilon_i^N(t) \quad (128)$$

where

$$\epsilon_i^N(t) = \int_\Omega G(t, x_i^N, x', y(t, x_i^N), y(t, x')) d\nu(x') - \frac{1}{N} \sum_{j=1}^N G(t, x_i^N, x_j^N, y(t, x_i^N), y(t, x_j^N)) \quad (129)$$

with  $r_i^N(0) = 0$ , for every  $i \in \{1, \dots, N\}$ . On the one part, we have

$$\|G(t, x_i^N, x_j^N, y(t, x_i^N), y(t, x_j^N)) - G(t, x_i^N, x_j^N, \xi_i^N(t), \xi_j^N(t))\| \leq L_y^N(t)(\|r_i^N(t)\| + \|r_j^N(t)\|) \quad (130)$$

where  $L_y^N(t)$  is defined by (38). On the other part, using (114) in Lemma A.18 (see Appendix A.4), we have

$$\|\epsilon_i^N(t)\| \leq \frac{C_\Omega^\alpha}{N r^\alpha} \text{Hol}_\alpha(x' \mapsto G(t, x_i^N, x', y(t, x_i^N), y(t, x'))) \quad (131)$$

and we claim that

$$\text{Hol}_\alpha(x' \mapsto G(t, x_i^N, x', y(t, x_i^N), y(t, x'))) \leq L_y(t)(1 + e^{tL_y(t)}(\text{Hol}_\alpha(y^0) + 1)). \quad (132)$$

Indeed, writing for short  $g(x', y(t, x')) = G(t, x_i^N, x', y(t, x_i^N), y(t, x'))$ , we have

$$\begin{aligned} & \|g(x'_1, y(t, x'_1)) - g(x'_2, y(t, x'_2))\| \\ & \leq \|g(x'_1, y(t, x'_1)) - g(x'_2, y(t, x'_1))\| + \|g(x'_2, y(t, x'_1)) - g(x'_2, y(t, x'_2))\| \\ & \leq L_y(t)d_\Omega(x'_1, x'_2)^\alpha + L_y(t)\|y(t, x'_1) - y(t, x'_2)\| \\ & \leq L_y(t)d_\Omega(x'_1, x'_2)^\alpha + L_y(t)\text{Hol}_\alpha(y(t, \cdot))d_\Omega(x'_1, x'_2)^\alpha \end{aligned}$$

and (132) follows by using (35). Finally, setting  $R^N(t) = (r_1^N(t), \dots, r_N^N(t))$ , noting that  $L_y(t) \leq L_y^N(t)$ , we infer from (128), (130), (131) and (132) that

$$\frac{d}{dt}\|R^N(t)\|_\infty \leq \|\dot{R}^N(t)\|_\infty \leq L_y^N(t) \left( 2\|R^N(t)\|_\infty + \frac{C_\Omega^\alpha}{N r^\alpha} (1 + e^{tL_y^N(t)}(\text{Hol}_\alpha(y^0) + 1)) \right)$$

and, noting that  $\tau \mapsto L_y^N(\tau)$  (defined by (38)) is nondecreasing and by integration, we obtain (36).

Let us establish (37). For every  $x \in \Omega$  there exists  $i \in \{1, \dots, N\}$  such that  $x \in \Omega_i^N$ , and thus  $d_\Omega(x, x_i^N) \leq \text{diam}_\Omega(\Omega_i^N) \leq \frac{C_\Omega}{N^r}$  (by (27)). It follows from (35) that

$$\|y(t, x) - y(t, x_i^N)\| \leq \text{Hol}_\alpha(y(t, \cdot))d_\Omega(x, x_i^N)^\alpha \leq \frac{C_\Omega^\alpha}{N r^\alpha} e^{tL_y^N(t)} (\text{Hol}_\alpha(y(0, \cdot)) + 1)$$

and, noting that  $y^N(t, x) = \xi_i^N(t)$ , (37) follows by the triangular inequality, using (36).

Let us now prove the first item of Theorem 2.2. Starting as in the proof of Lemma B.1, by continuity of  $G$ , we infer from (127) that, for any  $\varepsilon > 0$ , if  $x$  and  $x'$  are sufficiently close then

$$\|\partial_t(y(t, x) - y(t, x'))\| \leq L_y(t) (\varepsilon + \|y(t, x) - y(t, x')\|)$$

and by integration we obtain

$$\|y(t, x) - y(t, x')\| \leq e^{tL_y(t)} (\|y^0(x) - y^0(x')\| + \varepsilon). \quad (133)$$

By assumption,  $y^0$  is continuous  $\nu$ -almost everywhere on  $\Omega$ . It follows from (133) that, for every  $t \geq 0$ ,  $y(t, \cdot)$  is continuous  $\nu$ -almost everywhere on  $\Omega$  with the same continuity set as  $y^0$  (thus, not depending on  $t$ ).

Let us finally establish (34). By the Riemann integration theorem (see (28)), we have  $\varepsilon_i^N(t) = o(1)$  (where  $\varepsilon_i^N(t)$  is defined by (129)) as  $N \rightarrow +\infty$ , uniformly with respect to  $t$  on every compact. Besides, we still have the inequality (130), but with  $L_y^N(t)$  now defined by

$$L_y^N(t) = \max_{\substack{x, x' \in \Omega \\ 0 \leq \tau \leq t}} \text{Lip}(G(\tau, x, x', \cdot, \cdot)|_{S_y^N(\tau)^2},$$

i.e., like in (38) but without the first term involving  $\text{Hol}_\alpha(G)$ . With this substitution, we obtain

$$\frac{d}{dt} \|R^N(t)\|_\infty \leq \|\dot{R}^N(t)\|_\infty \leq L_y^N(t) (2\|R^N(t)\|_\infty + o(1))$$

and integrating we get  $\|R^N(t)\|_\infty \leq e^{2tL_y^N(t)} o(1)$ , which yields (34), noting that  $L_y^N(t)$  is uniformly bounded with respect to  $t \in [0, T]$  and to  $N \in \mathbb{N}^*$  (as a consequence of Lemma 1.1). Then, (33) follows by the triangular inequality, using the  $\nu$ -almost everywhere continuity of  $y(t, \cdot)$ .

## B.2 Proof of Theorem 2.3

The proof is a slight adaptation of the proof of Theorem 2.2. We start by establishing (40). Hence, we assume that  $G$  is locally  $\alpha$ -Hölder continuous with respect to  $(x, x', \xi, \xi')$  (uniformly with respect to  $t$  on any compact).

**Lemma B.2.** *Let  $i \in \{1, \dots, N\}$  and  $x, x' \in \Omega_i^N$  be arbitrary. We have*

$$\|y_N(t, x) - y_N(t, x')\| \leq e^{tL_{y_N}(t)} d_\Omega(x, x')^\alpha \quad (134)$$

where  $y^0(\cdot) = y(0, \cdot)$ , where  $L_{y_N}(t)$  is defined as  $L_y(t)$  in Theorem 2.2 with  $y$  replaced by  $y_N$ .

*Proof.* Following the proof of Lemma B.1, we arrive at

$$\|\partial_t(y_N(t, x) - y_N(t, x'))\| \leq L_{y_N}(t) (d_\Omega(x, x')^\alpha + \|y_N(t, x) - y_N(t, x')\|)$$

and (134) follows by integration, noting that  $y_N(0, x) - y_N(0, x') = 0$  if  $x, x' \in \Omega_i^N$ .  $\square$

It follows from Lemma B.1 that  $y_N(t, \cdot)$  is  $\alpha$ -Hölder continuous in each  $\Omega_i$ , with Hölder constant  $e^{tL_{y_N}(t)}$ .

We set  $r^N(t, x) = y_N(t, x) - y^N(t, x)$  for every  $x \in \Omega$ . By definition, if  $x \in \Omega_i^N$  then  $y^N(t, x) = \xi_i^N(t)$  and thus

$$\partial_t r^N(t, x) = \frac{1}{N} \sum_{j=1}^N (G(t, x, x_j^N, y_N(t, x), y_N(t, x_j^N)) - G(t, x_i^N, x_j^N, \xi_i^N(t), \xi_j^N(t))) + \epsilon^N(t, x)$$

where

$$\epsilon^N(t, x) = \int_\Omega G(t, x, x', y_N(t, x), y_N(t, x')) d\nu(x') - \frac{1}{N} \sum_{j=1}^N G(t, x, x_j^N, y_N(t, x), y_N(t, x_j^N))$$

with  $r^N(0, x) = 0$ . We have on the one part

$$\begin{aligned} & \|G(t, x, x_j^N, y_N(t, x), y_N(t, x_j^N)) - G(t, x_i^N, x_j^N, \xi_i^N(t), \xi_j^N(t))\| \\ & \leq L_{y_N}^N(t)(d_\Omega(x, x_i^N)^\alpha + \|r^N(t, x)\| + \|r^N(t, x_j^N)\|) \quad \forall x \in \Omega_i^N \end{aligned}$$

and on the other part, proceeding like in the proof of Theorem 2.2 (see Appendix B.1),

$$\|\epsilon^N(t, x)\| \leq \frac{C_\Omega^\alpha}{N^{r\alpha}} L_{y_N}^N(t)(1 + e^{tL_{y_N}^N(t)}) \quad \forall x \in \Omega_i^N.$$

Using that  $d_\Omega(x, x_i^N) \leq \text{diam}_\Omega(\Omega_i^N) \leq \frac{C_\Omega}{N^r}$  (see (27)), we finally obtain

$$\frac{d}{dt} \|r^N(t, \cdot)\|_{L^\infty(\Omega)} \leq \|\partial_t r^N(t, \cdot)\|_{L^\infty(\Omega)} \leq L_{y_N}^N(t) \left( 2\|r^N(t, \cdot)\|_{L^\infty(\Omega)} + \frac{C_\Omega^\alpha}{N^{r\alpha}} (2 + e^{tL_{y_N}^N(t)}) \right)$$

and by integration, noting that  $\tau \mapsto L_{y_N}^N(\tau)$  is nondecreasing, (40) follows.

Finally, (39) is established as in the proof of Theorem 2.2.

### B.3 Proof of Theorem 3.1

Let  $T > 0$  be arbitrary. Let  $F$  be either equal to  $\Omega \times \mathbb{R}^d$ , or a compact subset of  $\Omega \times \mathbb{R}^d$  that is the closure of an open set. Let  $\mathcal{C}_0(F)$  be the Banach space of continuous functions on  $F$  vanishing at infinity (when  $F$  is compact we have  $\mathcal{C}_0(F) = \mathcal{C}^0(F)$ ), and let  $\mathcal{M}^1(F) = \mathcal{C}^0(F)'$  be the Banach space of Radon measures on  $F$ , endowed with the total variation norm  $\|\cdot\|_{TV}$  (which is the dual norm). We have  $\mathcal{P}_c(F) \subset \mathcal{M}^1(F)$  and  $\mathcal{C}^0([0, T], \mathcal{P}_c(F)) \subset L^\infty([0, T], \mathcal{M}^1(F))$ .

The Banach space  $L^\infty([0, T], \mathcal{M}^1(F))$  is endowed with its strong topology, induced by the  $L^\infty$  norm in time of the total variation in space, but can also be endowed with its weak star topology, as follows. Recall the general fact of 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, where the prime is the topological dual. Applying this fact to the Banach space  $E = \mathcal{C}^0(F)$ , observing that  $E' = \mathcal{M}^1(F)$  is separable in weak star topology (because the set of rational convex combinations of Dirac measures over points with rational coordinates is dense in it), the Banach space  $L^\infty([0, T], \mathcal{M}^1(F))$  coincides with  $L^1([0, T], \mathcal{C}^0(F))'$ , i.e., with the dual of a Banach space, and thus is endowed with the corresponding weak star topology.

We have the following preliminary lemma.

**Lemma B.3.** *Let  $K$  be a compact subset of  $\Omega \times \mathbb{R}^d$ , let  $\mu_0 \in \mathcal{P}_c(K)$  and let  $T > 0$  be arbitrary. Assume that there exists a sequence of measures  $\mu^k \in \mathcal{C}^0([0, T], \mathcal{P}_c(K))$  solutions of the Vlasov equation  $\partial_t \mu^k + L_{\mathcal{X}[\mu^k]} \mu^k = 0$  in the sense (47), such that:*

- $\mu^k(0)$  converges weakly to  $\mu_0$  in  $\mathcal{P}_c(K)$ ,
- $\mu^k$  converges to  $\mu \in L^\infty([0, T], \mathcal{M}^1(\Omega \times \mathbb{R}^d))$  for the weak star topology,

as  $k \rightarrow +\infty$ . Then  $\mu \in \mathcal{C}^0([0, T], \mathcal{P}_c(K))$  and  $t \mapsto \mu(t)$  is Lipschitz continuous in  $W_p$  distance (for any  $p \in [1, +\infty)$ ) and is a solution of the Vlasov equation  $\partial_t \mu + L_{\mathcal{X}[\mu]} \mu = 0$  (in the sense (47)) such that  $\mu(0) = \mu_0$ . Moreover,  $\mu^k(t)$  converges weakly to  $\mu(t)$  as  $k \rightarrow +\infty$  (equivalently,  $W_p(\mu^k(t), \mu(t)) \rightarrow 0$ ), uniformly with respect to  $t \in [0, T]$ .

*Proof.* For every  $k \in \mathbb{N}^*$ , since  $\mu^k \in \mathcal{C}^0([0, T], \mathcal{P}_c(K))$  is solution of the Vlasov equation in the sense (47) and thus in the distributional sense, we have  $\langle \partial_t \mu^k + L_{\mathcal{X}[\mu^k]} \mu^k, f \rangle = 0$  for any



$f \in \mathcal{C}^\infty([0, T] \times K)$  and thus, integrating by parts,

$$\begin{aligned} \int_K f(0, x, \xi) d\mu_0^k(x, \xi) + \int_0^T \int_K \partial_t f(t, x, \xi) d\mu^k(t, x, \xi) \\ + \int_0^T \int_K \langle \nabla_\xi f(t, x, \xi), G(\tau, x, x', \xi, \xi') \rangle d\mu^k(t, x', \xi') d\mu^k(t, x, \xi) d\tau = 0. \end{aligned}$$

Passing to the limit, we obtain the same equation with  $\mu^k$  replaced by  $\mu$  and  $\mu_0^k$  replaced by  $\mu_0$ , for any  $f \in \mathcal{C}^\infty([0, T] \times K)$ . Hence  $\mu$  is solution of the Vlasov equation  $\partial_t \mu + L_{\mathcal{X}[\mu]} \mu = 0$  in the distributional sense, with  $\mu(0) = \mu_0$ .

Let us prove that  $\mu \in \mathcal{C}^0([0, T], \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  and that  $\mu$  is a solution of the Vlasov equation in the sense (47) and that  $\mu^k(t)$  converges weakly to  $\mu(t)$  for every  $t \in [0, T]$ .

By assumption,  $\text{supp}(\mu^k) \subset [0, T] \times K$  for every  $k \in \mathbb{N}^*$ , hence, by the Prokhorov theorem, a subsequence of  $\mu^k$  converges weakly (i.e., in  $(\mathcal{C}^0([0, T] \times K))'$ ) to some measure, which must then be equal to  $\mu$ . Then  $\text{supp}(\mu)$  is contained in the Kuratowski liminf of  $\text{supp}(\mu^k)$  (see, e.g., [3, Proposition 5.1.8]), i.e., for every  $(t, x, \xi) \in \text{supp}(\mu)$  there exists a sequence of  $(t_k, x_k, \xi_k) \in \text{supp}(\mu^k)$  such that  $(t_k, x_k, \xi_k) \rightarrow (t, x, \xi)$  as  $k \rightarrow +\infty$ . We infer that  $\text{supp}(\mu) \subset [0, T] \times K$ . Since  $\mu \in L^\infty([0, T], \mathcal{M}^1(\Omega \times \mathbb{R}^d))$ , we must have  $\text{supp}(\mu(t)) \subset K$  for almost every  $t \in [0, T]$ .

For every  $k \in \mathbb{N}^*$ , we consider on  $[0, T] \times K$  the time-dependent vector field  $v^k(t, x, \xi) = \mathcal{X}[\mu_t^k](t, x, \xi)$ , which is continuous and Lipschitz with respect to  $\xi$  (thanks to (G)). Since  $\text{supp}(\mu^k) \subset [0, T] \times K$ , we have  $\|v^k\|_{\mathcal{C}^0([0, T] \times K)} \leq C$  for some  $C > 0$  not depending on  $k$ . Let  $(\Phi_{v^k}(t))_{t \in [0, T]}$  be the flow on  $[0, T] \times K$  generated by  $v^k$ . Since  $\mu^k$  is a solution of the transport equation  $\partial_t \mu^k + L_{v^k} \mu^k = 0$ , by the usual existence and uniqueness theorem for linear transport equations (see, e.g., [77, Theorem 5.34]), we have  $\mu^k(t) = \Phi_{v^k}(t)_* \mu_0^k$  for every  $t \in [0, T]$ . This means, denoting by  $\nu^k = \pi_* \mu^k(t)$  the (constant in time) marginal of  $\mu^k(t)$  on  $\Omega$  and disintegrating  $\mu_t^k = \mu^k(t) = \int_\Omega \mu_{t,x}^k d\nu^k(x)$ , that  $\mu_{t,x}^k = \Phi_{v^k}(t, x, \cdot)_* \mu_{0,x}^k$  for  $\nu^k$ -almost every  $x \in \pi(K)$ , for every  $t \in [0, T]$ . It follows from Lemma A.9 of Appendix A.1.6 that  $W_p(\mu^k(t_1), \mu^k(t_2)) \leq C|t_1 - t_2|$  for all  $t_1, t_2 \in [0, T]$ , i.e.,  $t \mapsto \mu^k(t)$  is Lipschitz continuous on  $[0, T]$  in  $W_p$  distance, uniformly with respect to  $k$ . Since  $\mu^k$  converges to  $\mu \in L^\infty([0, T], \mathcal{M}^1(K))$  for the weak star topology, it follows from the Ascoli theorem that  $\mu^k(t)$  converges weakly to  $\mu(t)$  in  $\mathcal{M}^1(K) = \mathcal{C}^0(K)'$ , uniformly with respect to  $t \in [0, T]$ , and thus that  $\mu \in \mathcal{C}^0([0, T], \mathcal{P}_c^1(K))$  and that  $t \mapsto \mu(t)$  is Lipschitz continuous in  $W_p$  distance. Lemma B.3 is proved.  $\square$

In view of establishing Item (A), let us first prove the existence of a solution of the Vlasov equation. Given  $\mu_0 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$ , we consider a sequence of empirical measures  $\mu_{(X^N, \Xi_0^N)}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_{0,i}^N}$  converging weakly to  $\mu_0$  as  $N \rightarrow +\infty$ . Setting  $X^N = (x_1^N, \dots, x_N^N) \in \Omega^N$  and  $\Xi_0^N = (\xi_1^N, \dots, \xi_N^N) \in (\mathbb{R}^d)^N$ , let  $t \mapsto \Xi^N(t) = (\xi_1^N(t), \dots, \xi_N^N(t))$  be the unique solution of the particle system (9) with parameter  $X^N$  such that  $\Xi^N(0) = \Xi_0^N$ . It is well defined on  $[0, T]$  for any  $T \in (0, T_{\max}(\text{supp}(\mu_0)))$  thanks to Assumption (G) and Lemma 1.1. Using the first part of Proposition 3.1, which does not use anything from Theorem 3.1 (see its proof),  $t \mapsto \mu_{(X^N, \Xi^N(t))}^e = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N(t)}$  is a solution of the Vlasov equation (45) in the sense (47).

Without loss of generality, we can assume that  $(X^N, \Xi_0^N) \in (\text{supp}(\mu_0))^N$  for every  $N \in \mathbb{N}^*$ , where we recall that  $\text{supp}(\mu_0)$  is compact. Since  $\mu_{(X^N, \Xi^N(t))}^e$  is supported on the corresponding solutions of the particle system, it follows from Lemma 1.1 that there exists a compact subset  $K \subset \Omega \times \mathbb{R}^d$  such that  $\text{supp}(\mu_{(X^N, \Xi^N(t))}^e) \subset K$  for every  $t \in [0, T]$  and for every  $N \in \mathbb{N}^*$ , i.e., the measures  $\mu_{(X^N, \Xi^N(\cdot))}^e$  are equi-compactly supported on  $[0, T]$ , uniformly with respect to  $N$ .

Besides, since  $\mu_{(X^N, \Xi^N(t))}^e$  is a probability measure, we have  $\|\mu_{(X^N, \Xi^N(t))}^e\|_{TV} = 1 < +\infty$  for every  $t \in [0, T]$ , and thus the sequence  $(\mu_{(X^N, \Xi^N(\cdot))}^e)_{N \in \mathbb{N}^*}$  is bounded in  $L^\infty([0, T], \mathcal{M}^1(\Omega \times \mathbb{R}^d))$

for the strong topology, i.e., in  $(L^1([0, T], \mathcal{C}_0(\Omega \times \mathbb{R}^d)))'$  for the strong (dual norm) topology. By the Banach-Alaoglu theorem, there exists a subsequence of  $(\mu_{(X^N, \Xi^N(\cdot))}^e)_{N \in \mathbb{N}^*}$  converging to some  $\mu \in L^\infty([0, T], \mathcal{M}^1(\Omega \times \mathbb{R}^d))$  for the weak star topology.

Therefore, a subsequence of the sequence of measures  $\mu_{(X^N, \Xi^N(\cdot))}^e \in \mathcal{C}_{\text{comp}}^0([0, T], \mathcal{P}_c(K))$  satisfies all assumptions of Lemma B.3. It follows from that lemma that  $\mu \in \mathcal{C}^0([0, T], \mathcal{P}_c(K))$  and that  $\mu$  is a solution on  $[0, T]$  of the Vlasov equation  $\partial_t \mu + L_{\mathcal{X}[\mu]} \mu = 0$  (in the sense (47)) such that  $\mu(0) = \mu_0$ , and is Lipschitz continuous with respect to  $t$  in  $W_p$  distance.

At this step, we have obtained the existence of solutions in  $\mathcal{C}_{\text{comp}}^0([0, T_{\max}(\text{supp}(\mu_0))], \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  (not yet uniqueness).

**Remark B.2.** In [65, 66, 67, 68], the proof of the existence is done by constructing a sequence of piecewise constant measures converging to a solution, under the stronger assumption that  $G$  be globally Lipschitz continuous. The proof given above relies on approximation by empirical measures and propagation of them, in the spirit of [37] (see also [61] and [75, Part I, Theorem 5.1]), which is more appropriate to exploit Lemma 1.1. For the Cucker–Smale model, the proof done in [16] relies on a fixed point argument in the metric space of solutions, by adequately estimating the propagation of supports.

**Remark B.3.** Before going ahead, let us observe that, when  $G$  is locally Lipschitz with respect to all variables  $(x, x', \xi, \xi')$ , we have, for all  $\mu^1, \mu^2 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$  and for every  $(t, x, \xi) \in [0, T] \times \Omega \times \mathbb{R}^d$ ,

$$\begin{aligned} \|\mathcal{X}[\mu^1](t, x, \xi) - \mathcal{X}[\mu^2](t, x, \xi)\| &= \left\| \int_{\Omega \times \mathbb{R}^d} G(t, x, x', \xi, \xi') d(\mu^1(x', \xi') - \mu^2(x', \xi')) \right\| \\ &\leq \text{Lip}(G(t, x, \cdot, \xi, \cdot)|_S) W_1(\mu^1, \mu^2) \leq \text{Lip}(G(t, x, \cdot, \xi, \cdot)|_S) W_p(\mu^1, \mu^2) \end{aligned} \quad (135)$$

where  $S = \text{supp}(\mu^1) \cup \text{supp}(\mu^2)$  (compact set). We have used that  $W_1 \leq W_p$ .

In the case (A), there is however a weaker assumption: under Assumption (G),  $G$  is locally Lipschitz only with respect to  $(\xi, \xi')$  and thus the classical Wasserstein distance  $W_1$  cannot be used as above. The main difference then comes from the following observation: given any  $\mu^1, \mu^2 \in \mathcal{P}_c(\Omega \times \mathbb{R}^d)$  having the same marginal  $\nu \in \mathcal{P}_c(\Omega)$  on  $\Omega$ , we have, by disintegration,

$$\mathcal{X}[\mu^1](t, x, \xi) - \mathcal{X}[\mu^2](t, x, \xi) = \int_{\Omega} \int_{\mathbb{R}^d} G(t, x, x', \xi, \xi') d(\mu_{x'}^1(\xi') - \mu_{x'}^2(\xi')) d\nu(x')$$

and thus

$$\begin{aligned} \|\mathcal{X}[\mu^1](t, x, \xi) - \mathcal{X}[\mu^2](t, x, \xi)\| &\leq \max_{x' \in \text{supp}(\nu)} \text{Lip}(G(t, x, x', \xi, \cdot)|_{S_{x'}}) L_\nu^1 W_1(\mu^1, \mu^2) \\ &\leq \max_{x' \in \text{supp}(\nu)} \text{Lip}(G(t, x, x', \xi, \cdot)|_{S_{x'}}) L_\nu^1 W_p(\mu^1, \mu^2) \end{aligned} \quad (136)$$

where  $S_{x'} = \text{supp}(\mu_{x'}^1) \cup \text{supp}(\mu_{x'}^2)$  (compact) and  $L_\nu^1 W_p(\mu^1, \mu^2) = \int_{\Omega} W_p(\mu_{x'}^1, \mu_{x'}^2) d\nu(x')$  is defined by (26).

Let us now establish (51) in the item (A<sub>2</sub>) (which also entails uniqueness). Let  $\mu^1, \mu^2 \in \mathcal{C}_{\text{comp}}^0([0, T], \mathcal{P}_c(\Omega \times \mathbb{R}^d))$  be two solutions of the Vlasov equation for some  $T > 0$ , having the same (constant in time) marginal  $\nu = \pi_* \mu^i \in \mathcal{P}_c(\Omega)$  on  $\Omega$ . Let  $K \subset \Omega \times \mathbb{R}^d$  be a compact subset containing  $\text{supp}(\mu^i(t))$  for  $i = 1, 2$  and for every  $t \in [0, T]$ .

For  $i = 1, 2$ , we consider on  $[0, T] \times K$  the continuous time-dependent vector field  $v^i(t, x, \xi) = \mathcal{X}[\mu_t^i](t, x, \xi)$  (which is  $\mathcal{C}^1$  with respect to  $\xi$ ), so that  $\mu^i$  is a solution of the transport equation  $\partial_t \mu^i + L_{v^i} \mu^i = 0$ . Since we are going to apply Lemma A.8 (in Appendix A.1.6) with  $t_0 \neq 0$ , for every  $t_0 \in [0, T]$  we consider the flow  $(\Phi_{v^i}(t, t_0))_{t \in [0, T]}$  on  $[0, T] \times K$  generated by  $v^i$ , i.e., defined as

the unique solution of  $\partial_t \Phi_{v^i}(t, t_0, x, \cdot) = v^i(t, x, \cdot) \circ \Phi_{v^i}(t, t_0, x, \cdot)$  such that  $\Phi_{v^i}(t_0, t_0, x, \cdot) = \text{id}_{\mathbb{R}^d}$  for  $\nu$ -almost every  $x \in \text{supp}(\nu)$ . Then, we have  $\mu^i(t) = \Phi_{v^i}(t, t_0)_* \mu^i(t_0)$  for every  $t \in [0, T]$ . This means, disintegrating  $\mu_t^i = \mu^i(t) = \int_{\Omega} \mu_{t,x}^i d\nu(x)$ , that  $\mu_{t,x}^i = \Phi_{v^i}(t, t_0, x, \cdot)_* \mu_{t_0,x}^i$  for  $\nu$ -almost every  $x \in \text{supp}(\nu)$ , for every  $t \in [0, T]$ .

It follows from Lemma A.8 (in Appendix A.1.6), applied with  $\Lambda = \emptyset$  and  $E = \mathbb{R}^d$  to the vector fields  $v^i(t, x, \cdot)$  for any fixed  $x \in \text{supp}(\nu)$ , that

$$W_p(\mu_{t,x}^1, \mu_{t,x}^2) \leq e^{(t-t_0)L([t_0,t])} W_p(\mu_{t_0,x}^1, \mu_{t_0,x}^2) + M([t_0,t]) \frac{e^{(t-t_0)L([t_0,t])} - 1}{L([t_0,t])} \quad \forall t \in [t_0, T]$$

where, setting  $S(t) = \text{supp}(\nu) \times \Phi_{v^1}(t, t_0, \text{supp}(\mu_{t_0}^1) \cup \text{supp}(\mu_{t_0}^2)) \cup \text{supp}(\mu^2(t))$  (compact),

$$L([t_0, t]) = \text{ess sup} \{ \|(\partial_\xi G, \partial_{\xi'} G)(\tau, x, x', \xi, \xi')\| \mid t_0 \leq \tau \leq t, (x, \xi), (x', \xi') \in S(\tau) \},$$

$$M([t_0, t]) = \max \{ \|\mathcal{X}[\mu_\tau^1](\tau, x, \xi) - \mathcal{X}[\mu_\tau^2](\tau, x, \xi)\| \mid t_0 \leq \tau \leq t, (x, \xi) \in \text{supp}(\mu_\tau^2) \}.$$

Since  $\mu_\tau^1$  and  $\mu_\tau^2$  have the same marginal  $\nu$  on  $\Omega$ , it follows from (136) and from the above definition of  $L([t_0, t])$  and of  $S(t)$  that

$$M([t_0, t]) \leq L([t_0, t]) \max_{t_0 \leq \tau \leq t} L_\nu^1 W_p(\mu_\tau^1, \mu_\tau^2).$$

Therefore

$$W_p(\mu_{t,x}^1, \mu_{t,x}^2) \leq e^{(t-t_0)L([t_0,t])} W_p(\mu_{t_0,x}^1, \mu_{t_0,x}^2) + (e^{(t-t_0)L([t_0,t])} - 1) \max_{t_0 \leq \tau \leq t} L_\nu^1 W_p(\mu_\tau^1, \mu_\tau^2).$$

Integrating with respect to  $x \in \Omega$  for the measure  $\nu$ , we obtain

$$\begin{aligned} L_\nu^1 W_p(\mu^1(t), \mu^2(t)) &\leq e^{(t-t_0)L([t_0,t])} L_\nu^1 W_p(\mu^1(t_0), \mu^2(t_0)) \\ &\quad + (e^{(t-t_0)L([t_0,t])} - 1) \max_{t_0 \leq \tau \leq t} L_\nu^1 W_p(\mu^1(\tau), \mu^2(\tau)). \end{aligned} \quad (137)$$

We have the following general lemma.

**Lemma B.4.** *For every  $t_0 \in \mathbb{R}$ , let  $a_{t_0} : [t_0, +\infty) \rightarrow [0, +\infty)$  be a nondecreasing function, continuous at  $t_0$ , depending continuously on  $t_0$ . Let  $h : \mathbb{R} \rightarrow [0, +\infty)$  be an absolutely continuous function such that*

$$h(t) \leq e^{(t-t_0)a_{t_0}(t)} h(t_0) + (e^{(t-t_0)a_{t_0}(t)} - 1) \max_{t_0 \leq \tau \leq t} h(\tau) \quad \forall t \geq t_0 \quad \forall t_0 \in \mathbb{R}.$$

Then

$$h(t) \leq h(0) \exp \left( 2 \int_0^t a_\tau(\tau) d\tau \right) \quad \forall t \in \mathbb{R}.$$

*Proof.* Taking  $t_0 < t < t_1$ , writing

$$\frac{h(t) - h(t_0)}{t - t_0} \leq \frac{e^{(t-t_0)a_{t_0}(t_1)} - 1}{t - t_0} h(t_0) + \frac{e^{(t-t_0)a_{t_0}(t_1)} - 1}{t - t_0} \max_{t_0 \leq \tau \leq t} h(\tau)$$

and taking the limit as  $t \rightarrow t_0$ , since  $t_1$  is arbitrary, we obtain  $h'(t_0) \leq 2h(t_0)a_{t_0}(t_0)$ , for almost every  $t_0 \in \mathbb{R}$ . The lemma follows by integration.  $\square$

Applying Lemma B.4 to  $h(t) = L_\nu^1 W_p(\mu^1(t), \mu^2(t))$  and  $a_{t_0}(t) = L([t_0, t])$ , and using (137), we obtain (51). In particular, the uniqueness statement follows.

At this step, we have proved existence and uniqueness of solutions of the Vlasov equation in the space  $\mathcal{C}_{\text{comp}}^0([0, T], \mathcal{P}_c(\Omega \times \mathbb{R}^d))$ . We can thus now define the Vlasov flow by (49), and we obtain (48) by uniqueness.

Establishing (53) in the item (B) follows straightforwardly the same lines as above, by applying Lemma A.8 (in Appendix A.1.6) with  $\Lambda = \Omega$  and  $E = \mathbb{R}^d$ , and using (135) instead of (136). We do not give any further details.

It remains to establish the item (A<sub>1</sub>). For  $K \subset \Omega \times \mathbb{R}^d$  compact and  $T \in (0, T_{\max}(K))$ , we consider a sequence of measures  $\mu^k \in \mathcal{C}^0([0, T], \mathcal{P}_c(K))$  solutions of the Vlasov equation such that  $\mu_0^k = \mu^k(0)$  converges weakly to  $\mu_0 = \mu(0)$  as  $k \rightarrow +\infty$ . Our objective is to prove that  $\mu^k(t)$  converges weakly to  $\mu(t)$ , uniformly with respect to  $t \in [0, T]$ .

Since  $\mu^k(t)$  is a probability measure, we have  $\|\mu^k(t)\|_{TV} = 1 < +\infty$  for every  $t \in [0, T]$ , 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), i.e., in  $(L^1([0, T], \mathcal{C}_0(\Omega \times \mathbb{R}^d)))'$  for the strong (dual norm) topology. By the Banach-Alaoglu theorem, a subsequence of  $(\mu^k(\cdot))_{k \in \mathbb{N}^*}$  converges to some  $\tilde{\mu}(\cdot) \in L^\infty([0, T], \mathcal{M}^1(\Omega \times \mathbb{R}^d))$  for the weak star topology.

It follows from Lemma B.3 that  $\tilde{\mu} \in \mathcal{C}^0([0, T], \mathcal{P}_c(K))$ , that  $\tilde{\mu}$  is a solution of the Vlasov equation such that  $\tilde{\mu}(0) = \mu_0$ , and that  $\mu^k(t)$  converges weakly to  $\tilde{\mu}$  uniformly with respect to  $t$ . By uniqueness, we must have  $\tilde{\mu} = \mu$ . This finishes the proof of the theorem.

## B.4 Proof of Theorem 4.1

We have  $\rho^N(t) = \delta_{X^N} \otimes \delta_{\Xi^N(t)}$ . By (104) in Lemma A.14 of Appendix A.2.3 (applied with  $\gamma_i = \delta_{\xi_i^N(t)}$ ), we have

$$\rho^N(t)_{N:1}^s = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \delta_{\xi_i^N(t)} = \mu_{(X^N, \Xi^N(t))}^e,$$

which gives the preliminary remark to Theorem 4.1. The statement (A) for  $k = 1$  then follows from the item (A<sub>1</sub>) of Theorem 3.1, and the estimate (66) follows from the item (B) of Theorem 3.1.

For any  $k \in \mathbb{N}^*$ , the  $k^{\text{th}}$ -order marginal  $\rho^N(t)_{N:k}^s$  is given by (105) in Appendix A.2.3 (applied with  $\mu_i = \delta_{x_i^N} \otimes \delta_{\xi_i^N(t)}$ ). By the triangular inequality, we have

$$W_p^{[q]}(\rho^N(t)_{N:k}^s, \mu(t)^{\otimes k}) \leq W_p^{[q]}(\rho^N(t)_{N:k}^s, (\mu_{(X^N, \Xi^N(t))}^e)^{\otimes k}) + W_p^{[q]}((\mu_{(X^N, \Xi^N(t))}^e)^{\otimes k}, \mu(t)^{\otimes k}) \quad (138)$$

For the first term at the right-hand side of (138), noting that  $(X^N, \Xi_0^N) \in (\text{supp}(\mu_0))^N$  and thus

$$\text{diam}_{\Omega \times \mathbb{R}^d} \left( \bigcup_{i=1}^N \text{supp}(\delta_{x_i^N} \otimes \delta_{\xi_i^N(t)}) \right) \leq \text{diam}_\Omega(\text{supp}(\nu)) + \text{diam}_{\mathbb{R}^d}(\Xi^N(t)),$$

and noting that  $\mu_{(X^N, \Xi^N(t))}^e = \rho^N(t)_{N:1}^s$ , we infer from (109) in Lemma A.14 of Appendix A.2.3 that

$$W_p^{[q]}(\rho^N(t)_{N:k}^s, (\mu_{(X^N, \Xi^N(t))}^e)^{\otimes k}) \leq 2k^{1/q} \left( \frac{k^2}{N} \right)^{1/p} (\text{diam}_\Omega(\text{supp}(\nu)) + \text{diam}_{\mathbb{R}^d}(\Xi^N(t))). \quad (139)$$

For the second term at the right-hand side of (138), as a consequence, successively, of the estimate (93) of Lemma A.6 in Appendix A.1.4 and of Theorem 3.1, we have

$$W_p^{[q]} \left( (\mu_{(X^N, \Xi^N(t))}^e)^{\otimes k}, \mu(t)^{\otimes k} \right) \leq k^{1/q} W_p \left( \mu_{(X^N, \Xi^N(t))}^e, \mu(t) \right) \leq k^{1/q} C_\mu^N(t) W_p \left( \mu_{(X^N, \Xi_0^N)}^e, \mu(0) \right) \quad (140)$$

where the constant  $C_\mu^N(t)$  is defined by (65) (or equivalently by  $C_{\mu, \mu_{(X^N, \Xi^N)}^e}^N(t)$ , with the notation used in (50) in Theorem 3.1). Therefore, (67) follows from (138), (139) and (140). Note that, for  $k = 1$ , the first term of the right-hand side of (138) is equal to 0, which gives (66) again.

The statement (A) for any  $k \in \mathbb{N}^*$  is proved by replacing the right inequality in (140) with the application of the item (A) of Theorem 3.1.

## B.5 Proof of Theorem 4.2

First of all, since  $\rho_0^N = \delta_{X^N} \otimes \rho_{0, X^N}^N$ , with  $\delta_{X^N} = \delta_{x_1^N} \otimes \cdots \otimes \delta_{x_N^N}$  and  $\rho_{0, X^N}^N = \mu_{0, x_1^N} \otimes \cdots \otimes \mu_{0, x_N^N}$ , it follows from (104) in Lemma A.14 of Appendix A.2.3 (applied with  $\mu_i = \delta_{x_i^N} \otimes \mu_{0, x_i^N}$ ) that  $(\rho_0^N)_{N:1}^s = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \mu_{0, x_i^N} = (\mu_0)_{X^N}^{se}$  (semi-empirical measure), which gives (71), and the weak convergence to  $\mu_0$  stated in Item (A) for  $k = 1$  is obtained by Lemma A.19 of Appendix A.4.2, which gives the preliminary remark to the theorem.

Recall that  $\rho^N(t) = \Phi^N(t)_* \rho_0^N$  and  $\mu(t) = \varphi_{\mu_0}(t)_* \mu_0$ . Setting

$$\tilde{\rho}^N(t) = \varphi_{\mu_0}(t)_*^{\otimes N} \rho_0^N = \delta_{x_1^N} \otimes \cdots \otimes \delta_{x_N^N} \otimes \mu_{t, x_1^N} \otimes \cdots \otimes \mu_{t, x_N^N}$$

(the latter equality is because  $\varphi_{\mu_0}(t, x_i, \cdot)_* \mu_{0, x_i^N} = \mu_{t, x_i^N}$ ), we note that  $\tilde{\rho}^N(t)^s = \varphi_{\mu_0}(t)_*^{\otimes N} (\rho_0^N)^s$  and that

$$\tilde{\rho}^N(t)_{N:k}^s = \varphi_{\mu_0}(t)_*^{\otimes k} (\rho_0^N)_{N:k}^s \quad \forall k \in \{1, \dots, N\}.$$

Indeed, this follows from the following obvious lemma.

**Lemma B.5.** *Let  $E$  be a measure space,  $\varphi : E \rightarrow E$  be a measurable mapping,  $N \in \mathbb{N}^*$  and  $\rho \in \mathcal{P}(E^N)$ . Then*

$$(\varphi_*^{\otimes N} \rho)_{N:k} = \varphi_*(\rho_{N:k}) \quad \forall k \in \{1, \dots, N\}.$$

*Proof of Lemma B.5.* Denoting by  $\pi_k : E^N = E^k \times E^{N-k} \rightarrow E^k$  the canonical projection, the lemma straightforwardly follows from the fact that  $\pi_k \circ \varphi^{\otimes N} = \varphi^{\otimes k} \circ \pi_k$ .  $\square$

In particular, we have

$$\tilde{\rho}^N(t)_{N:1}^s = \varphi_{\mu_0}(t)_* (\rho_0^N)_{N:1}^s = \varphi_{\mu_0}(t)_* (\mu_0)_{X^N}^{se} = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^N} \otimes \mu_{t, x_i^N} = \mu(t)_{X^N}^{se}.$$

In order to establish (73), we start by applying the triangular inequality:

$$W_p^{[q]} \left( \rho^N(t)_{N:k}^s, \mu(t)^{\otimes k} \right) \leq W_p^{[q]} \left( \rho^N(t)_{N:k}^s, \tilde{\rho}^N(t)_{N:k}^s \right) + W_p^{[q]} \left( \tilde{\rho}^N(t)_{N:k}^s, \mu(t)^{\otimes k} \right), \quad (141)$$

and we next show how to estimate each of the two terms of the sum at the right-hand side of (141).

**First term.** Applying successively Lemma A.13 in Appendix A.2.2 and Lemma A.4 in Appendix A.1.2, and using that  $W_p^{[q]} \leq W_2^{[q]} \leq W_2^{[1]}$  (see (23) and (88)) because  $p \leq 2$ , we have

$$W_p^{[q]} \left( \rho^N(t)_{N:k}^s, \tilde{\rho}^N(t)_{N:k}^s \right) \leq \left( \frac{k}{N} \right)^{1/q} W_2^{[1]} \left( \rho^N(t), \tilde{\rho}^N(t) \right) \quad (142)$$

where we insist that the latter Wasserstein distance  $W_2$  is computed with respect to the distance  $d_{(\Omega \times \mathbb{R}^d)^N}^{[1]}$  defined by (63): the choice of  $q = 1$  is important. Given that  $\rho^N(t) = \Phi^N(t)_* \rho_0^N$  and  $\tilde{\rho}^N(t) = \varphi_{\mu_0}(t)^{\otimes N} \rho_0^N$ , to estimate this distance, we apply Lemma A.8 (in Appendix A.1.6) with  $\Lambda = \Omega^N$  and  $E = \mathbb{R}^{dN}$  to the flows  $\Phi^N(t)$  and  $\varphi_{\mu_0}(t)^{\otimes N}$  in the space  $\Omega^N \times \mathbb{R}^{dN}$  endowed with the distance  $d_{(\Omega \times \mathbb{R}^d)^N}^{[1]}$ , respectively generated by the time-dependent vector fields  $Y^N(t, \cdot, \cdot)$  (defined by (10)) and  $\mathcal{X}[\mu_t](t, \cdot, \cdot)^{\otimes N}$  (with  $\mathcal{X}[\mu_t]$  defined by (44)), obtaining from the alternative estimate of that lemma, with  $p = 2$ , that

$$W_2^{[1]}(\rho^N(t), \tilde{\rho}^N(t)) \leq M_2(t) \sqrt{t} \left( \frac{e^{tL_2(t)} - 1}{L_2(t)} \right)^{1/2} \quad (143)$$

where

$$L_2(t) = \max_{0 \leq \tau \leq t} \text{Lip}(Y^N(\tau, \cdot, \cdot)|_{\text{supp}(\rho^N(\tau)) \cup \text{supp}(\tilde{\rho}^N(\tau))})$$

and, using (98),

$$\begin{aligned} M_2(t) &= \max_{0 \leq \tau \leq t} \left( \int_{\Omega^N \times \mathbb{R}^{dN}} \|Y^N(\tau, \cdot, \cdot) - \mathcal{X}[\mu_\tau](\tau, \cdot, \cdot)^{\otimes N}\|_{\ell^1}^2 d\tilde{\rho}_\tau^N \right)^{1/2} \\ &= \max_{0 \leq \tau \leq t} \left( \int_{\Omega^N \times \mathbb{R}^{dN}} \left( \sum_{i=1}^N \|Y_i^N(\tau, X, \Xi) - \mathcal{X}[\mu_\tau](\tau, x_i, \xi_i)\| \right)^2 d\tilde{\rho}_\tau^N(X, \Xi) \right)^{1/2} \end{aligned}$$

where we recall that  $\|\Xi\|_{\ell^1} = \sum_{i=1}^N \|\xi_i\|$  for any  $\Xi = (\xi_1, \dots, \xi_N) \in (\mathbb{R}^d)^N$ . Let us estimate  $L_2(t)$  and  $M_2(t)$ .

Since the  $\ell^1$  distance  $d_{(\Omega \times \mathbb{R}^d)^N}^{[1]}$  has been used, according to Lemma A.1 in Section A.1 we have, using the definition (11) of  $Y_i$ ,

$$\begin{aligned} L_2(t) &= \max_{0 \leq \tau \leq t} \max_{1 \leq i \leq N} \text{Lip}(Y_i^N(\tau, \cdot, \cdot)|_{\text{supp}(\rho^N(t)) \cup \text{supp}(\tilde{\rho}^N(t))}) = \max_{0 \leq \tau \leq t} \text{Lip}(G(\tau, \cdot, \cdot, \cdot, \cdot)|_{S_\mu^N(\tau)^2}) \\ &\leq \max_{0 \leq \tau \leq t} \|G(\tau, \cdot, \cdot, \cdot, \cdot)|_{S_\mu^N(\tau)^2}\|_{\mathcal{C}^{0,1}} = L(t) \end{aligned}$$

where  $S_\mu^N(\tau)$  is defined by (64). The choice  $q = 1$  has been crucial here (for a choice  $q > 1$  we would get a positive power of  $N$  in the exponential term in (143), which is not desirable).

Besides, by Lemma A.22 in Appendix A.6 (the choice of  $p = 2$  has been done to be able to apply this lemma), we have

$$M_2(t) \leq 2L(t) \left( \sqrt{N} \sqrt{1 + 70 \max_{0 \leq \tau \leq t} \text{diam}_{\Omega \times \mathbb{R}^d}(\text{supp}(\mu(t)))} + N \sqrt{5 W_1(\mu(t)_{X^N}^{se}, \mu(t))} \right).$$

Since the map  $s \mapsto \frac{e^{ts} - 1}{s}$  is increasing for  $s > 0$ , and since  $\sqrt{y(e^y - 1)} \leq e^y$  for every  $y \geq 0$ , we infer from (142) and (143) that

$$W_2^{[q]}(\rho^N(t)_{N:k}^s, \tilde{\rho}^N(t)_{N:k}^s) \leq 2 \left( \frac{k}{N} \right)^{1/q} \left( \sqrt{N} C'_\mu(t) + N \sqrt{5 W_1(\mu(t)_{X^N}^{se}, \mu(t))} \right) e^{tL(t)}.$$

where  $C'_\mu(t) = (1 + 70 \max_{0 \leq \tau \leq t} \text{diam}_{\Omega \times \mathbb{R}^d}(\text{supp}(\mu(t))))^{1/2}$ , for every  $t \geq 0$ . Applying Lemma A.8 (in Appendix A.1.6) with  $\Lambda = \Omega$  and  $E = \mathbb{R}^d$  to the Vlasov flow  $\varphi_{\mu_0}(t)$  in  $\Omega \times \mathbb{R}^d$  generated by the vector field  $\mathcal{X}[\mu_t](t, \cdot, \cdot)$ , we obtain

$$W_1(\mu(t)_{X^N}^{se}, \mu(t)) \leq e^{tL(t)} W_1((\mu_0)_{X^N}^{se}, \mu_0).$$

Finally,

$$W_p^{[q]}(\rho^N(t)_{N:k}^s, \tilde{\rho}^N(t)_{N:k}^s) \leq 2k^{1/q} \left( \frac{C'_\mu(t)}{N^{\frac{1}{q}-\frac{1}{2}}} + N^{1-\frac{1}{q}} \sqrt{5 W_1((\mu_0)_{X^N}^{se}, \mu_0)} \right) e^{2tL(t)} \quad (144)$$

**Second term.** Applying Lemma A.8 (in Appendix A.1.6) with  $\Lambda = \Omega^k$  and  $E = (\mathbb{R}^d)^k$  to the Vlasov flow  $\varphi_{\mu_0}(t)^{\otimes k}$  in the space  $\Omega^k \times (\mathbb{R}^d)^k$  endowed with the distance  $d_{(\Omega \times \mathbb{R}^d)^k}^{[q]}$  defined by (63), generated by the vector field  $\mathcal{X}[\mu_t](t, \cdot, \cdot)^{\otimes k}$ , we obtain

$$W_p^{[q]}(\tilde{\rho}^N(t)_{N:k}^s, \mu(t)^{\otimes k}) = W_p^{[q]}(\varphi_{\mu_0}(t)^{\otimes k}(\rho_0^N)_{N:k}^s, \varphi_{\mu_0}(t)^{\otimes k} \mu_0^{\otimes k}) \leq e^{tL(t)} W_p^{[q]}((\rho_0)_{N:k}^s, \mu_0^{\otimes k})$$

where  $L(t)$  is defined as before.

Similarly as in the proof of Theorem 4.1 (see Appendix B.4), we note that, for any  $k \in \{1, \dots, N\}$ , the measure  $(\rho_0^N)_{N:k}^s$  is given by the formula (105) of Lemma A.14 in Appendix A.2.3 with  $\beta_k$  given by (107) and  $\gamma_i = \mu_{0, x_i^N}$ . Hence we infer from (109) in Lemma A.14 (Appendix A.2.3) that, if  $k^2 \leq 2N \ln(1 + \frac{1}{2^p})$ , since  $(\rho_0^N)_{N:1}^s = (\mu_0)_{X^N}^{se}$ ,

$$W_p^{[q]}((\rho_0^N)_{N:k}^s, ((\mu_0)_{X^N}^{se})^{\otimes k}) \leq 2k^{1/q} \left( \frac{k^2}{N} \right)^{1/p} \text{diam}_{\Omega \times \mathbb{R}^d}(\text{supp}(\mu_0))$$

(the above term is zero and thus does not appear in the final estimate when  $k = 1$ ). Therefore, by the triangular inequality and by (93) in Lemma A.6 (Appendix A.1.4),

$$\begin{aligned} W_p^{[q]}(\tilde{\rho}^N(t)_{N:k}^s, \mu(t)^{\otimes k}) &\leq \left( W_p^{[q]}((\rho_0)_{N:k}^s, ((\mu_0)_{X^N}^{se})^{\otimes k}) + W_p^{[q]}(((\mu_0)_{X^N}^{se})^{\otimes k}, \mu_0^{\otimes k}) \right) e^{tL(t)} \\ &\leq k^{1/q} \left( 2 \left( \frac{k^2}{N} \right)^{1/p} \text{diam}_{\Omega \times \mathbb{R}^d}(\text{supp}(\mu_0)) + W_p((\mu_0)_{X^N}^{se}, \mu_0) \right) e^{tL(t)}. \end{aligned} \quad (145)$$

**Conclusion.** From (141), (144) and (145), we conclude that, for every  $t \geq 0$ ,

$$\begin{aligned} W_p^{[q]}(\rho^N(t)_{N:k}^s, \mu(t)^{\otimes k}) &\leq 2k^{1/q} \left( \left( \frac{k^2}{N} \right)^{1/p} C'_\mu(0) + \frac{C'_\mu(t)}{N^{\frac{1}{q}-\frac{1}{2}}} \right. \\ &\quad \left. + N^{1-\frac{1}{q}} \sqrt{5 W_1((\mu_0)_{X^N}^{se}, \mu_0)} + W_p((\mu_0)_{X^N}^{se}, \mu_0) \right) e^{2tL(t)} \end{aligned}$$

and (73) finally follows.

To obtain the statement (A) for any  $k \in \mathbb{N}^*$ , we have to adapt all the above arguments and in particular Lemma A.8 in Appendix A.1.6 to the case where  $G$  is locally Lipschitz only with respect to  $(\xi, \xi')$ . This is lengthy but straightforward and we do not give any details.

## References

- [1] G. Albi, Y.-P. Choi, M. Fornasier, D. Kalise, *Mean field control hierarchy*, Appl. Math. Optim. **76** (2017), no. 1, 93–135.
- [2] G. Albi, L. Pareschi, *Modeling of self-organized systems interacting with a few individuals: from microscopic to macroscopic dynamics*, Applied Mathematics Letters **401** (2013), no. 26, 397–401.

- [3] L. Ambrosio, N. Gigli, G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Second edition. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008. x+334 pp.
- [4] 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.
- [5] 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.
- [6] N. Ayi, N. Pouradier Duteil, *Graph limit for interacting particle systems on weighted random graphs*, arXiv:2307.12801 (2023).
- [7] 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.
- [8] 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.
- [9] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., Wiley, 1999.
- [10] 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.
- [11] V.I. Bogachev, *Measure theory*, Vol. I, II. Springer-Verlag, Berlin, 2007. Vol. I: xviii+500 pp., Vol. II: xiv+575 pp.
- [12] F. Bolley, J.A. Canizo, J.A. Carrillo, *Stochastic mean-field limit: non-Lipschitz forces and swarming*, Math. Models Methods Appl. Sci. **21** (2011), no. 11, 2179–2210.
- [13] B. Bonnet, N. Pouradier Duteil, M. Sigalotti, *Consensus formation in first-order graphon models with time-varying topologies*, Math. Models Methods Appl. Sci. **32** (2022), no. 11, 2121–2188.
- [14] 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.
- [15] M. Caponigro, B. Piccoli, F. Rossi, E. Trélat, *Mean-field sparse Jurdjevic-Quinn control*, Math. Models Methods Appl. Sci. **27** (2017), no. 7, 1223–1253.
- [16] J.A. Cañizo, J.A. Carrillo, J. Rosado, *A well-posedness theory in measures for some kinetic models of collective motion*, Math. Models Methods Appl. Sci. **21** (2011), no. 3, 515–539.
- [17] 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.
- [18] 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.
- [19] J.A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, D. Slepcev, *Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations*, Duke Math. J. **156** (2011), no. 2, 229–271.



- [20] J. A. Carrillo, M. Fornasier, J. Rosado, G. Toscani, *Asymptotic flocking dynamics for the kinetic Cucker-Smale model*, SIAM J. Math. Anal. **42** (2010), 218–236.
- [21] 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.
- [22] H. Chiba, G.S. Medvedev, *The mean field analysis of the Kuramoto model on graphs I. The mean field equation and transition point formulas*, Discrete Contin. Dyn. Syst. **39** (2019), no. 1, 131–155.
- [23] H.-H. Cotte, P.D. Koumoutsakos, *Vortex methods*, Cambridge University Press, Cambridge, 2000, xiv+313 pp.
- [24] F. Cucker, S. Smale, *Emergent behavior in flocks*, IEEE Trans. Automat. Control **52** (2007), no. 5, 852–862.
- [25] 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.
- [26] 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.
- [27] R.L. Dobrushin, *Vlasov equations*, Funct. Anal. Appl. **13** (1979), 115–123.
- [28] 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.
- [29] 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.
- [30] M. Fornasier, F. Solombrino, *Mean-field optimal control*, ESAIM Control Optim. Calc. Var. **20** (2014), no. 4, 1123–1152.
- [31] 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.
- [32] D. H. Fremlin, *Measure theory, Vol. 2, Broad foundations*, Corrected second printing of the 2001 original. Torres Fremlin, Colchester, 2003. 563+12 pp.
- [33] 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.
- [34] B. Geshkovski, C. Letrouit, Y. Polyanskiy, P. Rigollet, *A mathematical perspective on transformers*, arXiv:2312.10794 (2023).
- [35] F. Golse, *On the dynamics of large particle systems in the mean field limit*, in: Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity, 1–144. Lect. Notes Appl. Math. Mech., 3, Springer, Cham, 2016.
- [36] F. Golse, C. Mouhot, T. Paul, *On the mean-field and classical limits of quantum mechanics*, Comm. Math. Physics **343** (2016), 165–205.

- [37] S.-Y. Ha, J.-G. Liu, *A simple proof of the Cucker-Smale flocking dynamics and mean-field limit*, Commun. Math. Sci. **7** (2009), no. 2, 297–325.
- [38] S.-Y. Ha, E. Tadmor, *From particle to kinetic and hydrodynamic description of flocking*, Kinet. Relat. Models **1** (2008), 415–435.
- [39] P. Hartman, *Ordinary differential equations*, Birkhäuser, Boston, MA, 1982. xv+612 pp.
- [40] 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.
- [41] M. Hauray, P.-E. Jabin, *N-particles approximation of the Vlasov equations with singular potential*, Arch. Ration. Mech. Anal. **183** (2007), no. 3, 489–524.
- [42] R. Hegselmann, U. Krause, *Opinion dynamics and bounded confidence: models, analysis and simulation*, J. Artif. Soc. Soc. Sim. **5** (2002), no. 3.
- [43] P.-E. Jabin, *A review of the mean field limits for Vlasov equations*, Kinet. Relat. Models **7** (2014), no. 4, 661–711.
- [44] P.-E. Jabin, D. Poyato, J. Soler, *Mean-field limit of non-exchangeable systems*, arXiv:2112.15406 (2022).
- [45] P.-E. Jabin, Z. Wang, *Mean field limit and propagation of chaos for Vlasov systems with bounded forces*, J. Funct. Anal. **271** (2016), no. 12, 3588–3627.
- [46] P.-E. Jabin, Z. Wang, *Quantitative estimates of propagation of chaos for stochastic systems with  $W^{-1,\infty}$  kernels*, Invent. Math. **214** (2018), no. 1, 523–591.
- [47] M. Kac, *Foundations of kinetic theory*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III, 171–197. University of California Press, Berkeley (1956).
- [48] D. Kaliuzhnyi-Verbovetskyi, G.S. Medvedev, *The mean field equation for the Kuramoto model on graph sequences with non-Lipschitz limit*, SIAM J. Math. Anal. **50** (2018), no. 3, 2441–2465.
- [49] Y. Kuramoto, *Self-entrainment of a population of coupled non-linear oscillators*, International Symposium on Mathematical Problems in Theoretical Physics (Kyoto Univ., Kyoto, 1975), Springer, Berlin, Lecture Notes in Phys., **39** (1975), 420–422.
- [50] C. Lancellotti, *On the Vlasov limit for systems of nonlinearly coupled oscillators without noise*, Transport Theory and Statistical Physics **34** (2005), no. 7, 523–535.
- [51] E. Mariucci, M. Reiß, *Wasserstein and total variation distance between marginals of Lévy processes*, Electron. J. Stat. **12** (2018), no. 2, 2482–2514.
- [52] R.J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J. **80** (1995), no. 2, 309–323.
- [53] H.P. McKean, *Propagation of chaos for a class of non-linear parabolic equations*, Stochastic Differential Equations, Lecture Series in Differential Equations, Session 7, 1967, 41–57.
- [54] G.S. Medvedev, *The nonlinear heat equation on  $W$ -random graphs*, Arch. Ration. Mech. Anal. **212** (2014), no. 3, 781–803.

- [55] G.S. Medvedev, *The nonlinear heat equation on dense graphs and graph limits* SIAM J. Math. Anal. **46** (2014), no. 4, 2743–2766.
- [56] S. Mischler, C. Mouhot, *Kac’s program in kinetic theory*, Invent. Math. **193** (2013), no. 1, 1–147.
- [57] S. Motsch, E. Tadmor, *Heterophilous dynamics enhances consensus*, SIAM Rev. **56** (2014), no. 4, 577–621.
- [58] 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.
- [59] R. Natalini, T. Paul, *The mean-field limit for hybrid models of collective motions with chemotaxis*, SIAM J. Math. Anal. **55** (2023), no. 2, 900–928.
- [60] R. Natalini, T. Paul, M. Menci, *Microscopic, kinetic and hydrodynamic hybrid models of collective motions with chemotaxis: a numerical study*, to appear in Math. Mech. Complex Syst. (MEMOCS).
- [61] H. Neunzert, *An introduction to the nonlinear Boltzmann-Vlasov equation*, Kinetic theories and the Boltzmann equation (Montecatini, 1981), 60–110, Lecture Notes in Math., 1048, Springer, Berlin, 1984.
- [62] T. Paul, E. Trélat, *From microscopic to macroscopic scale equations: mean field, hydrodynamic and graph limits*, Preprint arXiv:2209.08832.
- [63] T. Paul, E. Trélat, *Universal approximation of PDEs by finite distinguishable particle systems*, in preparation.
- [64] B. Piccoli, N. Pouradier Duteil, E. Trélat, *Sparse control of Hegselmann-Krause models: black hole and declustering*, SIAM J. Control Optim. **57** (2019), no. 4, 2628–2659.
- [65] B. Piccoli, F. Rossi, *Transport equation with nonlocal velocity in Wasserstein spaces: convergence of numerical schemes*, Acta Appl. Math. **124** (2013), 73–105.
- [66] 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.
- [67] B. Piccoli, F. Rossi, *On properties of the generalized Wasserstein distance*, Arch. Ration. Mech. Anal. **222** (2016), no. 3, 1339–1365.
- [68] 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.
- [69] M. Pulvirenti, S. Simonella, *On the evolution of the empirical measure for the hard-sphere dynamics*, Bull. Inst. Math. Acad. Sin. (N.S.) **10** (2015), no. 2, 171–204.
- [70] P.-A. Raviart, *An analysis of particle methods*, Numerical methods in fluid dynamics (Como, 1983), 243–324. Lecture Notes in Math., 1127, Springer-Verlag, Berlin, 1985.
- [71] G.M. Rotskoff, E. Vanden-Eijnden, *Trainability and accuracy of artificial neural networks: an interacting particle system approach*, Comm. Pure Appl. Math. **75** (2022), no. 9, 1889–1935.
- [72] H. Sakaguchi, *Cooperative phenomena in coupled oscillator systems under external fields*, Prog. Theor. Phys. **79** (1988), 39–46.

- [73] 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.
- [74] S. Serfaty, *Mean field limit for Coulomb-type flows*, With an appendix by Mitia Duerinckx and Sylvia Serfaty, *Duke Math. J.* **169** (2020), no. 15, 2887–2935.
- [75] H. Spohn, *Large scale dynamics of interacting particles*, Texts and Monographs in Physics, 1991, Berlin, Springer-Verlag, xi+342 pp.
- [76] A.-S. Sznitman, *Topics in propagation of chaos*, Ecole d’été de Probabilités de Saint-Flour XIX, 1989, 165–251. Lecture Notes in Math., 1464, Springer-Verlag, Berlin, 1991.
- [77] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003. xvi+370 pp.
- [78] C. Villani, *Optimal transport. Old and new*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 338. Springer-Verlag, Berlin, 2009. xxii+973 pp.
- [79] A.A. Vlasov, *On the kinetic theory of an assembly of particles with collective interaction*, *Acad. Sci. USSR. J. Phys.* **9** (1945), 25–40.